On Weak Approximation of Stochastic Differential Equations through Hard Bounds by Mathematical Programming

KENJI KASHIMA* AND REIICHIRO KAWAI†

Abstract
Upper and lower hard bounds of the expected value on stochastic differential equations can be obtained with the help of the mathematical programming and the Dynkin formula, without recourse to Monte Carlo sample paths simulation. In this paper, we show that feasible solutions of those optimization approaches further provide useful additional information. Namely, feasible solutions provide upper and lower bounds for arbitrary intermediate times and/or different initial states. We also show that the optimization approach can also be applied to stochastic differential equations with a random initial state. Our theoretical analyses are illustrated by numerical results on the survival probability for a square-root diffusion, on the moment estimation of a Doléans-Dade stochastic exponential with jumps, and on pricing of a barrier basket option under the multi-asset Black-Scholes model.

Keywords: Dynkin formula, semidefinite programming, sum-of-squares relaxation, stochastic differential equation, weak approximation.

2010 Mathematics Subject Classification: 60H10, 65C30, 60G51, 90C22.

1 Introduction

Stochastic differential equations have often been employed to build realistic models in various fields, such as economics, finance, biology, the social sciences, chemistry and physics. For practical applications such as moment and tail probability estimations, we need to estimate the expected value on stochastic differential equations. The weak approximation via the time discretization of Euler-Maruyama type has long been the standard approach, that is, \( |E[V(X_T)] - E[V(X_T^\Delta)]| \leq C \Delta^\beta \), where \( \{X_t : t \in [0,T]\} \) is a solution of the stochastic differential equation of interest, \( V \) is a smooth function and \( X_T^\Delta \) indicates the time discretization approximation of the marginal \( X_T \) with time step \( \Delta > 0 \). The theoretical investigation of time discretization schemes in diffusion settings has been thoroughly presented in Kloeden and Platen [11], while stochastic differential equations with jumps have also been studied, for example, in Protter [17].

In contrast to Monte Carlo simulations of the time discretized sample paths, methodologies leading to the computation of bounds for the expectation have been investigated by several authors, for example, Bertsimas, Popescu and Sethuraman [3], Helmes, Röhll and Stockbridge [5], Lasserre and Prieto-Rumeau [12], Suzuki, Miyoshi and Kojima [19] to mention just a few. In principle, they maximize and minimize the expected value of interest where the underlying probability measure implicitly acts as a 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.

Meanwhile, the formulation developed by Primbs [16] and the authors [7,8,10] is known to be the dual of the method-of-moments approach in the context of mathematical programming, which we describe in brief. First, one employs a function with arguments both in time and in state, which bounds the value function at terminal time from above (or from below) over the support of the marginal \( X_T \). One further restricts the infinitesimal generator to be non-negative (or non-positive). Under all the constraints, the Dynkin formula guarantees that the bounding function at the deterministic initial point \((0,X_0)\) serves as an upper (or lower) bound of the expected value. 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 problem is solvable. A general class of stochastic differential equations with jumps is considered by the authors [7], for which the standard Monte Carlo simulations are not implementable due to the expensive computing cost or often even due to the non-availability of simulation methods. The formulation requires underlying stochastic differential equations to have the marginal with finite moments of up to a sufficiently large polynomial order. This moment condition may rule out interesting problem settings, such as stochastic differential equations driven by stable Lévy processes and the Heston stochastic volatility model. This issue was addressed in [8] by employing an exponential tempering on bounding functions to avoid the polynomial explosion which necessarily occurs at infinity without the exponential tempering.

*Email address: kashima@sys.es.osaka-u.ac.jp. Postal address: Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, 560-8531, Japan.
†Corresponding Author. Email address: reiichiro.kawai@maths.usyd.edu.au. Postal address: School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia. This work was carried out largely while RK was based at University of Leicester, UK.
In this paper, we show that feasible solutions obtained through our approach in [7, 8] provide some additional information of direct practical use. Namely, the bounding functions serve as upper and lower bounds for arbitrary intermediate times and/or for different initial states. By using this property, we can also reduce the optimality gap for an arbitrary intermediate time and/or for a different initial state, in return for giving up some portion of the tightness of the optimality gap at the terminal time and an originally set initial state. We also show that our optimization approach can also be applied to stochastic differential equations with a random initial state.

The rest of this paper is organized as follows. Section 2 is devoted to recalling the basic idea of the approach of [7, 8] based upon the mathematical programming and the Dynkin formula. Section 3 presents the main results; the weak approximations for arbitrary intermediate times and/or different initial states, and for the random initial state. Section 4 illustrates our results through numerical examples on estimation of the survival probability for a Gaussian Ornstein-Uhlenbeck process and on pricing of a barrier option under the Black-Scholes model. Finally, Section 5 concludes and indicates the direction of future research.

2 Problem Setup

Let us begin with general notations which will be used throughout the text. We define $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$ and $\mathbb{R}_+ := (0, +\infty)$ and write $\mathcal{B}(A)$ for the Borel $\sigma$-field of a set $A \subseteq \mathbb{R}$. For $k \in \mathbb{N}$, $\partial_k$ indicates the partial derivative with respect to $k$-th argument. We denote by $C_{k_1, k_2}^{k_3, k_4}$ the class of continuous functions which are $k_1$-times continuously differentiable with respect to the first variable and $k_2$-times continuously differentiable with respect to the second variable. We denote by $C_p$ the class of polynomial functions in the form of

$$f_p(t, x) = \sum_{B(0, 0)} c_{k_1, k_2} t^{k_1} x^{k_2},$$

(2.1)

where

$$B(l, m) := \{ (k_1, k_2) \in \mathbb{N}^2 : k_1 \geq l, k_2 \geq m, k_1 + k_2 \leq K \},$$

(2.2)

for a fixed even natural number $K$, while $\{ c_{k_1, k_2} \}_{B(0, 0)}$ 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 henceforth fix $(\Omega, \mathcal{F}, \mathbb{P})$ as our underlying probability space.

Let $X_0$ be in $\mathbb{R}$ a.s., and fix $T > 0$. Consider a one-dimensional stochastic differential equation

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

(2.3)

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 where $\mu$ is a Poisson random measure on $\mathbb{R}_0$ whose compensator is given by the Lévy measure $\nu$, that is, a $\sigma$-finite measure defined on $\mathbb{R}_0$ satisfying $\int_{|z| \leq 1} \nu(dz) < +\infty$. Here, we assume that for each $t \in [0, T]$, the coefficients $a_0(t, x)$, $a_1(t, x)$ and $b(t, x, z)$ in (2.3) satisfy the usual conditions such as at most linear growth and Lipschitz so that the solution of (2.3) is well defined. (For example, see Section 6.2 of Applebaum [2].) We henceforth equip our underlying probability space with the natural filtration $(\mathcal{F}_t)_{t \in [0, T]}$ generated by $\{ X_t : t \in [0, T] \}$. Moreover, we use the notation

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

where $\{ X_t : t \in [0, T] \}$ is defined in (2.3). Note that by imposing that the set $\mathcal{X}$ is convex, 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 $\mathcal{X} = [0, +\infty)$. This larger space will be required for optimization problems.

Our interest throughout this study is in approximating the expectation $\mathbb{E}[V(\tau, X_\tau)]$, where $V$ is a suitable function from $[0, T] \times \mathbb{R}$ to $\mathbb{R}$ such that $\mathbb{E}[V(\tau, X_\tau))] < +\infty$, and where $\tau$ is an $(\mathcal{F}_t)_{t \in [0, T]}$-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 the Poisson jump component. Meanwhile, our approach relies on the mathematical programming and is free of the random number generation, which we will formulate as follows. For $f \in C^{1,2}([0, T] \times \mathcal{X}; \mathbb{R})$, the Ito formula reads

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

where

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

(2.4)

and for $z \in \mathbb{R}_0$,

$$B_z f(t, x) := f(t, x + b(t, x, z)) - f(t, x),$$

(2.5)
provided that for each \( t \in [0, T] \) and each \( x \in \mathcal{X} \), \( \int_{|z|>1} |B_z f(t, x)| \, \mathbb{V}(dz) < +\infty \). A key building block is the so-called Dynkin formula; for any \( (\mathcal{F}_t)_{t \in [0,T]} \)-stopping time \( \tau \) taking values in \( [0,T] \),

\[
\mathbb{E} [f(\tau, X_\tau)] - f(0, X_0) = \mathbb{E} \left[ \int_0^\tau \mathcal{A} f(s, X_s) \, ds \right].
\]  

(2.6)

Throughout this paper, we write

\[
E_0 := \inf \{ B \subseteq [0,T] \times \mathbb{R} : \mathbb{P}((\tau, X_\tau) \in B) = 1, B \text{ convex} \},
\]

\[
E_1 := \inf \{ B \subseteq [0,T] \times \mathbb{R} : \mathbb{P}((\tau, X_\tau) \in B, t \in [0, \tau]) = 1, B \text{ convex} \},
\]

\[
E_2 := E_0 \cup E_1.
\]

We briefly summarize the conditions under which the formula makes sense. (See [8] for its proof.)

**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 \( \mathcal{A} f(t,x) \) in (2.4) is well defined. Then, the Dynkin formula (2.6) holds if at least one of the following conditions is satisfied:

(i) \( \mathbb{E} [f(\tau, X_\tau)] \leq +\infty \), and for each \( t \in [0, \tau) \), \( \mathbb{E} \left[ \mathcal{A} f(t, X_t) \right] \leq +\infty \), a.s.,

(ii) \( \mathbb{E} \left[ \int_0^\tau (\mathcal{A} f(s, X_s)a_1(s, X_s))^2 \, ds \right] < +\infty \) and \( \mathbb{E} \left[ \int_0^\tau B_2 f(s, X_s) \mathbb{V}(dz) \, ds \right] < +\infty \).

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 \( \tau \) is the first exit time for a bounded subset of \( \mathcal{X} \), and so on. Those conditions are no longer readily valid when the Poisson jump is involved, due to the local difference operator \( B_2 \) of (2.5).

**Proposition 2.2.** Let \( \mathbb{F} \) be a subclass of \( C^{1,2}(E_2; \mathbb{R}) \)-functions for which the Dynkin formula (2.6) holds and assume that \( \mathbb{F} \) is not empty.

(i) If \( f \in \mathbb{F} \) satisfies

\[
f(t,x) \geq V(t,x) \quad \text{on } E_0,
\]

and if there exists a measurable function \( h_1 \) mapping from \( E_2 \) to \( \mathbb{R} \) such that

\[
\mathcal{A} f(t,x) \leq h_1(t,x) \quad \text{on } E_1,
\]

then it holds that

\[
\mathbb{E} [V(\tau, X_\tau)] \leq f(0, X_0) + \mathbb{E} \left[ \int_0^\tau h_1(t, X_t) \, dt \right].
\]

(2.10)

(ii) If \( g \in \mathbb{F} \) satisfies

\[
g(t,x) \leq V(t,x) \quad \text{on } E_0,
\]

and if there exists a measurable function \( h_2 \) mapping from \( E_2 \) to \( \mathbb{R} \) such that

\[
\mathcal{A} g(t,x) \geq h_2(t,x) \quad \text{on } E_1,
\]

then it holds that

\[
\mathbb{E} [V(\tau, X_\tau)] \geq g(0, X_0) + \mathbb{E} \left[ \int_0^\tau h_2(t, X_t) \, dt \right].
\]

Proof. The desired result of (i) holds by

\[
\mathbb{E} [V(\tau, X_\tau)] \leq \mathbb{E} [f(\tau, X_\tau)] = f(0, X_0) + \mathbb{E} \left[ \int_0^\tau \mathcal{A} f(t, X_t) \, dt \right] \leq f(0, X_0) + \mathbb{E} \left[ \int_0^\tau h_1(t, X_t) \, dt \right].
\]

where the inequality in the first line holds by (2.8), the equality in the second line holds by (2.6) and the last inequality holds by (2.9). The claim (ii) can be proved in a similar manner.

### 3 Main Results

We are now in a position to discuss our main results. In this section, we show that feasible solutions obtained through the optimization approach described in Section 2 provide some additional information of direct practical use, when underlying stochastic differential equations are homogeneous in time. Namely, the bounding functions serve as upper and lower bounds for arbitrary intermediate times and/or for different initial states. By using this property, we can also reduce the optimality gap for an arbitrary intermediate time and/or for a different initial state, in return for giving up some portion of the tightness of the optimality gap at the terminal time and
an originally set initial state. We also show that our optimization approach can also be applied to stochastic differential equations with random initial state.

Recall that $\mathcal{F}$ is a subclass of $C^{1,2}(E_2; \mathbb{R})$-functions for which the Dynkin formula (2.6) holds. We henceforth write

$$\tau_t := \tau \wedge (T - t),$$  \hspace{1cm} (3.1)

that is, the stopping time condensed into the shorter interval $[0, T - t]$. Clearly, $\tau_0 = \tau$. For later convenience, we write

$$E_3 := [0, T] \times \mathcal{F},$$
$$E_4 := \inf \{ B \subseteq \mathcal{F} : \mathbb{P}(X_t \in B, t \in [0, \tau]) = 1, B \text{ connected} \},$$
$$E_5(t) := \inf \{ B \subseteq E_3 : \mathbb{P}((\tau_1, X_1) \in B) = 1, B \text{ convex} \}, \quad t \in [0, T].$$

Finally, we use the notations $\mathbb{E}_x$ and $\mathbb{P}_x$, when emphasizing that the expectation and the probability are considered under which the initial state of the stochastic differential equation (2.2) is set deterministically $X_0 = x$.

### 3.1 Simultaneous Weak Approximation of Intermediate Times

Let us first state the following result indicating that the function $f \in \mathcal{F}$ in Proposition 2.2 can also serve as bounds at arbitrary intermediate time points.  

**Theorem 3.1.** Assume the following:

(i) The coefficients $a_0(t, x)$, $a_1(t, x)$ and $b(t, x, z)$ in (2.7) are independent of $t$,
(ii) the function $f \in \mathcal{F}$ satisfies (2.8) and (2.9) with $h_1 \equiv 0$,
(iii) there exists $t \in [0, T]$ such that for every $(s, x) \in E_5(t)$, the point $(t + s, x)$ is in $E_0$.

It then holds that

$$\mathbb{E}[V(t + \tau_t, X_{\tau_1})] \leq f(t, X_0).$$

**Remark 3.2.** Before proceeding to the proof, let us remark a few important points. First, the assumption (i) is crucial when considering intermediate times, while it is redundant otherwise. Such problem settings will be investigated later in Corollary 3.8 and Proposition 3.9. Next, the assumption (ii) can be relaxed to the case of non-zero $h_i(t, x)$, while this generalization does not add much practical value. Finally, by the time in the left hand side of the above inequality, that is, $t + \tau_t = (t + \tau) \wedge T$, we indicate that the underlying stochastic process restarts at time $t$ and consider the remaining time interval $(t, T]$. Our proof below is probabilistic based on such conditioning.

**Proof.** Let $x \in E_4$ and define the event $A := \{ \tau \geq t \} \cap \{ X_t = x \}$. First, note that

$$\mathbb{E} \left[ V(\tau, X_{\tau_1}) | A \right] = \mathbb{E}_x \left[ V(t + \tau, X_{\tau_1}) | A \right],$$

due to the time homogeneity and the Markovian property of $\{ X_t : t \in [0, T] \}$. Next, it holds by (2.8) that given the event $A$,

$$\mathbb{E} \left[ V(\tau, X_{\tau_1}) | A \right] \leq \mathbb{E} \left[ f(\tau, X_{\tau_1}) | A \right].$$

Finally, again given the event $A$, we have

$$\mathbb{E} \left[ f(\tau, X_{\tau_1}) | A \right] = f(t, x) + \mathbb{E} \left[ \int_t^\tau \mathcal{A} f(s, X_s) ds | A \right] \leq f(t, x),$$

where the last inequality holds by the non-positivity of $\mathcal{A} f$ over $E_1$. This concludes the proof.

To give better intuition of the condition (iii) of Theorem 3.1, we provide Figure 1, where the left figure corresponds to a case where the condition is satisfied, while the right figure to a case violating the condition. In short, the problem setting over $(t, T]$ needs to be perfectly identical to that over $[0, T - t]$. Hence, the result requires the time-homogeneity of the underlying stochastic differential equation (2.2) as well as that of the sets $E_0$ and $E_1$.

**Remark 3.3.** A similar claim holds for the lower bound with the function $g$ in Proposition 2.2. Let $f$ and $g$ be functions in $\mathcal{F}$ of Proposition 2.2 with $h_i \equiv 0$. It then obviously holds that

$$g(0, X_0) \leq \mathbb{E}[V(\tau, X_{\tau_1})] \leq f(0, X_0),$$

while Theorem 3.1 guarantees that feasible functions $f$ and $g$ also serve as bounds of the expected value $\mathbb{E}[V(t + \tau, X_{\tau_1})]$ for every $t \in [0, T]$ in such a way that

$$g(t, X_0) \leq \mathbb{E}[V(t + \tau, X_{\tau_1})] \leq f(t, X_0).$$

As described in Remark 3.2, the time $t + \tau_t$ can be thought of as a stopped timing after the underlying stochastic process restarts at time $t$.

\[ \square \]
Based on the definition \((3.1)\) of the stopping time, we have derived Theorem \(3.1\) in a backward manner (that is, \(T - t\)) using the conditioning argument. Note that the result can also be formulated in a forward manner, as given in Kashima and Kawai [9]. For the reader’s convenience, we state the result in this way and give a different proof without recourse to the probabilistic conditioning argument.

**Corollary 3.4.** For each \(t \in [0,T]\), define \(\tau(t) := \tau \wedge t\), which is the stopping time condensed into the shorter interval \([0,t]\), and define for each \(t \in [0,T]\),

\[
E_0(t) := \inf \{ B \subseteq [0,T] \times \mathbb{R} : \mathbb{P}( (\tau(t), X_{\tau(t)}) \in B ) = 1, B \text{ convex} \} ,
\]

\[
E_1(t) := \inf \{ B \subseteq [0,T] \times \mathbb{R} : \mathbb{P}( (s, X_s) \in B, s \in [0, \tau(t)) ) = 1, B \text{ convex} \},
\]

in accordance with the sets defined in \((2.7)\). Impose the conditions (i) and (ii) of Theorem \(3.1\) and assume that for each \((s,x) \in E_0(t)\), the following two

\[
(s + (T - t), x) \in E_0,
\]

\[
V(s,x) = V(s + (T - t), x),
\]

hold. Then, it holds that

\[
\mathbb{E} \left[ V(\tau(t), X_{\tau(t)}) \right] \leq f(T - t, X_0).
\]

**Proof.** Define \(\tilde{f}(s,x) := f(s + (T - t), x)\). Due to the time homogeneity, we have

\[
\mathcal{A} \tilde{f}(s,x) = \mathcal{A} f(s + (T - t), x) \leq 0, \quad (s,x) \in E_1(t),
\]

where the last inequality holds by \((2.9)\) the inclusion \(E_1(t) \subseteq E_1\). We also have

\[
\tilde{f}(s,x) \leq V(s + (T - t), x) = V(s,x), \quad (s,x) \in E_0(t),
\]

where the first inequality holds by \((2.9)\) and \((3.2)\). By combining these inequalities with the Dynkin formula \((2.6)\), we get

\[
\mathbb{E} \left[ V(\tau(t), X_{\tau(t)}) \right] \leq \mathbb{E} \left[ \tilde{f}(\tau(t), X_{\tau(t)}) \right] = \tilde{f}(0, X_0) + \mathbb{E} \left[ \int_0^{\tau(t)} \mathcal{A} \tilde{f}(s, X_s) \, ds \right] \leq f(T - t, x),
\]

which concludes. 

Before proceeding to the next section, let us state a direct consequence of Theorem \(3.1\) in the case where the value function \(V\) is homogeneous in time or is independent of time.

**Corollary 3.5.** In the setting of Theorem \(3.1\) if \(V(s,x) = V(t + s,x)\) for \((s,x) \in [0,T - t] \times \mathcal{X}\), then it holds that

\[
\mathbb{E} [V(\tau(t), X_{\tau(t)})] \leq f(t, X_0).
\]

In particular, if there exists \(\tilde{V} : \mathcal{X} \rightarrow \mathbb{R}\) such that \(V(s,x) = \tilde{V}(x)\) for \((s,x) \in E_0\), then it holds that

\[
\mathbb{E} \left[ \tilde{V}(X_0) \right] \leq f(t, X_0).
\]
In order for Proposition 2.2 and Theorem 3.1 to be of practical use, we need to be able to find functions $f$ and $g$ satisfying the conditions (2.8), (2.9), (2.11) and (2.12), or much more preferably to find a minimized upper bound and a maximized lower bound through the optimization problems

$$
\begin{align*}
\min_{f \in \mathcal{F}} & \quad f(0, X_0) \\
\text{s.t.} & \quad f(t, x) \geq V(t, x) \quad \text{on } E_0, \\
& \quad \sigma f(t, x) \leq 0 \quad \text{on } E_1,
\end{align*}
$$
\begin{align*}
\max_{g \in \mathcal{G}} & \quad g(0, X_0) \\
\text{s.t.} & \quad g(t, x) \leq V(t, x) \quad \text{on } E_0, \\
& \quad \sigma g(t, x) \geq 0 \quad \text{on } E_1.
\end{align*}

(3.3)

Suppose for a moment that we are capable of finding optimal solutions of the optimization problems (3.3) with $f^*$ and $g^*$ being optimal solutions, respectively. (We will come back to this point in Section 4.)

### 3.2 Weak Approximation with Different Time Points and Initial States

We next show that the gap at an intermediate time, that is $f(t, X_0) - g(t, X_0)$ in the context of Theorem 3.1 may be tightened, in return for giving up some portion of the tightness of the gap at the terminal time, that is, $f(0, X_0) - g(0, X_0)$.

**Theorem 3.6.** Let $\varphi_1$ and $\varphi_2$ be finite measures defined on $E_3$. Consider the following optimization problems

$$
\begin{align*}
\min_{f \in \mathcal{F}} & \quad \int_{E_3} f(t, x) \varphi_1(dt, dx) \\
\text{s.t.} & \quad f(t, x) \geq V(t, x) \quad \text{on } E_0, \\
& \quad \sigma f(t, x) \leq 0 \quad \text{on } E_1,
\end{align*}
$$

and

$$
\begin{align*}
\min_{f \in \mathcal{F}} & \quad \int_{E_3} f(t, x) (\varphi_1 + \varphi_2)(dt, dx) \\
\text{s.t.} & \quad f(t, x) \geq V(t, x) \quad \text{on } E_0, \\
& \quad \sigma f(t, x) \leq 0 \quad \text{on } E_1,
\end{align*}
$$

(3.4)

(3.5)

and assume that both are feasible with optimal solutions $f^*$ for (3.4) and $f^*$ for (3.5). Then, it holds that

$$
\int_{E_3} f^*(t, x) \varphi_1(dt, dx) \leq \int_{E_3} f^*(t, x) \varphi_1(dt, dx),
$$

(3.6)

and

$$
\int_{E_3} f^*(t, x) \varphi_2(dt, dx) \geq \int_{E_3} f^*(t, x) \varphi_2(dt, dx),
$$

(3.7)

with equality when there exists a positive constant $c$ satisfying $\varphi_1 = c \varphi_2$.

**Proof.** We only consider the case $\varphi_1 \neq c \varphi_2$ for any $c > 0$, since the optimization problems (3.4) and (3.5) are simply equivalent otherwise.

The inequality (3.6) is nothing but the definition of the optimization problem (3.4). To derive the inequality (3.7), observe that

$$
\int_{E_3} f^*(t, x) (\varphi_1 + \varphi_2)(dt, dx) \leq \int_{E_3} f^*(t, x) (\varphi_1 + \varphi_2)(dt, dx),
$$

by the definition of the optimization problem (3.5). By combining this with (3.6), we get (3.7). \qed

Let us here clarify what the optimization problem (3.4) is looking for. First, the optimal solution $f^*$ satisfies that for each $(t, x) \in E_2$,

$$
E_x [V(\tau_t, X_\tau_t)] \leq E_x [f^*(\tau_t, X_\tau_t)] = f^*(t, x) + E_x \left[ \int_0^{\tau_t} \sigma f^*(s, X_s)ds \right].
$$

By integrating the above equation with respect to the finite measure $\varphi_1$ over $E_2$, we get

$$
\int_{E_2} E_x [V(\tau_t, X_\tau_t)] \varphi_1(dt, dx) \leq \int_{E_2} E_x [f^*(\tau_t, X_\tau_t)] \varphi_1(dt, dx),
$$

\begin{align*}
&\leq \int_{E_2} f^*(t, x) \varphi_1(dt, dx) + \int_{E_2} E_x \left[ \int_0^{\tau_t} \sigma f^*(s, X_s)ds \right] \varphi_1(dt, dx).
\end{align*}

This relation indicates that the optimization problem (3.4) is aimed at minimizing an upper bound of $\int_{E_2} E_x [V(\tau_t, X_\tau_t)] \varphi_1(dt, dx)$. It is noteworthy that Theorem 3.6 does not require time-homogeneity of the underlying stochastic differential equation (2.3).

The following corollary demonstrates one example of the above result for practical use.
Corollary 3.7. (Simultaneous Weak Approximation at Different Time Points) Consider the optimization problems (3.4) and (3.5), impose all the conditions in Theorem 3.1 and 3.6, fix $s_1$ and $s_2$ in $[0, T]$ and set

$$
\varphi_1(dt, dx) = \delta_{[s_1 \times [0,T]}(dt, dx),
\varphi_2(dt, dx) = \delta_{[s_2 \times [0,T]}(dt, dx).
$$

Then, we have

$$
\mathbb{E}\left[V(s_1 + \tau_{s_1}, X_{\tau_{s_1}})\right] \leq f^*(s_1, X_0),
$$

and

$$
\mathbb{E}\left[V(s_2 + \tau_{s_2}, X_{\tau_{s_2}})\right] \leq f^*(s_2, X_0).
$$

In particular, if $\tau = T$, then it holds that

$$
\mathbb{E}\left[V(T, X_{T-t_0})\right] \leq f^*(s_1, X_0) \leq f^*(s_2, X_0).
$$

Proof. The inequality (3.8) and the second inequality of (3.9) are obvious from Theorem 3.1 and 3.6. To show first inequality of (3.9), let $f^0$ be the optimal solution of (3.4) when $\varphi_1(dt, dx) = \delta_{[s_1 \times [0,T]}(dt, dx)$. It holds by Theorem 3.6 that $f^0(s_2, X_0) \leq f^*(s_2, X_0)$, while $\mathbb{E}[V(s_2 + \tau_{s_2}, X_{\tau_{s_2}})] \leq f^0(s_2, X_0)$ since $f^0(t, x) \leq 0$ is guaranteed over $E_1$ and due to the time-homogeneity of $\{X_t : t \in [0, T]\}$.

The above result indicates that by replacing the objective function $f(0, X_0)$ with $f(0, X_0) + f(\delta, X_0)$, we can intentionally tighten the gap in the inequality $\mathbb{E}[V(s + \tau, X_\tau)] \leq f(s, X_0)$ for the estimation at the shortened stopping time $\tau$, through a single implementation of the optimization problem (3.5). We however have to give up some portion of the tightness of the inequality $\mathbb{E}[V(\tau, X_\tau)] \leq f(0, X_0)$ at the the original stopping time $\tau$. It is worthwhile noting the constant $s$ can be arbitrarily chosen.

Next, we demonstrate another practical use of Theorem 3.6. Note that the condition (i) of Theorem 3.1 (in other words, the time homogeneity) is not required here, as we do not consider any intermediate times.

Corollary 3.8. (Simultaneous Weak Approximation of Different Initial States) Consider the optimization problems (3.4) and (3.5), set the conditions (ii) and (iii) of Theorem 3.1 and set $\varphi_1(dt, dx) = \delta_{[0_1 \times [0,T]}(dt, dx)$ and $\varphi_2(dt, dx) = \delta_{[0_2 \times [0,T]}(dt, dx)$, where $x_1$ and $x_2$ are in $E_4$. Then, we have

$$
\mathbb{E}_{\alpha_1} [V(\tau, X_\tau)] \leq f^*(0, x_1) \leq f^*(0, x_2),
$$

and

$$
\mathbb{E}_{\alpha_2} [V(\tau, X_\tau)] \leq f^*(0, x_2) \leq f^*(0, x_2).
$$

Proof. This can be proved in a very similar manner to Corollary 3.7.

This result indicates that by introducing the objective function $f(0, x_1) + f(0, x_2)$ instead of solely $f(0, x_1)$, we can simultaneously deal with the estimation problem for $\mathbb{E}[V(T, X_T)]$ with different initial states through a single implementation of the optimization problem (3.5). Just as before, the gap for the initial state $x_2$ can be tight, while the result for the original initial state $x_1$ will no longer be as good as the one with objective function $f(0, x_1)$.

It may possibly yield some meaningful outcomes to further put weights on more components both in time and in initial state, for example, $f(0, x_1) + f(t, x_1) + f(t, x_2)$ for some $t \neq 0$ and $x_1 \neq x_2$. It seems difficult to predict the exact effectiveness in advance.

Note that the setting of Corollary 3.8 can also be interpreted as the one with the initial state of a stochastic differential equation being random, as the objective function is unique up to constant multiple. In particular, in Corollary 3.8 the initial state $X_0$ can be thought of as a uniform random variable taking values in $\{x_1, x_2\}$. Let us state this concept in a rigorous manner.

Proposition 3.9. (Weak Approximation with Random Initial State) Consider the stochastic differential equations (2.3) with $X_0$ being a random variable with distribution $F$ on $\Omega$ and measurable with respect to $\mathcal{F}_0$. Let $\varphi_3$ be a finite measure on $[0, T]$ and assume that for each $t \in [0, T]$, $t + x, x \in E_0$ for every $(x, s) \in E_2(t)$. Then, optimal solutions $f^*$ and $g^*$ for the optimization problems

\[
\begin{align*}
\min_{f \in F} \int_{E_1} f(t, x) \varphi_3(dt) F(dx) \\
\text{s.t.} \quad f(t, x) \geq V(t, x) \text{ on } E_0, \quad \text{and} \quad g(t, x) \leq 0 \text{ on } E_1,
\end{align*}
\]

satisfies

$$
\int_{E_3} g^*(t, x) \varphi_3(dt) F(dx) \leq \int_{[0, T]} \mathbb{E}[V(t + \tau, X_{\tau})] \varphi_3(dt) \leq \int_{E_3} f^*(t, x) \varphi_3(dt) F(dx).
$$

In particular, if $\varphi_3(dt) = \delta_{[0]}(dt)$, then it holds that

$$
\int_{\Omega} g^*(0, x) F(dx) \leq \mathbb{E}[V(\tau, X_\tau)] \leq \int_{\Omega} f^*(0, x) F(dx).
$$
Proof. By Theorem 3.1 it holds that
\[ \mathbb{E}_{t} [V(t + \tau, X_{\tau})] \leq \mathbb{E}_{t} [f^{*}(t + \tau, X_{\tau})] \leq f^{*}(t, x). \]
By integrating the above inequalities with respect to \( \Phi_{3}(dt)F(dx) \) on \( E_{3} \), we get
\[ \int_{E_{3}} \mathbb{E}_{t} [V(t + \tau, X_{\tau})] \Phi_{3}(dt)F(dx) \leq \int_{E_{3}} \mathbb{E}_{t} [f^{*}(t + \tau, X_{\tau})] \Phi_{3}(dt)F(dx) \leq \int_{E_{3}} f^{*}(t, x) \Phi_{3}(dt)F(dx), \]
With the equality
\[ \int_{\mathcal{X}} \mathbb{E}_{t} [V(t + \tau, X_{\tau})]F(dx) = \mathbb{E} [V(t + \tau, X_{\tau})], \]
the desired result holds.

It would be interesting to introduce randomness to the initial state, in particular for the case where a stochastic process of interest admits an asymptotic stationary distribution since then the stochastic process with initial state having the stationary distribution may have the same distribution for every time. Examples are Ornstein-Uhlenbeck processes with a Gaussian or a selfdecomposable stationary distribution. (The background driving process is a Wiener process in the former, while a pure-jump Lévy process in the latter.)

Remark 3.10. Since objective functions are equivalent up to positive constant factors, it certainly makes perfect sense to begin with a finite (not necessary probability) measure \( F(dx) \) in the formulation (3.10) and then interpret retrospectively that \( X_{0} \) is a random variable with distribution \( F(dx)/F(\mathcal{X}) \) defined on \( \mathcal{X} \). We will indeed encounter this situation in the next section. Although it is not necessary to assume that \( F(\mathcal{X}) \) is finite in order to assure that the formulation (3.10) is well defined itself, we can then no longer interpret that \( X_{0} \) is a random variable.

4 Implementation with Illustrative Examples

In general, the optimization problems of the previous section with respect to functions in the \( \mathbb{F} \) are intractable in practice, as the class \( \mathbb{F} \) is too abstract. To address this practical issue, the authors [7] restrict the bounding functions \( f \) and \( g \) to polynomial functions \( C_{p}(E_{2}; \mathbb{R}) \) in the form (2.1) to reduce the problems (3.3) to the so-called polynomial programming. To this end, assume that both \( a_{0} \) and \( a_{1} \) are polynomial and the coefficient \( b \) is decomposed as
\[ b(x, z) = b_{1}(x)b_{2}(z), \]
where \( b_{1} \) is polynomial and where \( b_{2} : \mathbb{R}_{0} \rightarrow \mathbb{R} \) such that \( \int_{\mathbb{B}_{0}} \mid b_{2}(z) \mid^{k} v(dz) < +\infty, k = 2, \ldots, K. \) By further assuming \( \mathbb{E} \mid X_{t} \mid^{K} < +\infty \) for \( t \in [0, T] \), it holds that
\[ \mathcal{A} f_{p}(t, x) = \sum_{B(0, 1)} c_{k, k_{1}} k_{1}^{k_{1} - 1} x^{k_{1}} + \sum_{B(0, 1)} c_{k, k_{1}} k_{1}^{k_{1} - 1} a_{0}(x) + \frac{1}{2} \sum_{B(0, 2)} c_{k, k_{1}} k_{1}^{k_{1} - 1} a_{1}(x)^{2} \]
\[ + \sum_{B(0, 2)} c_{k, k_{1}} k_{1}^{k_{1} - 2} \sum_{k = 0} k_{2} k_{2}^{k_{2} - 1} \int_{\mathbb{B}_{0}} b_{2}(z)^{k_{2} - k} v(dz), \]
that is polynomial in \( t \) and \( x \). Thus, the optimization problems
\[ \begin{array}{ll}
\min & f(0, X_{0}) \\
\text{s.t.} & f(t, x) \geq V(t, x) \text{ on } E_{0}, \quad \mathcal{A} f_{p}(t, x) \leq 0 \text{ on } E_{1}, \quad f \in C_{p}(E_{2}; \mathbb{R}), \end{array} \quad \begin{array}{ll}
\max & g(0, X_{0}) \\
\text{s.t.} & g(t, x) \leq V(t, x) \text{ on } E_{0}, \quad \mathcal{A} g(t, x) \geq 0 \text{ on } E_{1}, \quad g \in C_{p}(E_{2}; \mathbb{R}), \end{array} \]
are polynomial programming. Here, the decision variables are the coefficients \( \{ c_{k, k_{1}} \}_{B(0, 0)} \) in the definition of the polynomial (2.1).

It should be noted that there is no a priori theoretical error bound (for given \( K \)) available. That is, we cannot estimate the quality of the error before solving these optimization problems. Though we can show that the gap converges to zero as \( K \rightarrow \infty \) under suitable assumptions, this is not necessarily useful because the size of semi definite programming quickly increases. Further discussion on this subject is beyond the scope of this paper.

4.1 Survival Probability Estimation of Square-Root Diffusions

We first consider a survival probability estimation problem for the so-called square-root process, with a nonlinearity in coefficients. We set \( a_{0}(x) = \lambda (\theta - x), a_{1}(x) = \sigma \sqrt{x}, b_{1}(x) \equiv 0, b_{2}(z) \equiv 0, \) where \( \lambda > 0, \theta > 0 \) and \( \sigma > 0 \), so that the stochastic differential equation (2.3) reads
\[ dX_{t} = \lambda (\theta - X_{t}) dt + \sigma \sqrt{X_{t}} dW_{t}. \]
Suppose, for a moment, that \( X_0 \) is a positive constant. It will never be negative values and if \( 2\lambda \theta \geq \sigma^2 \), then it remains strictly positive. It is well known that

\[
\mathbb{E}[X_t] = e^{-\lambda t}X_0 + \theta \left( 1 - e^{-\lambda t} \right) \to \theta,
\]

as \( t \to +\infty \). The square-root diffusion has attracted great attention, in particular, of financial practitioners as a positive mean-reverting interest rate dynamics. The exact discrete sampling of the trajectory is possible but very expensive, while the Euler discretization requires some special care in such a way that the square root in the diffusion component is always well defined. (We refer the reader to Alfonsi \cite{alfonsi} and references therein for various issues around its sample paths generation.) In this subsection, we illustrate that our optimization approach and theorems are applicable to this square-root diffusion, without a prohibitive computation effort and a careful treatment of the square-root diffusion term.

Consider its finite horizon \( \{X_t : t \in [0, T]\} \) and the bounded set

\[ E_1 := [0, T] \times [0, U], \]

where \( U > 0 \) and \( X_0 \in [0, U] \). In what follows, we set \( \tau \) to be the \((\mathcal{F}_t)_{t \in [0, T]}\)-measurable 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 \((\tau, X_\tau)\) indicates the exit location, whose support \( E_0 \) can split into three disjoint sets

\[ E_U := [0, T] \times \{ U \}, \quad E_r := \{ T \} \times [0, U]. \]

It is enough to have \( E_U \cup E_r \) for the exit location, since sample paths are almost surely continuous and nonnegative. We mean by the survival probability the probability that the sample path reaches \( E_r \) without touching \( E_U \), that is, \( \mathbb{P}(\tau, X_\tau) \in E_r \). Hereafter, we fix \( \lambda = 1.0, \theta = 0.8, \sigma = 1.0, U = 1.0 \) and \( T = 0.4 \). This parameter setting indicates that the origin is unattainable. For illustration, we draw two typical sample paths in Figure 2. To formulate it in our framework, we set

\[ V(t, x) = \mathbbm{1}(t, x) \in E_r), \]

which yields

\[
\begin{align*}
\min_{f(0, X_0)} \quad & f(0, X_0) \\
\text{s.t.} \quad & f(T, x) \geq 1 \text{ on } [0, U], \\
& f(t, x) \geq 0 \text{ on } E_U, \quad \text{and} \\
& \mathcal{A} f(t, x) \leq 0 \text{ on } E_1, \\
& f \in C_p(E_1; \mathbb{R}), \\
\end{align*}
\]

\[
\begin{align*}
\max_{g(0, X_0)} \quad & g(0, X_0) \\
\text{s.t.} \quad & g(T, x) \leq 1 \text{ on } [0, U], \\
& g(t, x) \leq 0 \text{ on } E_U, \quad \text{and} \\
& \mathcal{A} g(t, x) \geq 0 \text{ on } E_1, \\
& g \in C_p(E_1; \mathbb{R}),
\end{align*}
\]

(4.4)

where the decision variables are the coefficients \( \{c_k, k\}_{k \geq 0} \) of the polynomials \( f \) and \( g \) as in (2.1) and where

\[ \mathcal{A} f(t, x) = \partial_t f(t, x) + \lambda (\theta - x) \partial_x f(t, x) + \frac{1}{2} \sigma^2 \lambda \partial^2_x f(t, x). \]

Figure 2: Two typical sample paths; the one (- - -) survives up to the time \( T \), while the other (-o-) exits from the set \( E_1 \) by the time \( T \).
As previously discussed, this optimization approach does not require the sample paths simulation at all towards the approximation of the expected value. Due to the curse of dimensionality of the semi-definite programming, we need to restrict the degree of polynomials in optimization problems. In our numerical experiments, we impose $c_{k,k_s} = 0$ whenever $k_t + k_s > 10$. Polynomial optimization problems are NP hard in general, while if the degrees of $f_p$ that is, $K$ is fixed (in (4.2)), sums-of-squares relaxation enables us to solve the problem efficiently. (For its complete details, see Parrilo [14].) Let us here briefly describe this relaxation procedure in a setting with $\mathcal{X} = \mathbb{R}_+$, the value function $V$ being polynomial and $c_{k_t,k_s} = 0$ for $k_t + k_s > 2n$, where $n$ is a positive integer. Let $q(t,x)$ be a vector consisting of $(n+1)(n+2)/2$ monomials of the form $t^{k_t}x^{k_s}$ such that $k_t \geq 0$, $k_s \geq 0$ and $k_t + k_s \leq n$. Instead of the above original minimization problem (4.4), we will solve

$$
\begin{align*}
\min & \quad f(0,X_0) \\
\text{s.t.} & \quad f(T,x) \geq 1 + (U - x)x(q(T,x),Q_1q(T,x)), \quad x \in \mathbb{R}, \\
& \quad f(t,U) \geq t(T-t)(q(t,U),Q_2q(t,U)), \quad t \in \mathbb{R}, \\
& \quad \forall f(t,x) \leq -(U - x)xt(t-t)(q(t,x),Q_3q(t,x)), \quad (t,x) \in \mathbb{R}^2, \\
& \quad f \in C_p(\mathbb{R}^2;\mathbb{R}), \\
& \quad \{Q_k\}_{k=1,...,A} \text{ are positive semidefinite matrices,}
\end{align*}
$$

where decision variables are not only the coefficients of the polynomial $f$ but also the entries of $Q_k$’s. This formulation serves as a relaxation of the original problem (4.4) since the equalities in (4.5) imply the inequalities in (4.4), due to the positive semidefiniteness of $Q_k$’s. The optimization problem (4.5) can be rewritten as an optimization problem under a set of linear equalities and semidefiniteness constraints. This problem conversion is automatically executed by SOSTOOLS [13] or YALMIP [13]. The resulting problem is solved with SeDuMi [18], a well-established semidefinite programming solver.

First, in the spirit of Theorem 3.1 we solve the above optimization problems (4.4) only once, and check how well their optimal bounding functions $f^*$ and $g^*$ can provide the information about the intermediate times. In the left of Figure 3 we draw the obtained optimal bounding functions $f^*$ and $g^*$ when we set $X_0 = 0.7$. Observe that the upper bounding function $f^*(s,X_0)$ is completely useless for any $s$ such that $f^*(s,X_0) \geq 1$. Second, in accordance with Corollary 3.7, we solve the optimization problems

$$
\begin{align*}
\min & \quad f(0,X_0) + f(0.6,X_0) \\
\text{s.t.} & \quad f(T,x) \geq 1 \text{ on } [0,U], \\
& \quad f(t,U) \geq 0 \text{ on } E_U, \quad \text{and} \\
& \quad \forall f(t,x) \leq 0 \text{ on } E_1, \\
& \quad f \in C_p(E_1;\mathbb{R}), \\
\max & \quad g(0,X_0) + g(0.6,X_0) \\
\text{s.t.} & \quad g(T,x) \leq 1 \text{ on } [0,U], \\
& \quad g(t,U) \leq 0 \text{ on } E_U, \quad \text{and} \\
& \quad \forall g(t,x) \geq 0 \text{ on } E_1, \\
& \quad g \in C_p(E_1;\mathbb{R}),
\end{align*}
$$

instead of the original optimization problems (4.4). By this, we check how much the optimization problems (4.6) may tighten the gap where the time to the terminal is 0.4$(= T - 0.6)$,

$$
f^*(0.6,X_0) - g^*(0.6,X_0) < f^*(0.6,X_0) - g^*(0.6,X_0),
$$

by giving up some portion of the tight gap

$$
f^*(0,X_0) - g^*(0,X_0) > f^*(0,X_0) - g^*(0,X_0).
$$

We draw, in the right of Figure 3 the obtained optimal bounding functions $f^*$ and $g^*$ through the optimization problems (4.6) again with $X_0 = 0.7$ for comparison. Observe that the optimality gap at $s = 0.6$ is well tightened, while that at $s = 0$ is slightly widened.

Figure 3: Optimal bounding functions $f^*(s,X_0)$ and $g^*(s,X_0)$ (left) and $f^*(s,X_0)$ and $g^*(s,X_0)$ (right) in Corollary 3.7 with $\phi_1(dt,dx) = \delta_{(0,0.6)}(dt,dx)$ and $\phi_2(dt,dx) = \delta_{(0.7)}(dt,dx)$. 

10
Next, in accordance with Corollary 3.8, we solve the optimization problems

\[
\begin{align*}
\min_{f} & \quad f(0,x_1) + f(0,x_2) \\
\text{s.t.} & \quad f(T,x) \geq 1 \quad \text{on } [0,U], \\
& \quad f(t,x) \geq 0 \quad \text{on } E_U, \quad \text{and} \\
& \quad \mathcal{A} f(t,x) \leq 0 \quad \text{on } E_1, \\
& \quad f \in C_p(E_1; \mathbb{R}),
\end{align*}
\]

\[
\begin{align*}
\max_{g} & \quad g(0,x_1) + g(0,x_2) \\
\text{s.t.} & \quad g(T,x) \leq 1 \quad \text{on } [0,U], \\
& \quad g(t,x) \leq 0 \quad \text{on } E_U, \quad \text{and} \\
& \quad \mathcal{A} g(t,x) \geq 0 \quad \text{on } E_1, \\
& \quad g \in C_p(E_1; \mathbb{R}),
\end{align*}
\]

instead of the original optimization problems (4.4) with \( X_0 = x_1 \). By this, we check how much the optimization problems (4.7) may tighten the gap for \( x_2 = 0.1 \),

\[ f^*(0,x_2) - g^*(0,x_2) \leq f^*(0,x_2) - g^*(0,x_2), \]

by giving up some portion of the tight gap

\[ f^*(0,x_1) - g^*(0,x_1) > f^*(0,x_1) - g^*(0,x_1). \]

We draw, in the left of Figure 4, the optimal bounding functions \( f^*(0,x) \) and \( g^*(0,x) \) through the optimization problems (4.4) with \( X_0 = x_1 = 1.0 \), while \( f^*(0,x) \) and \( g^*(0,x) \) through (4.7) with \( (x_1,x_2) = (1.0,0.1) \) in the right figure. We can observe that the optimality gap at \( x = 0.1 \) is well tightened, while that at \( x = 1.0 \) is slightly widened.

![Figure 4: Optimal bounding functions](image)

It is straightforward to see that the optimization problems (4.7) cover the case of random initial state whose distribution is (discrete) uniform defined on \( \{x_1,x_2\} \) in Proposition 3.9 with \( \varphi_1(dt, dx) = \delta_{[0] \times [1.0]}(dt, dx) \) and \( \varphi_2(dt, dx) = \delta_{[0]}(dt) \). We can reuse the result just as

\[
0.0921 = \frac{1}{2} \sum_{k=1}^{2} g^*(0,x_k) \leq \mathbb{P}((\tau,X_\tau) \in E) \leq \frac{1}{2} \sum_{k=1}^{2} f^*(0,x_k) = 0.1664,
\]

where the probability measure here considers \( X_0 \) as a random variable with the aforementioned distribution.

Any large sample size in Monte Carlo simulations for square-root diffusions can never be in competition with our results since the upper and lower bounds obtained through our method form nothing but the 100%-confidence interval. By taking into account the computing time required for the Monte Carlo simulations, the superiority of our optimization approach is evident. Indeed, each computation in this subsection requires only a few seconds.

### 4.2 Stochastic Differential Equations with Jumps: Doléans-Dade stochastic exponential driven by gamma process

Next, we demonstrate that our method is applicable to a nonlinear moment estimation and stochastic differential equations with jumps. Set \( X_0 > 0, \ a_0(t,x) = a_1(t,x) = 0, \ b_1(t,x) = x, \ b_2(z) = z, \) and

\[ \nu(dz) = a e^{-bz} \frac{dz}{z}, \quad z > 0, \]

for \( a > 0 \) and \( b > 0 \), that is a gamma Lévy measure. In this setting, the stochastic differential equation (2.3) reduces to the so-called Doléans-Dade stochastic exponential

\[ dX_t = X_t \int_{\mathbb{R}_+} z(\mu - \nu)(dz,dt), \quad X_0 > 0. \]
It is clear that $\mathbb{E}[X_T] = X_0$. Moreover, we have $\mathbb{E}[X_2^2] = X_0^2 e^{2\sigma^2 t}$, since by the Ito-Wiener isometry,

$$\mathbb{E} [X_2^2] = X_0^2 + \int_{\mathbb{R}_+} z^2 \nu(dz) \mathbb{E} \left[ \int_0^T X_2^2 dt \right] = X_0^2 + \frac{b}{\mathcal{B}^2} \int_0^T \mathbb{E} [X_2^2] dt,$$

where the interchange of the integrals holds by the Fubini theorem with the almost sure non-negativity of $X_2^2$. (For more details, we refer to Applebaum [2].)

On one hand, Euler-Maruyama schemes do not guarantee the non-negativity of sample paths. To illustrate this, let $N \in \mathbb{N}$ and $\Delta := T/N$, and consider the equidistant time discretization approximation of $\{X_t : t \in [0,T]\}$, that is,

$$X_{k\Delta} \frac{X_{(k-1)\Delta}}{X_{(k-1)\Delta}} = 1 + \gamma_k (\Delta, b) - \Delta/b,$$

where $\{\gamma_k(a,b)\}_{k \in \mathbb{N}}$ is a sequence of iid gamma random variables with the common probability density function $b^a/\Gamma(a)e^{-by}dy$ on $\mathbb{R}_+$. With a choice of $(a,b,\Delta)$ satisfying $1 - \Delta/b < 0$, discretized sample paths may drop below zero. Such numerical experiments produce a misleading result $\mathbb{E}[X_T] < X_0$. On the other hand, based upon the canonical form

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

with an infinite shot noise series representation due to Bondesson [4], sample paths over a finite horizon can be simulated as

$$X_t = X_0 \exp \left[-t a + \sum_{k=1}^{+\infty} \ln \left(1 + e^{-\frac{\Gamma_k}{b}} \right) \mathbb{I} (T_k \leq t) \right], \quad t \in [0,T],$$

where $\{\Gamma_k\}_{k \in \mathbb{N}}$ are arrival times of a standard Poisson process, where $\{V_k\}_{k \in \mathbb{N}}$ is a sequence of iid standard exponential random variables, and where $\{T_k\}_{k \in \mathbb{N}}$ is a sequence of iid uniform random variables on $[0,T]$. It is clearly not sensible to generate this infinite series for each sample path, even with some finite truncation of the series. (See, for example, Imai and Kawai [6] for numerical issues around the infinite shot noise series representation.)

With the notations [2,1] and [2,2] and $\mathbb{R}_+$, $z^k \nu(dz) = a(k-1)!/b^k$ for $k = 2, 3, \ldots$, in mind, it holds that for $f \in C_{\mathcal{P}}([0,T] \times \mathbb{R}_+; \mathbb{R})$,

$$\mathcal{A} f(t,x) = \sum_{k=1}^{\infty} c_{k,k,k-1} x^{k-1} x^k + \sum_{k=0}^{\infty} c_{k,k,k} x^k \sum_{k=0}^{\infty} k_i C_k a(k_i - k - 1)!.$$ 

We test our optimization approach on moment estimation of a nonlinear unbounded function. Here, we consider the second moment $\mathbb{E}[X_2^2] = X_0^2 e^{2\sigma^2 t}$ derived above and fix the model parameters $a = 0.1$, $b = 1.5$ and $X_0 = 1.0$. Two optimal bounding functions $f^*$ and $g^*$ are obtained for the terminal time point $T = 1.0$ with the polynomial degree $K = 6$. In Figure 5 two bounding functions $f^*(s,X_0)$, $g^*(s,X_0)$ and the true value $X_0^2 e^{2\sigma^2 (T-t)}$ (backward in time) are indistinguishably close to each other. This indicates that those bounding functions provide a very good approximation of the second moment uniformly over the time interval $[0,T]$ without additional computation.

### 4.3 Barrier Basket Option under the Multi-Asset Black-Scholes Model

Let us close this study by demonstrating the applicability of our method in the multivariate setting. To this end, consider the two-asset Black-Scholes model, that is,

$$dX_t = X_t \left[ r dt + \sigma_1 dW^{(1)}_t \right],$$

$$dY_t = Y_t \left[ r dt + \sigma_2 \left( \rho dW^{(1)}_t + \sqrt{1 - \rho^2} dW^{(2)}_t \right) \right],$$

where $r$ is the (non-negative) risk-free rate, $\sigma_1$ and $\sigma_2$ are positive constants, $\rho \in (-1, +1)$, and $\{W^{(1)}_t : t \geq 0\}$ and $\{W^{(2)}_t : t \geq 0\}$ are independent standard Brownian motions. We simply fix the initial states $X_0$ and $Y_0$ positive constants here. (For simplicity of notation, we use the notation $X_t$ and $Y_t$ above, although $S_t$ is much more familiar in this context.) Then, it holds by the multivariate Ito formula that for $f \in C^{1,2,\gamma}\{[0,T] \times \mathbb{R} \times \mathbb{R}; \mathbb{R}\}$,

$$df(t,x,y) = \partial_1 f(t,x,y) dt + \partial_2 f(t,x,y) dX_t + \frac{1}{2} \partial_3^2 f(t,x,y) d[X,X]_t$$

$$+ \partial_3 f(t,x,y) dY_t + \frac{1}{2} \partial_3^2 f(t,x,y) d[Y,Y]_t + \partial_3 \partial_3 f(t,x,y) d[X,Y]_t$$

$$= \mathcal{A} f(t,x,y) dt + [\sigma_1 X \partial_2 f(t,x,y) + \rho \sigma_2 Y \partial_3 f(t,x,y)] dW^{(1)}_t + \sqrt{1 - \rho^2} \sigma_2 Y \partial_3 f(t,x,y) dW^{(2)}_t,$$
We are interested in evaluating the lower exit boundary \( e \). Here, we omit the (deterministic) discount factor (again, the first exit time out of \( E \) is \( \tau \)). Hence, by denoting by \( U \) suitable constants which is very similar to the one-dimensional counterpart (2.6).

Hence, by denoting by \((\mathcal{F}_t)_{t \geq 0}\) the filtration generated by \(\{X_t : t \geq 0\}\) and \(\{Y_t : t \geq 0\}\) and by \(\tau\) a suitable \((\mathcal{F}_t)_{t \geq 0}\)-measurable stopping time, the Dynkin formula is given by

\[
\mathbb{E}[f(0,X_\tau,Y_\tau)] = \mathbb{E}[f(0,X_0,Y_0)] + \mathbb{E}\left[ \int_0^\tau \mathcal{A} f(t,X_t,Y_t) \, dt \right],
\]

which is very similar to the one-dimensional counterpart (2.6).

Here, we are concerned with a up-and-out barrier option on a linear combination of the two dynamics \(\{c_sX_t + c_yY_t : t \geq 0\}\), with suitable constants \(c_s\) and \(c_y\). Define the bounded set

\[
E_1 := [0,T) \times \{(x,y) \in \mathbb{R}_+^2 : c_sx + c_yy \in [0,U]\},
\]

with \(U > 0\), and set the random variable \(\tau\) to be the \((\mathcal{F}_t)_{t \in [0,T]}\)-stopping time defined by

\[
\tau := \inf\{t \geq 0 : (t,x,y) \notin E_1\} \wedge T,
\]

again, the first exit time out of \(E_1\) on the finite time horizon \([0,T]\). The barrier is set to be constant in time. The random vector \((\tau,X_\tau,Y_\tau)\) indicates the exit location, whose support \(E_0\) can be split into two disjoint sets

\[
E_U := [0,T) \times \{(x,y) \in \mathbb{R}_+^2 : c_sx + c_yy = U\},
\]

\[
E_r := \{T\} \times \{(x,y) \in \mathbb{R}_+^2 : c_sx + c_yy \in [0,U]\}.
\]

The lower exit boundary \(E_D\) can be omitted in this context, as the sample path is known to be almost surely positive componentwise. We are interested in evaluating

\[
\mathbb{E}\left[ (c_sX_T + c_yY_T - K)^+ \mathbb{1} (\tau = T) \right],
\]

with some strike price \(K \in (0,U)\), that is,

\[
V(t,x,y) = (c_sx + c_yy - K)^+ \mathbb{1} ((t,x,y) \in E_r), \quad (t,x,y) \in [0,T] \times (0,\infty) \times (0,\infty).
\]

Here, we omit the (deterministic) discount factor \(e^{-rT}\) as it is irrelevant to this optimization procedure. With the notation

\[
E_{0,K} := \{(x,y) \in \mathbb{R}_+^2 : c_sx + c_yy \in [0,K]\}, \quad E_{K,U} := \{(x,y) \in \mathbb{R}_+^2 : c_sx + c_yy \in (K,U)\},
\]

Figure 5: Optimal bounding functions \(f^*(s,X_0)\) and \(g^*(s,X_0)\) and the true value \(X_0^2 e^{\frac{\sigma^2}{2}(T-s)}\) (backward in time). Three functions are visually indistinguishable.
we can formulate optimization problems

\[
\begin{align*}
\min & \quad f(0, X_0, Y_0) \\
\text{s.t.} & \quad f(t, x, y) \geq 0 \text{ on } E_U, \\
& \quad f(T, x, y) \geq c_x x + c_y y - K \text{ on } E_{K, U}, \quad \text{and} \\
& \quad \nabla f(t, x, y) \leq 0 \text{ on } E_1, \\
& \quad f \in C_p(E_1; \mathbb{R}),
\end{align*}
\]

\[
\begin{align*}
\max & \quad g(0, X_0, Y_0) \\
\text{s.t.} & \quad g(t, x, y) \leq 0 \text{ on } E_U, \\
& \quad g(T, x, y) \leq c_x x + c_y y - K \text{ on } E_{K, U}, \\
& \quad \nabla g(t, x, y) \geq 0 \text{ on } E_1, \\
& \quad g \in C_p(E_1; \mathbb{R}).
\end{align*}
\]

(4.10)

It is noteworthy that in this problem setup, the trajectory \{c_x X_t + c_y Y_t : t \in [0, T]\} will be monitored continuously, while it is difficult for our optimization approach to deal with discretely-monitored barrier options. Barrier options have often been meant for discrete monitoring ironically, as the continuous monitoring is physically impossible in practice.

We fix \(X_0 = Y_0 = 1.0, r = 0.05, \sigma_1 = \sigma_2 = 0.2, \rho = 0.3, c_x = c_y = 0.5\) (a simple average of two assets), \(U = 1.5, K = 0.6\) and \(T = 0.5\). Note that this setting can be transformed into \((X_0, Y_0, U, K) = (100, 100, 150, 60)\) for example, by multiplying the solutions by 100 afterwards. In the middle of Figure 6 we plot the discounted optimal bounding functions, where \(f_1\) and \(g_1\) are the polynomial functions obtained through the optimization problems (4.10) with a suitable relaxation like (4.5). The bounds at time \(s\) should be interpreted as the bounds of the same barrier basket option price, but with a shorter maturity \(T - s\). The circle at time \(T\) indicates \(c_x X_0 + c_y Y_0 - K\), that is, the option price which instantaneously terminates. As can be seen, the upper and lower bounding functions are sufficiently close to each other all over the time region \([0, T]\). With this small optimality gap, those bounds can act as reliable estimates for the option price with every length of the maturity from 0 to \(T\). Let us emphasize that we have implemented the set of the optimization problems (4.10) only once (more precisely, once for \(f_1\) and once for \(g_1\)) to provide the bounds in time. Although Monte Carlo simulation of the two-asset Black-Scholes model is rather simple, it is not straightforward to give good estimates of the option price with every shorter maturity under continuous monitoring.

![Figure 6: Bounding functions \((f_1, g_1)\) for the up-and-out barrier basket option price under the multi-asset Black-Scholes model through optimization (4.10).](image)

Next, along Corollary 3.8 we check the applicability of the optimal bounding functions \(f_1\) and \(g_1\) at different initial states. We draw in Figure 7 the discounted bounding functions evaluated at \((0.8X_0, 0.8Y_0)\) in the leftmost, and the discounted bounding functions evaluated at \((1.2X_0, 1.2Y_0)\) in the rightmost. Note that the functions \(f_1\) and \(g_1\) are the polynomial functions obtained through (4.10) used in the middle of Figure 6. In the other words, no additional optimization problems are implemented to draw those two figures. The quality, in terms of optimality gap, gets worse for the different initial states. In particular, optimality gap for the initial state \((1.2X_0, 1.2Y_0)\) with a longer maturity is large very badly. This quality loss is an inevitable phenomenon and rather obvious, as the optimal bounding functions \(f_1\) and \(g_1\) are chosen to be the best when the initial state is \((X_0, Y_0)\). Still, depending on the required extent of accuracy, the original results \(f_1\) and \(g_1\) for the initial state \((X_0, Y_0)\) would provide useful estimates for the option pricing with the same structure with a relatively short maturity, without additional implementation of the optimization problems or Monte Carlo simulation.

In general, with greater sets \(E_0\) and \(E_1\) (which is caused mainly by a longer maturity \(T\) in this case), the optimization problems tend to provide wider optimality gaps since then the bounding polynomials are required to satisfy the constraints over a greater domain. Obviously, this drawback turns out to be more remarkable in the multivariate setting. It is difficult to predict quality of optimal solutions a priori as several complex factors are involved. For example, it is partially a matter of the fitness of bounding polynomials to the problem structure, such as the value function, the infinitesimal generator, the domain, etc, as the polynomial degree \(K\) in the definition (2.2) need to be finite (and relatively small) due to a limitation of computing resource.
5 Concluding Remarks

In this paper, we have shown that feasible solutions obtained through the known optimization approach gives additional information for arbitrary intermediate times and/or for different initial states. We have proved that by using this property, the optimality gap for an arbitrary intermediate time and/or for a different initial state may also be tightened, in return for widening the optimality gap at the terminal time and an originally set initial state. We have further shown that this optimization approach can also be applied to stochastic differential equations with random initial state. It is remarkable that randomness of initial state is confined into deterministic optimization problems. We have provided numerical results to illustrate the effectiveness of our theoretical results.

We have not yet fully explored duality results of the obtained ones. That is, it is not trivial to formulate these problems as 
\textit{generalized moment problems}. It would be an interesting future work to investigate this duality, and also to extend the obtained theories to wider class of problems such as two-point boundary value problems.

References


