SIAM J. CONTROL OPTIM.
Vol. 51, No. 5, pp. 4189–4210




# WEAK CONVERGENCE METHODS FOR APPROXIMATION OF THE EVALUATION OF PATH-DEPENDENT FUNCTIONALS*

QINGSHUO SONG†, GEORGE YIN‡, AND QING ZHANG§

**Abstract.** In many applications, one needs to evaluate a path-dependent objective functional $V$ associated with a continuous-time stochastic process $X$. Due to the nonlinearity and possible lack of Markovian property, more often than not, $V$ cannot be evaluated analytically, and only Monte Carlo simulation or numerical approximation is possible. In addition, such calculations often require the handling of stopping times, the usual dynamic programming approach may fall apart, and the continuity of the functional becomes the main issue. Denoting by $h$ the stepsize of the approximation sequence, this work develops a numerical scheme so that an approximating sequence of path-dependent functionals $V^h$ converges to $V$. By a natural division of labors, the main task is divided into two parts. Given a sequence $X^h$ that converges weakly to $X$, the first part provides sufficient conditions for the convergence of the sequence of path-dependent functionals $V^h$ to $V$. The second part constructs a sequence of approximations $X^h$ using Markov chain approximation methods and demonstrates the weak convergence of $X^h$ to $X$, when $X$ is the solution of a stochastic differential equation. As a demonstration, combining the results of the two parts above, approximation of option pricing for the discrete-monitoring-barrier option underlying stochastic volatility model is provided. Different from the existing literature, the weak convergence analysis is carried out by using the Skorohod topology together with the continuous mapping theorem. The advantage of this approach is that the functional under study may be a function of stopping times, projection of the underlying diffusion on a sequence of random times, and/or maximum/minimum of the underlying diffusion.

**Key words.** path-dependent functional, weak convergence, Monte Carlo optimization, Skorohod topology, continuous mapping theorem

**AMS subject classifications.** 93E03, 93E20, 60F05, 60J60, 60J05

**DOI.** 10.1137/130913158

## 1. Introduction and examples.

In many applications, one needs to evaluate path-dependent objective functionals. They arise, for example, in derivative pricing, networked system analysis, and Euler's approximation to solution of SDEs. This paper is concerned with approximation methods for computing such objective functions. At first glance, the problem may appear as a standard approximation of a stopping time problem involving traditional techniques. Nevertheless, a closer scrutiny reveals that the problem under consideration is far more challenging and difficult. The difficulties are in the following three aspects:

(i) The path dependence leads to fundamental difficulty in the evaluation of the underlying functionals with stopping times.

(ii) The traditional dynamic programming approach falls apart, not to mention any hope for a closed-form solution or any viable numerical solutions for the associated partial differential equations (PDEs). Thus one naturally turns to

*Received by the editors March 15, 2013; accepted for publication (in revised form) September 16, 2013; published electronically October 30, 2013. This research was supported in part by the Research Grants Council of Hong Kong, CityU 109613, and in part by the Army Research Office under grant W911NF-12-1-0223.

http://www.siam.org/journals/sicon/51-5/91315.html

†Department of Mathematics, City University of Hong Kong, Kowloon Tong, Hong Kong (song.qingshuo@cityu.edu.hk).

‡Department of Mathematics, Wayne State University, Detroit, MI 48202 (gyin@math.wayne.edu).

§Department of Mathematics, University of Georgia, Athens, GA 30602 (qingz@math.uga.edu).

approximation methods using Monte Carlo optimization. To the best of our knowledge, convergence analysis is available only in a few special cases due to the complexity of the nature of path dependence.

(iii) In evaluating the path-dependent functionals, operations involving max and min are used. As a result, continuity becomes a major issue; more illustrations and counterexamples will be provided in section 1.4. The standard weak convergence argument does not apply.

To begin, there are virtually no closed-form solutions involving path-dependent objective functionals. One has to look for alternatives. The next possibility is numerical approximation using PDE-based techniques. Nevertheless, such possibility is ruled out due to the lack of the Markovian properties. The only often used technique left is the Monte Carlo method. While most of the existing methods for treating Monte Carlo optimization are somewhat ad hoc, this work develops a systematic alternative method for analyzing the convergence of the approximation algorithm. In this paper, we use a weak convergence approach and carry out the convergence analysis under the Skorohod topology. One of the main ingredients is the use of a generalized projection operator. With the use of such projections, we proceed to explore the intrinsic properties of the Skorohod space. The convergence analysis developed in this paper provides a thorough understanding of the nature of path dependence and a general framework for handling many path-dependent problems.

**1.1. Path-dependent objective functions.** We are interested in approximating path-dependent functions in a finite time horizon. For simplicity, we focus our attention on the time interval $[0,1]$ to avoid using more complex notation. Let $C[0,1]$ be the collection of continuous real-valued functions defined on $[0,1]$, and let $D[0,1]$ be the collection of all right continuous with left-hand limits (RCLL) functions on $[0,1]$. With $\mathbb{F} = \{\mathcal{F}_t : t \in [0,1]\}$ satisfying the usual conditions and $\mathcal{F} = \mathcal{F}_1$, let $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F})$ be a filtered probability space; see [10]. As a convention, we use capital letters to denote random elements and use lowercase letters for deterministic (or nonrandom) elements. For instance, $x \in D[0,1]$ is an RCLL function defined on $[0,1]$, while $X : \Omega \mapsto D[0,1]$ is a random RCLL process with sample paths in $D[0,1]$. As usual, $X$ can be regarded as a function of two variables (time $t$ and sample point $\omega$). That is, for each fixed $\omega \in \Omega$, $X(\cdot,\omega)$ denotes a sample path and for each fixed $t$, $X(t,\cdot)$ is a random element in $\mathbb{R}$.

Throughout the paper, $\alpha, \beta \in C[0,1]$ are given satisfying $\alpha(t) < \beta(t)$ for all $t \in [0,1]$. The domain of our interest is defined by all the points bounded by $\alpha$ and $\beta$, that is,

$$\Gamma = \{(a,t) : \alpha(t) < a < \beta(t), t \in [0,1]\}.$$

The cross section $\Gamma(t) = (\alpha(t), \beta(t)) \subset \mathbb{R}$ may be time-dependent. We also denote its boundary by

$$\partial\Gamma = \{(\alpha(t),t) : t \in [0,1]\} \cup \{(\beta(t),t) : t \in [0,1]\}.$$

Consider an $\mathbb{F}$-adapted continuous process $X : [0,1] \times \Omega \mapsto C[0,1]$ with initial $X(0) \in (\alpha(0), \beta(0))$. Let $\tau$ be the first hitting time to the boundary of the time-dependence domain

$$\tau = \inf\{t > 0 : X(t) \notin \Gamma(t)\} \wedge 1. \tag{1.1}$$

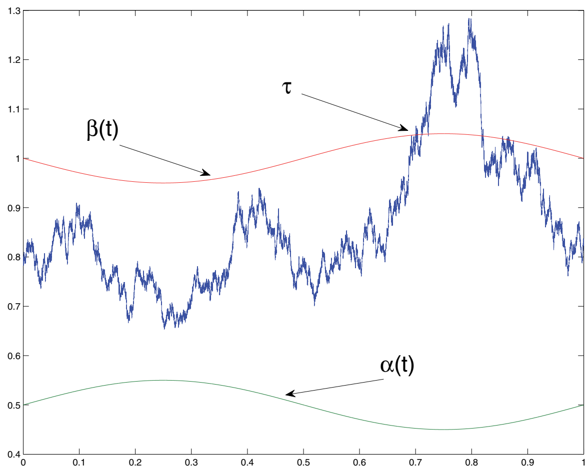


FIG. 1.1. *A demonstration of a sample path $X$ and domain $\Gamma$ bounded by lower barrier $\alpha$ and upper barrier $\beta$.*

See Figure 1.1 for its illustration of the above problem setup. Now we introduce the notion of projection operator (see [3]) $\Pi : D[0,1] \times [0,1]^m \mapsto \mathbb{R}^m$ as

$$\Pi(x,\nu) = (x(\nu_1),\ldots,x(\nu_m))' \quad \forall x \in D[0,1], \nu \in [0,1]^m. \tag{1.2}$$

For $x \in D[0,1]$, we define a maximum process $x^*$ by $x^*(t) = \sup_{0\le s\le t} x(t)$. It is easy to check that $x^* \in D[0,1]$. For some measurable function $g : \mathbb{R}^{4m+1} \to \mathbb{R}$, we are interested in the computation of the objective functional $V : C[0,1] \mapsto \mathbb{R}$ defined by, for given $\{\nu^{(i)} \in [0,1]^m : i = 1,2,3,4\}$,

$$V = \mathbb{E}[g(\Pi(X,\tau\nu^{(1)}),\Pi(X,\nu^{(2)}),\Pi(X^*,\tau\nu^{(3)}),\Pi(X^*,\nu^{(4)}),\tau)]. \tag{1.3}$$

Note that random times $\tau_j = \tau\nu_j^{(1)}$ are only measurable with respect to $\mathcal{F}_1$ but may not be stopping times whenever $\nu_j^{(1)} < 1$. Moreover, the value function may be a measurable function depending on the initial states $V(x,t)$ if the functional $V$ is given as a non-path-dependent form of $\mathbb{E}[f(X(T))]$ for some function $f : \mathbb{R} \mapsto \mathbb{R}$. However, we emphasize that our main interest is the computation when the functional is path-dependent in the form of (1.3), which has a wide range of applications in mathematical finance.

*Remark* 1. Related literature in connection with a hitting time under a Markovian framework can be found in Dufour and Piunovskiy [6], de Saporta, Dufour, and Gonzalez [5], and Szpruch and Higham [20], among others. In particular, in [6], the existence of optimal stopping with constraints is established using a convex analytic approach. A numerical method for optimal stopping of piecewise deterministic Markov processes is considered in [5]. Bounds for the convergence rate of their algorithms are obtained by introducing quantization of the post jump location and path-adapted time discretization grids. In [20], an application in physics (thermodynamic limit) involving mean hitting time behavior is considered.

**1.2. Examples.** The evaluation of path-dependent objective functions is of great interest in many networked systems as well as in financial applications such as option pricing. In particular, the value $V$ of (1.3) can be considered as a general form of a class of option prices including the look-back option, rebate/barrier option, Asian option, and Bermuda option. To illustrate, consider the underlying stock price $X$ that is a nonnegative continuous martingale process under $\mathbb{P}$ that is the corresponding *equivalent local martingale measure*. Throughout this subsection, we also assume that the interest rate is $r \geq 0$, and $\nu^{(i)} = \frac{1}{m}(1, 2, \ldots, m)'$, $-\alpha(t) = \infty$, and $\beta(t) = 1$. Thus, the first hitting time can be written

$$\tau = \inf\{s > 0 : X(s) \notin (-\infty, 1)\} \wedge 1.$$

We also assume the distribution of $X$ under $\mathbb{P}$ has no atom. In particular, $\mathbb{P}\{X(t) = c\} = 0$ for all $c > 0$ and $t > 0$.

*Example* 2 (barrier option). A barrier option is an option with a payoff depending on whether, within the life of the option, the price of the underlying asset reaches a specified level (the so-called barrier); see more details in [16]. There are three basic types of barrier options: knock-out options, knock-in options, and rebate options. As an illustration, we consider an up-and-in barrier call with strike 1/2 at maturity $T = 1$, which is one special case of knock-in options. Roughly speaking, the up-and-in call activates (knocks-in) its payoff $(X(1)-1/2)^+$ if the upper barrier is touched before the expiration $T = 1$. The precise price formula of the up-and-in call is given by

$$V = e^{-r}\mathbb{E}\left[\left(X(1) - \frac{1}{2}\right)^+ I_{[0,1)}(\tau)\right] = e^{-r}\mathbb{E}\left[\left(X(1) - \frac{1}{2}\right)^+ I_{[1,\infty)}(X^*(1))\right].$$

In this case, the payoff function is

$$g(x_1, x_2, \ldots, x_{4m+1}) = e^{-r}\left(x_{2m} - \frac{1}{2}\right)^+ I_{[1,\infty)}(x_{4m}),$$

which is discontinuous at $x_{4m} = 1$.

*Example* 3 (discrete-monitoring-barrier option). In the related literature, most works assume continuous monitoring of the barrier like Example 2. However, in practice most barrier options traded in markets are monitored at discrete time. Unlike their continuous-time counterparts, there is essentially no closed-form solution available, and even numerical pricing is difficult; see more related discussions in [4] and [12]. For instance, the price formula of up-and-in barrier call monitoring at discrete time instants $\{1/m, 2/m, \ldots, m-1/m, 1\}$ with strike 1/2 is

$$V = e^{-r}\mathbb{E}\left[\left(X(1) - \frac{1}{2}\right)^+ I_{[1,\infty)}\Big(\max_{1\leq i\leq m} X(i/m)\Big)\right]. \tag{1.4}$$

The payoff function $g$ is written as

$$g(x_1, x_2, \ldots, x_{4m+1}) = e^{-r}\left(x_{2m} - \frac{1}{2}\right)^+ I_{[1,\infty)}\Big(\max_{m+1\leq i\leq 2m} x_i\Big).$$

Note that $g$ has linear growth, is unbounded, and is discontinuous.

**1.3. Computational methods.** To evaluate the path-dependent objective functions, a major difficulty arises from the lack of Markovian property. This rules out the possibility of the conventional numerical-PDE-based methods including the finite difference or finite element methods. In addition, the time-dependent barriers $\alpha$ and $\beta$ also create additional layers of difficulty. In this paper, we consider Monte Carlo methods to obtain a feasible estimate of the path-dependent problems. The Monte Carlo method is a class of computational algorithms that relies on some repeated random sampling to evaluate its deterministic value using its probabilistic fact. This includes Euler–Maruyama (EM) approximation [11] and Markov chain approximation [8], [13], and [14], among others.

The general idea of the Monte Carlo method in this vein is the following. For each sample point $\omega \in \Omega$, use $X^h(\cdot, \omega) \in D[0,1]$ to denote a simulated path for the underlying process $X(\cdot, \omega) \in C[0,1]$ by a certain Monte Carlo method with a small parameter $h$ (maybe a step size). Define the approximated stopping time by

$$\tau^h = \inf\{t > 0 : X^h(t) \notin (\alpha(t), \beta(t))\} \wedge 1.$$

One can approximate $V$ in this way by computing

$$V^h = \mathbb{E}[g(\Pi(X^h, \tau^h \nu^{(1)}), \Pi(X^h, \nu^{(2)}), \Pi(X^{h,*}, \tau^h \nu^{(3)}), \Pi(X^{h,*}, \nu^{(4)}), \tau^h)]. \tag{1.5}$$

The main task of this approximating scheme is to design an appropriate Monte Carlo method so that the desired convergence takes place eventually, i.e.,

$$\lim_{h \to 0} V^h = V.$$

As the objective value is given as an expectation of (1.3), $V$ is invariant under the same distribution, and the usual requirements for the Monte Carlo method are intuitively given as follows:

(H1) $X^h$ converges weakly to $X$ as $h \to 0$, denoted by $X^h \Rightarrow X$.
(H2) $g$ is continuous.

A couple of natural questions are as follows:

- Are (H1)–(H2) sufficient to guarantee the desired convergence $\lim_{h\to 0} V^h = V$?
- Can (H2) be possibly weakened to some discontinuous function $g$ so that the barrier option pricing (see section 1.2) can be included?

**1.4. Counterexamples.** Interestingly, there exist circumstances that lead to counterexamples in connection with the desired convergence under (H1)–(H2).

A noticeable counterexample is given in Figure 1.2, which is motivated by the so-called tangency problem. Consider the two underlying processes $X^1$ (solid line) and $X^2$ (dotted line) in Figure 1.2. No matter how close $X^1$ and $X^2$ are at the initial states, the difference between their first exit time $\tau_1$ and $\tau_2$ could be far; see [2]. The above idea is illustrated in the following example.

*Example* 4. Let $\{X(s) : 0 \le s \le 1\}$ be a deterministic process given by $X(s) = 1 - (s - \frac{1}{2})^2$. Let $\alpha(s) = -\infty$ and $\beta_s = 1$ for all $s \ge 0$. Then the exit time of $X$ is

$$\tau = \inf\{s > 0 : X(s) \notin (\alpha(s), \beta(s))\} \wedge 1 = \frac{1}{2}.$$

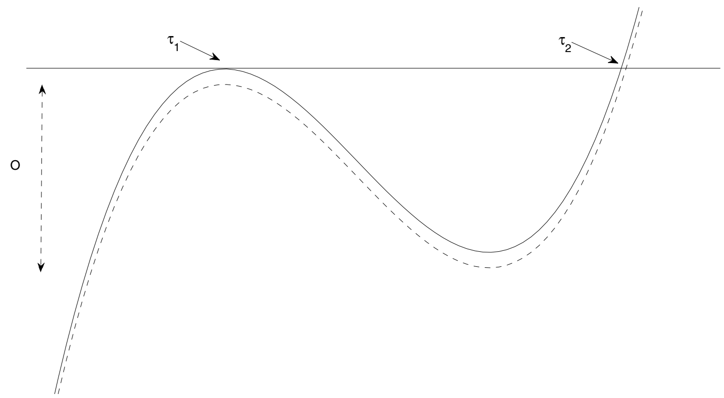


FIG. 1.2. *Demonstration of tangency problem: The sample paths of the two processes (solid line and dotted line) are fairly close to each other, but their first exit times are far apart, indicating a loss of continuity.*

Define a family of processes parameterized by $h$ with $X^h(s) = X(s) - h$. Although $X^h$ converges to $X$ in $L^\infty$ as $h \to 0^+$, we note that

$$\tau^h = \inf\{s > 0 : X^h(s) \notin (\alpha(s), \beta(s))\} \wedge 1 = 1$$

is not converging to $\tau = 1/2$.

The next example gives a very different aspect from the first one in the sense that there are no barriers and the underlying process is a Bessel process. This system is often used in finance to model the dynamics of asset prices, of the spot rate, and of the stochastic volatility, or as a computational tool. In particular, computations for the celebrated Cox–Ingersoll–Ross and constant elasticity variance models can be carried out using Bessel processes.

*Example* 5. Let $X(t)$ be a Bessel process of order 3 with initial $X(0) = 1$, i.e.,

$$X(t) = 1 + W(t) + \int_0^t \frac{ds}{X(s)}, \tag{1.6}$$

which is a well-known strict local martingale with $\mathbb{E}[X(1)] < 1$; see p. 336 of [9]. A sample path of $X(t)$ is given in Figure 1.3. Define $X^h(t) = X(t) \wedge \frac{1}{h}$. Fix a vector $\nu = (t_1 \le \cdots \le t_n)' \in [0,1]^n$; then we have by definition $\Pi(X^h, \nu) \to \Pi(X, \nu)$ a.s., and the weak convergence $\Pi(X^h, \nu) \Rightarrow \Pi(X, \nu)$ takes place. Moreover, the Kolmogorov consistency theorem together with the uniqueness of the weak solution of SDE (1.6) implies that $X^h \Rightarrow X$ due to the arbitrariness of $n$ and $\nu \in [0,1]^n$. However, since $X^h$ is a bounded local martingale, it is a martingale. Thus we have $\lim_{h\to 0} \mathbb{E}[X^h(1)] = 1 > \mathbb{E}[X(1)]$.

**1.5. Goal and outline of the paper.** The above examples show that (H1)–(H2) may not be sufficient for Monte Carlo simulation to be convergent to the right value of (1.3). This suggests the following question:

(Q1) Given $X^h \Rightarrow X$, what are sufficient conditions to ensure the convergence $\lim_{h\to 0} V^h = V$? Is it possible to weaken the continuity of $g$ to cover the barrier option?

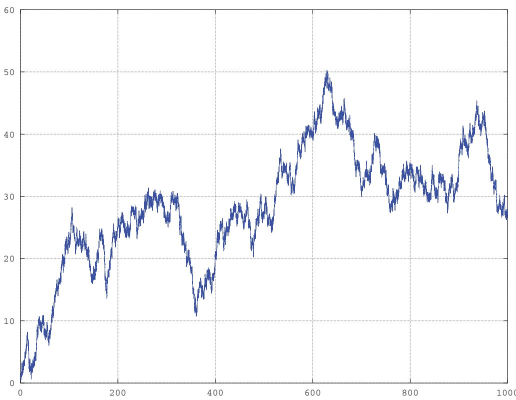


FIG. 1.3. *A sample path of Bessel process of order* 3.

In this paper, we first present a rigorous proof of the convergence in section 2. The results in Theorem 9 provide a set of sufficient conditions for the convergence. It can also explain why Examples 4 and 5 do not work.

Our approach is based on the actual computations using the Skorohod metric in the space $D[0,1]$. This is the first attempt in this context to the best of our knowledge. Such an approach is advantageous. For example, the study of Monte Carlo convergence usually varies with the particular form of the underlying payoff functions; see [8] and the references therein. However, Theorem 9 provides a unified theoretical basis for the convergence in a general path-dependent form with possibly discontinuous payoff function $g$ and finite time-dependent barrier $(\alpha(t), \beta(t))$. This covers a wider range of applications of Theorem 9 to various types of options, for instance, barrier options, Asian options, and look back options. It is also notable that the assumptions on the barrier in the path-dependent problem have some similarities to that imposed for stochastic controlled exit problems; see [2]. To complete the approximation of the value $V$ of (1.3), one should consider the following question in addition to (Q1):

(Q2) Given $X$, how does one construct an approximating sequence of processes $X^h$ such that $X^h \Rightarrow X$?

Of the many possible answers to (Q2), we will mainly focus on the construction of $X^h$ in the framework of Markov chain approximation in section 3. Recall that from the book [14], a family of continuous approximating processes $X^h$ constructed using Markov chain is convergent to $X$ in distribution if the Markov chain satisfies the local consistency [14, equation (9.4.2)]; see also [14, Theorem 10.4.1]. Compared to [14], we investigate the weak convergence for the more generalized local consistency condition. This enables us to cover various types of Monte Carlo simulations in the framework of Markov chain approximation to verify its weak convergence using the generalized local consistency. It essentially extends the use of Markov chain approximation.

To proceed, the rest of the paper is arranged as follows. Section 2 provides sufficient conditions for the convergence of the approximating path-dependent functional $V^h$ to the original functional $V$ given that there exists a sequence of random processes $X^h$ converging to the original process $X$ weakly. Section 3 addresses the question of how to find a sequence of weak convergent $X^h$ when the original process $X$ is the solution of an SDE. Since section 3 does not use any result obtained in section 2, these two sections are independently accessible to the reader. Finally, the paper is concluded with a demonstrating example on discrete-monitoring-barrier options together with some further remarks in section 4.

## 2. Sufficient conditions for convergence.

**2.1. Preliminaries.** We use the notation given in [3]. Define a metric on $D[0,1]$ by

$$\|x-y\| = \sup_{t\in[0,1]} |x(t)-y(t)| \ \forall x, y \in D[0,1].$$

Then, the continuous function space on $[0,1]$, denoted by $C[0,1]$, is complete with respect to the uniform topology with the above metric $\|\cdot\|$. On the other hand, the RCLL function space defined on $[0,1]$, denoted by $D[0,1]$, is equipped with the Skorohod topology with the metric

$$\|x-y\|_s = \inf_{\lambda\in\Lambda} \{\|\lambda - \mathbb{I}\|, \|x\circ\lambda - y\|\} \ \forall x, y \in D[0,1],$$

where $x \circ y$ denotes the composite function of $x$ and $y$, $\Lambda$ is the collection of all continuous increasing functions $\lambda$ on $[0,1]$ with $\lambda(0) = 0$ and $\lambda(1) = 1$, and $\mathbb{I} \in \Lambda$ is identity mapping. It is often useful to use the following fact: $x_n \to x$ in Skorohod topology if and only if there exists $\lambda_n \in \Lambda$ such that

$$\|\lambda_n - \mathbb{I}\| \to 0, \quad \|x_n - x\circ\lambda_n\| \to 0 \text{ as } n\to\infty. \tag{2.1}$$

Note that $D[0,1]$ is not complete under the metric $\|\cdot\|_s$, but it is complete under an equivalent metric $\|\cdot\|_s^o$ defined by

$$\|x-y\|_s^o = \inf_{\lambda\in\Lambda} \left\{ \sup_{s<t} \left| \log \frac{\lambda(t)-\lambda(s)}{t-s} \right|, \|x-y\| \right\} \ \forall x, y \in D[0,1].$$

For notational convenience, we will use $\|\cdot\|_s$ for the metric of $D[0,1]$ in the rest of the paper. In particular, for a mapping $F : D[0,1] \mapsto \mathcal{M}$ with metric space $(\mathcal{M}, \|\cdot\|_{\mathcal{M}})$, $F$ is said to be continuous at some $x \in D[0,1]$ if

$$\lim_{n\to\infty} \|F(x_n) - F(x)\|_{\mathcal{M}} = 0 \ \text{ whenever } \ \lim_{n\to\infty} \|x_n - x\|_s = 0.$$

In contrast, $x \in D[0,1]$ is said to be continuous at some $t \in (0,1)$ if

$$\lim_{n\to\infty} |x(t_n) - x(t)| = 0 \ \text{ whenever } \ \lim_{n\to\infty} |t_n - t| = 0.$$

**2.2. Some continuous mappings under Skorohod topology.** In this section, we will discuss continuity of some useful mappings on $D[0,1]$ with respect to Skorohod topology. Recall that for $x \in D[0,1]$, we define $x^* \in D[0,1]$ by $x^*(t) = \sup_{0\le s\le t} x(s)$. We also define a mapping $\mathcal{M} : D[0,1] \mapsto D[0,1]$ as

$$\mathcal{M}(x) = x^*.$$

In addition, we are interested in the projection $\Pi$ in (1.2) given by

$$\Pi(x,\nu) = (x(\nu_1), \ldots, x(\nu_m))' \quad \forall x \in D[0,1], \nu \in [0,1]^m$$

and a mapping $\pi : D[0,1] \mapsto [0,1]$ defined by

$$\pi(x) = \inf\{t : x(t) \notin (\alpha(t), \beta(t))\} \wedge 1. \tag{2.2}$$

LEMMA 6. *$\mathcal{M}$ is continuous on $D[0,1]$.*

*Proof.* Since $\lambda$ is strictly increasing, $x^* \circ \lambda = (x \circ \lambda)^*$. Therefore,

$$\begin{aligned}\|x^* - y^*\|_s &= \inf_{\lambda \in \Lambda} \left\{ \|\lambda - \mathbb{I}\|, \sup_{t \in [0,1]} |x^*(\lambda(t)) - y^*(t)| \right\} \\ &= \inf_{\lambda \in \Lambda} \left\{ \|\lambda - \mathbb{I}\|, \sup_{t \in [0,1]} |(x \circ \lambda)^*(t) - y^*(t)| \right\} \\ &\leq \inf_{\lambda \in \Lambda} \left\{ \|\lambda - \mathbb{I}\|, \sup_{t \in [0,1]} \sup_{0 \leq s \leq t} |(x \circ \lambda)(s) - y(s)| \right\} \\ &= \inf_{\lambda \in \Lambda} \{ \|\lambda - \mathbb{I}\|, \|x \circ \lambda - y\| \} = \|x - y\|_s.\end{aligned}$$

So, $\mathcal{M}$ is continuous in $D[0,1]$ with respect to the Skorohod topology. ∎

LEMMA 7. *$\Pi$ is continuous at $(x, \nu) \in D[0,1] \times [0,1]^m$ whenever $x$ is continuous at each $\nu_i$ of $i = 1, 2, \ldots, m$.*

*Proof.* Given a sequence of vectors $\nu^{(n)} \to \nu$, we observe that

$$|\Pi(x, \nu^{(n)}) - \Pi(x, \nu)| \leq \sum_{i=1}^m |x(\nu_i^{(n)}) - x(\nu_i)|,$$

and the continuity of $\Pi$ in the variable $\nu$ follows directly from the continuity of $x$ at $\nu_i$ for each $i$. Next, we show the continuity of $\Pi$ in the variable $x$. To proceed, we fix $\nu$ and an arbitrary sequence $\{x_n\}$ satisfying $\lim_n \|x_n - x\|_s = 0$. Then, there exists $\{\lambda_n\}$ satisfying (2.1). Therefore, we have

$$\begin{aligned}|\Pi(x_n, \nu) - \Pi(x, \nu)| &= \sum_{i=1}^m |x_n(\nu_i) - x(\nu_i)| \\ &\leq \sum_{i=1}^m (|x_n(\nu_i) - x(\lambda_n(\nu_i))| + |x(\lambda_n(\nu_i)) - x(\nu_i)|) \\ &\leq \|x_n - x \circ \lambda_n\| + \sum_{i=1}^m |x(\lambda_n(\nu_i)) - x(\nu_i)|.\end{aligned}$$

Note that the first term $\|x_n - x \circ \lambda_n\| \to 0$ as $n \to \infty$ by (2.1). On the other hand, since $\|\lambda_n - \mathbb{I}\| \to 0$ as $n \to \infty$ by (2.1), we have $|\lambda_n(\nu_i) - \nu_i| \to 0$ as $n \to \infty$. Hence, the second term $\sum_{i=1}^m |x(\lambda_n(\nu_i)) - x(\nu_i)| \to 0$ by continuity of $x$ at $\nu_i$. Thus, we conclude the continuity of $\Pi$ in the variable $x$. ∎

The continuity of $\pi$ of (2.2) is a rather tricky part. To proceed, let us partition the space $C[0,1]$ as follows: Define

$$C_1 = \{x \in C[0,1] : \pi(x) < 1,\ x(\pi(x)) = \beta(\pi(x)), \inf\{t > \pi(x) : x(t) > \beta(t)\} = \pi(x)\},$$
$$C_2 = \{x \in C[0,1] : \pi(x) < 1,\ x(\pi(x)) = \alpha(\pi(x)), \inf\{t > \pi(x) : x(t) < \alpha(t)\} = \pi(x)\},$$

and

$$C_3 = \{x \in C[0,1] : \pi(x) = 1\}, \text{ and } C_4 = C[0,1] \setminus (\cup_{i=1}^3 C_i).$$

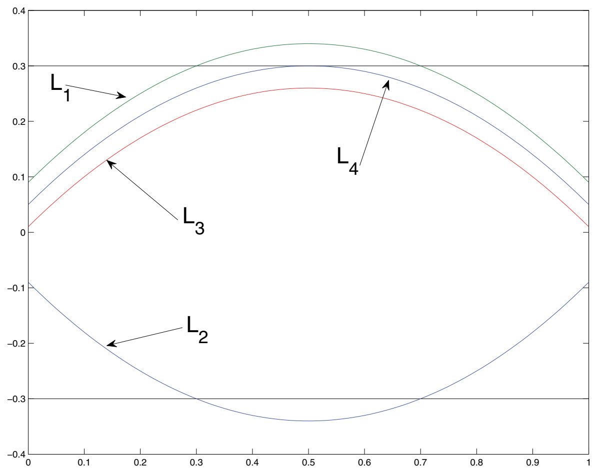


Fig. 2.1. *Illustration for the partition of* $C[0,1]$.

Accordingly, we can write $C[0,1] = \cup_{i=1}^4 C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$. As for the illustration, one can see that the four curves depicted in Figure 2.1 belong to four different subsets separately, that is, $L_i \in C_i$ for $i = 1, 2, 3, 4$.

Lemma 8. *The* $\pi$ *defined in* (2.2) *is continuous at each* $x \in \cup_{i=1}^3 C_i$.

*Proof.* First, $\pi$ is continuous at $x \in C_3$ thanks to Lemma 6. We assume $-\alpha(t) + \beta(t) < \infty$ for each $t$ without loss of generality. Fix $x \in C_1$, and let $\{x_n\}$ be an arbitrary sequence in $D[0,1]$ satisfying $\|x_n - x\|_s \to 0$ as $n \to \infty$. By the definition of $\|\cdot\|_s$, it implies that there exists $\lambda_n \in \Lambda$ such that

$$\|\lambda_n - \mathbb{I}\| \to 0, \quad \|x_n \circ \lambda_n - x\| \to 0 \text{ as } n \to \infty. \tag{2.3}$$

- Case 1. Define, for any $\varepsilon > 0$,

  $$\delta_\varepsilon := \frac{1}{2} \sup_{\pi(x) < t < \pi(x)+\varepsilon} (x(t) - \beta(t)).$$

  By the definition of $C_1$, we have $\delta_\varepsilon > 0$ for all $\varepsilon > 0$. Moreover, by continuity of $x$, there exists a time $t_\varepsilon \in (\pi(x), \pi(x) + \varepsilon)$ such that $x(t_\varepsilon) > \beta(t_\varepsilon) + \delta_\varepsilon$. Next, we observe that

  $$|x_n(\lambda_n(t_\varepsilon)) - x(t_\varepsilon)| \leq \|x_n \circ \lambda_n - x\| \to 0 \text{ as } n \to \infty.$$

  This implies that for some large $N_{\varepsilon,1}$

  $$x_n(\lambda_n(t_\varepsilon)) \geq x(t_\varepsilon) - \frac{\delta_\varepsilon}{2} > \beta(t_\varepsilon) + \frac{\delta_\varepsilon}{2} \; \forall n \geq N_{\varepsilon,1}. \tag{2.4}$$

  On the other hand, there exists $r_\varepsilon > 0$ due to the continuity of $\beta$ such that

  $$\sup_{t_\varepsilon - r_\varepsilon < t < t_\varepsilon + r_\varepsilon} |\beta(t) - \beta(t_\varepsilon)| < \frac{\delta_\varepsilon}{4}.$$

Moreover, since $\|\lambda_n - \mathbb{I}\| \to 0$, we have for some large enough $N_{\varepsilon,2}$,

$$\|\lambda_n - \mathbb{I}\| < r_\varepsilon \ \forall n \geq N_{\varepsilon,2}.$$

Therefore, it leads to

$$|\beta(\lambda_n(t_\varepsilon)) - \beta(t_\varepsilon)| \leq \sup_{t_\varepsilon - r_\varepsilon < t < t_\varepsilon + r_\varepsilon} |\beta(t) - \beta(t_\varepsilon)| < \frac{\delta_\varepsilon}{4} \ \forall n \geq N_{\varepsilon,2}. \tag{2.5}$$

Inequality (2.4) together with (2.5) implies

$$x_n(\lambda_n(t_\varepsilon)) \geq \beta(\lambda_n(t_\varepsilon)) + \frac{\delta_\varepsilon}{4} \ \forall n \geq N_\varepsilon := \max\{N_{\varepsilon,1}, N_{\varepsilon,2}\},$$

or equivalently,

$$\pi(x_n) \leq \lambda_n(t_\varepsilon) \leq \lambda_n(\pi(x) + \varepsilon) \ \forall n \geq N_\varepsilon.$$

Finally, taking $\limsup_{n\to\infty}$ on each side of the above inequality and using $\|\lambda_n - \mathbb{I}\| \to 0$ of (2.3), we have

$$\limsup_{n\to\infty} \pi(x_n) \leq \limsup_{n\to\infty} \lambda_n(\pi(x) + \varepsilon) = \pi(x) + \varepsilon.$$

So we conclude that by the arbitrariness of $\varepsilon$,

$$\limsup_{n\to\infty} \pi(x_n) \leq \pi(x).$$

- Case 2. We prove the reverse inequality. By the definition of the first hitting time, for any $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ such that

$$(\beta - x)^*(\pi(x) - \varepsilon) > \delta_\varepsilon, \ (x - \alpha)^*(\pi(x) - \varepsilon) > \delta_\varepsilon. \tag{2.6}$$

Also, (2.3) implies that $\|\alpha \circ \lambda_n - \alpha\| \to 0$, $\|\beta \circ \lambda_n - \beta\| \to 0$, and $\|x_n \circ \lambda_n - x\| \to 0$. Using the triangle inequality, we have

$$\|(x_n - \alpha) \circ \lambda_n - (x - \alpha)\| \to 0, \ \|(x_n - \beta) \circ \lambda_n - (x - \beta)\| \to 0.$$

So, there exists a large $N_\varepsilon$ such that

$$\|(x_n - \alpha) \circ \lambda_n - (x - \alpha)\| < \frac{\delta_\varepsilon}{2}, \ \|(x_n - \beta) \circ \lambda_n - (x - \beta)\| < \frac{\delta_\varepsilon}{2} \ \forall n \geq N_\varepsilon.$$

In particular, this leads to

$$\begin{aligned}(\alpha - x_n) \circ \lambda_n(t) < (\alpha - x)(t) + \frac{\delta_\varepsilon}{2}, \ (x_n - \beta) \circ \lambda_n(t) \\ < (x - \beta)(t) + \frac{\delta_\varepsilon}{2} \ \forall t \in [0,1], \ n \geq N_\varepsilon.\end{aligned}$$

By taking $\sup_{[0,\pi(x)-\varepsilon]}$ on both sides of the above inequalities and using (2.6), we have

$$(x_n - \beta)^* \lambda_n(\pi(x) - \varepsilon) < (x - \beta)^*(\pi(x) - \varepsilon) + \frac{\delta_\varepsilon}{2} < -\frac{\delta_\varepsilon}{2} \ \forall n \geq N_\varepsilon$$

and

$$(\alpha - x_n)^* \lambda_n(\pi(x) - \varepsilon) < (\alpha - x)^*(\pi(x) - \varepsilon) + \frac{\delta_\varepsilon}{2} < -\frac{\delta_\varepsilon}{2} \ \forall n \geq N_\varepsilon.$$

Thus, we have

$$\pi(x_n) \geq \lambda_n(\pi(x) - \varepsilon) \ \forall n \geq N_\varepsilon.$$

Taking $\liminf_n$ and using (2.3) and the fact of the arbitrariness of $\varepsilon$, we have

$$\liminf_n \pi(x_n) \geq \pi(x).$$

Summarizing the above, we have proved the continuity of $\pi$ at $x \in C_1$ with respect to the Skorohod topology, i.e., $\lim_{n\to\infty} \pi(x_n) = \pi(x)$ whenever $x_n \to x \in C_1$ with respect to the Skorohod topology. Similar arguments as in Cases 1 and 2 yield that $\pi$ is also continuous at $x \in C_2$. □

**2.3. Main convergence results.** In this section, we utilize the continuity properties under Skorohod topology together with the continuous mapping theorem to obtain the main convergence result. To proceed, we make the following assumptions.

(A1) If $\beta < \infty$, then $\mathbb{P}\{(X - \beta)^*(1) \geq 0\} > 0$. If $\alpha > -\infty$, then $\mathbb{P}\{(-X + \alpha)^*(1) \geq 0\} > 0$.

Note that condition (A1) means that the process hits both barriers $\alpha$ and $\beta$ with positive probability whenever $-\alpha$ or $\beta$ are bounded above. This is not a restriction; it is imposed only for technical convenience since one can simply set $\alpha(t) = -\infty$ (resp., $\beta(t) = \infty$) for all $t \in [0, 1]$ if $X$ never hits $\alpha$ (resp., $\beta$) almost surely in $\mathbb{P}$. If $-\alpha(t) = \beta(t) = \infty$, then $\tau = 1$.

(A2) $X$ satisfies $\inf\{t > \tau : X(t) \notin [\alpha(t), \beta(t)]\} = \tau$ almost surely in $\mathbb{P}$.

Condition (A2) requires that the boundary $\partial\Gamma$ is regular with respect to the process $X$. Note that for any small $\varepsilon > 0$, $X$ exits from $\bar{\Gamma}$ in the interval $(\tau, \tau + \varepsilon)$ under (A2). Loosely speaking, condition (A2) means that the process $X(t)$ exits $\bar{\Gamma}$ immediately after it hits the boundary at $\tau$. Note that (A2) also implies that

$$\mathbb{P}\{X \in \cup_{i=1}^3 C_i\} = 1 \text{ or equivalently } \mathbb{P}\{X \in C_4\} = 0.$$

More discussions appear in Remark 10.

(A3) $g : \mathbb{R}^{4m+1} \to \mathbb{R}$ is an almost surely continuous function with respect to $\mathbb{P}Z^{-1}$, where $Z := \Pi(X, \tau\nu^{(1)}), \Pi(X, \nu^{(2)}), \Pi(X^*, \tau\nu^{(3)}), \Pi(X^*, \nu^{(4)}), \tau)$. In fact, (A3) is equivalent to the following statement: $g$ is discontinuous only at points in a set $\mathcal{N}$ satisfying $\mathbb{P}\{Z \in \mathcal{N}\} = 0$; see also Remark 11.

(A4) One of the following conditions holds:
1. $g$ is a bounded function;
2. $g$ is a function with linear growth and $\{X^h(t) : h > 0, t \in [0, 1]\}$ is uniformly integrable.

Now, we are ready to answer question (Q1).

THEOREM 9. *Assume* (A1)–(A4)*. Let $X$ be an $\mathcal{F}_t$-adapted continuous process with initial $X(0) = x$ and $X^h$ be a sequence of RCLL processes satisfying $X^h \Rightarrow X$ as $h \to 0$. Then,* $\lim_{h\to 0} V^h = V$.

*Proof.* We rewrite $V$ and $V^h$ by

$$V = \mathbb{E}[G(X)] \text{ and } V^h = \mathbb{E}[G(X^h)],$$

where $G : D[0,1] \mapsto \mathbb{R}$ is defined by

$$G(x) = g(\Pi(x, \pi(x)\nu^{(1)}), \Pi(x, \nu^{(2)}), \Pi(\mathcal{M}(x), \pi(x)\nu^{(3)}), \Pi(\mathcal{M}(x), \nu^{(4)}), \pi(x)).$$

Note that (A2) implies that $\mathbb{P}\{X \in \cup_{i=1}^3 C_i\} = 1$. Together with (A3) and Lemmas 6, 7, and 8, we have continuity of $G$ almost surely in $\mathbb{P}X^{-1}$. By the continuous mapping theorem [3, Theorem 2.7], we conclude that

$$G(X^h) \Rightarrow G(X) \text{ as } h \to 0.$$

Together with (A4), it results in $\lim_{h\to 0} V^h = V$; see [3, pp. 25, 31]. ∎

Theorem 9 holds under assumptions (A1)–(A4). Recall that (A1) is not a restriction. We elaborate on (A2), (A3), and (A4) in what follows.

*Remark* 10 (discussions on (A2) and Example 4). In fact, (A2) is a requirement on the regularity of the boundary $\partial\Gamma$ with respect to the process $X$, and it is referred to as $\tau'$-regularity for simplicity; see [19]. Note that since $X$ in Example 4 violates $\tau'$-regularity (A2), by observing

$$\inf\{t > \tau : X(t) \notin [\alpha(t), \beta(t)]\} \wedge 1 = 1 > 1/2 = \tau,$$

it yields the convergence to the wrong value. In other words, (A2) is crucial for the investigation of the convergence.

*Remark* 11 (discussions on (A3) and Example 2). Assumption (A3) is the requirement on the function $g$. First, it allows discontinuity of $g$, but it cannot be too much discontinuous in the sense that it is at least required to be almost surely continuous. However, it is already enough to include option pricing for the discontinuous payoff such as the barrier option in Example 2. In fact, $g$ of Example 2 given by

$$g(x_1, x_2, \ldots, x_{4m+1}) = e^{-r}\left(x_{2m} - \frac{1}{2}\right)^+ I_{[1,\infty)}(x_{4m})$$

is continuous only at $\{x_{4m} = 1\}$. Suppose the stock price $X$ follows a geometric Brownian motion; then the probability measure $\mathbb{P}$ satisfies $\mathbb{P}\{X^*(1) = 1\} = 0$, and $g$ is continuous almost surely in $\mathbb{P}Z^{-1}$.

*Remark* 12 (discussions on (A4) and Example 5). In (A4), another issue yet mentioned is the growth condition of $g$. In particular, if $g$ is is a function of linear growth, then one shall verify the uniform integrability. We have already seen that, for instance, Example 5 converges to a wrong value, since $\{X(t) : t \in (0,1)\}$ is not uniformly integrable, while the payoff function $g(x) = x$ is linear growth. Recall the definition of uniform integrability. A set of random variables $\{Y_\gamma : \gamma \in \Gamma\}$ is said to be uniformly integrable if for any $\varepsilon > 0$, there exists a compact set $K$ such that

$$\sup_{\gamma\in\Gamma} \mathbb{E}[Y_\gamma I_{\{Y_\gamma \notin K\}}] < \varepsilon.$$

Note that to ease the verification of the uniform integrability, one can often use Proposition 14 practically rather than the above definition.

## 3. Weak convergence of Markov chain approximation.

**3.1. Markov chain approximation.** In this section, we establish the weak convergence of Markov chain approximation for multidimensional SDEs. In what follows, $K$ is a generic constant whose value may change at each line.

(A5) $b$ and $\sigma$ are Lipschitz in $y$ and Hölder-1/2 continuous in $t$, i.e., with $\phi = b$, $\sigma$,

$$|\phi(y_1, t_1) - \phi(y_2, t_2)| \le K(|y_1 - y_2| + |t_1 - t_2|^{1/2}).$$

Let $Y = \{Y(t) : t \in [0, 1]\}$ be the unique solution of

$$dY(t) = b(Y(t), t)dt + \sigma(Y(t), t)dW(t);\ Y(0) = y, \tag{3.1}$$

where $b : \mathbb{R}^{d+1} \mapsto \mathbb{R}^d$, $W$ is a standard $\mathbb{R}^{d_1}$ Brownian motion, and $\sigma : \mathbb{R}^{d+1} \mapsto \mathbb{R}^{d\times d_1}$. Let $t_0^h = 0 \le t_1^h \le \cdots \le t_N^h = 1$ be a sequence of increasing predictable (i.e., $t_i^h$ is $\mathcal{F}_{i-1}^h$-measurable) random times with respect to a discrete filtration $\{\mathcal{F}_i^h : i = 0, 1, \ldots\}$, and let $\{Y_i^h : i = 1, 2, \ldots N\}$ be a sequence of $\{\mathcal{F}_i^h\}$-adapted Markov chain in $\mathbb{R}^d$ with transition probability

$$\mathbb{P}\{Y_{i+1}^h \in dy | Y_i^h = x, t_i^h = t\} = p^h(t, x, y).$$

We use $Y^h = \{Y^h(t) : t \in [0, 1]\}$ to denote piecewise constant interpolation

$$Y^h(t) = \sum_{i=0}^{n-1} Y_i^h I_{\{t_i^h \le t < t_{i+1}^h\}}. \tag{3.2}$$

For notational simplicity, we set $\Delta t_n^h = t_{n+1}^h - t_n^h$ and $\Delta Y_n^h = Y_{n+1}^h - Y_n^h$, $\mathbb{E}_n^h[\,\cdot\,] := \mathbb{E}[\,\cdot\,|\mathcal{F}_n^h]$, and

$$z^h(t) := \max\{n \ge 0 : t_n^h \le t\}, \quad \Delta M_n^h = \Delta Y_n^h - \mathbb{E}_n^h[\Delta Y_n^h].$$

The interpolation of the Markov chain process $Y^h$ is said to be *locally consistent* if

(LC1) $\mathbb{E}_n^h[\Delta Y_n^h] = \mathbb{E}_n^h[\Delta t_n^h] \cdot (b(Y_n^h, t_n^h) + O(h))$;

(LC2) $\mathrm{cov}(\Delta Y_n^h | \mathcal{F}_n^h) = \mathbb{E}_n^h[\Delta t_n^h] \cdot ((\sigma\sigma')(Y_n^h, t_n^h) + O(h))$, where $O(h)$ is either a $d$-dimensional vector or a $d \times d$-dimensional matrix that is $\mathcal{F}_n^h$-measurable with each element being $O(h)$.

To proceed, we also require quasi-uniform step size.

(QU) The step size $\{\Delta t_i^h\}$ satisfies $\frac{h}{K} \le \inf_i \Delta t_i^h \le \sup_n \Delta t_n^h \le Kh$.

The (QU) condition yields

$$z^h(t) \le Kt/h, \quad z^h(t) \sup_i \Delta t_i^h \le Kt \ \text{ almost surely.} \tag{3.3}$$

The main goal of this section is to show that $\{Y^h(t) : t \in [0, 1], h > 0\}$ is uniformly integrable and $Y^h \Rightarrow Y$ as $h \to 0$. Since the entire proof is rather long, we first provide some useful estimates.

LEMMA 13. *For the SDE* (3.1)*, assume that* (A5) *is satisfied, and* $Y^h$ *of* (3.2) *is a sequence of random processes satisfying the local consistency conditions* (LC1)–(LC2) *and quasi-uniform step size* (QU)*. Then, the family of random variables* $\{Y^h(t) : t \in [0,1], h > 0\}$ *satisfies*

$$\mathbb{E}[|Y^h(t)|^2] \leq K \,\forall t \in [0,1] \tag{3.4}$$

*and*

$$\mathbb{E}\left[\sup_{0\leq n\leq z^h(t)-1}\left|\sum_{i=0}^{n}\Delta Y_i^h\right|^2\right] \leq Kt. \tag{3.5}$$

*Proof.* First, we separate the entire estimation into two parts:

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq n\leq z^h(t)-1}\left|\sum_{i=0}^{n}\Delta Y_i^h\right|^2\right] &= \mathbb{E}\left[\sup_{0\leq n\leq z^h(t)-1}\left|\sum_{i=0}^{n}\mathbb{E}_i^h[\Delta Y_i^h]+\Delta M_i^h\right|^2\right]\\ &\leq K\mathbb{E}\left[\sup_{0\leq n\leq z^h(t)-1}\left|\sum_{i=0}^{n}\mathbb{E}_i^h[\Delta Y_i^h]\right|^2\right]\\ &\quad + K\mathbb{E}\left[\sup_{0\leq n\leq z^h(t)-1}\left|\sum_{i=0}^{n}\Delta M_i^h\right|^2\right]\\ &:= K\widetilde{I}_1 + K\widetilde{I}_2.\end{aligned}$$

Note that $\widetilde{I}_1$ has the following upper bound by local consistency:

$$\begin{aligned}\widetilde{I}_1 &\leq \mathbb{E}\left[z^h(t)\sum_{i=0}^{z^h(t)-1}\left|\mathbb{E}_i^h[\Delta Y_i^h]\right|^2\right]\\ &\leq K\mathbb{E}\left[z^h(t)\sum_{n=0}^{z^h(t)-1}(\mathbb{E}_n^h[\Delta t_n^h])^2\cdot(b^2(Y_n^h,t_n^h)+O(h^2))\right].\end{aligned}$$

Owing to the (QU) condition, we can use inequality (3.3) to obtain

$$\begin{aligned}\widetilde{I}_1 &\leq Kt\mathbb{E}\left[\sum_{n=0}^{z^h(t)-1}\mathbb{E}_n^h[\Delta t_n^h]\cdot(b^2(Y_n^h,t_n^h)+O(h^2))\right]\\ &\leq Kt\mathbb{E}\left[\sum_{n=0}^{[Kt/h]-1}\mathbb{E}_n^h[\Delta t_n^h\cdot(b^2(Y_n^h,t_n^h)+O(h^2))\cdot I_{\{n\leq z^h(t)-1\}}]\right].\end{aligned}$$

In the last line above, the term $(b^2(Y_n^h,t_n^h)+O(h^2))\cdot I_{\{n\leq z^h(t)-1\}}$ can be included in $\mathbb{E}_n^h$, since it is $\mathcal{F}_n^h$-measurable. Now we are ready to use the tower property of conditional expectation and regularity condition (A5) and end up with

$$\widetilde{I}_1 \leq Kt\mathbb{E}\left[\int_0^t |Y_s|^2 ds\right] + Kt^2h^2. \tag{3.6}$$

On the other hand,

$$\begin{aligned}
\widetilde{I}_2 &= \mathbb{E}\left[\sup_{0\le n\le z^h(t)-1}\left|\sum_{i=0}^{n}\Delta M_i^h\right|^2\right]\\
&\le K\mathbb{E}\left[\left|\sum_{i=0}^{z^h(t)-1}\Delta M_i^h\right|^2\right] \quad \text{by Doob's maximal inequality}\\
&= K\mathbb{E}\left[\sum_{i=0}^{z^h(t)-1}|\Delta M_i^h|^2\right] \quad \text{since } \mathbb{E}[(\Delta M_i^h)'(\Delta M_j^h)] = 0\ \forall i\neq j\\
&\le K\mathbb{E}\left[\sum_{i=0}^{z^h(t)-1}\Delta t_i^h \mathrm{tr}(\sigma\sigma')(Y_i^h, t_i^h)\right] + Kht \quad \text{by (LC2)},
\end{aligned}$$

where $\mathrm{tr}(A)$ denotes the trace of $A$. Together with (A5), this implies that

$$\widetilde{I}_2 \le K\mathbb{E}\left[\int_0^t |Y_s^h|^2 ds\right] + Kht. \tag{3.7}$$

Combining (3.6) and (3.7) yields that

$$\mathbb{E}\left[\sup_{0\le n\le z^h(t)-1}\left|\sum_{i=0}^{n}\Delta Y_i^h\right|^2\right] \le Kth + K\mathbb{E}\left[\int_0^t |Y^h(s)|^2 ds\right]. \tag{3.8}$$

Therefore, we have

$$\mathbb{E}[|Y^h(t)|^2] \le K\mathbb{E}\left[|y|^2 + \left|\sum_{i=0}^{z^h(t)-1}\Delta Y_i^h\right|^2\right] \le K + K\mathbb{E}\left[\int_0^t |Y^h(s)|^2 ds\right].$$

Gronwall's inequality then yields the result of (3.4). Pluging (3.4) into (3.8), we conclude (3.5). ∎

To proceed, we also need the following convenient proposition for the uniform integrability, and the reader is referred to [7] for the proof of the general case.

PROPOSITION 14. *A family of real-valued random variable $\{Z_\theta : \theta\in\Theta\}$ for some index set $\Theta$ is uniformly integrable if $\sup_\theta \mathbb{E}[|Z_\theta|^p] < \infty$ for some $p > 1$.*

THEOREM 15. *Under the same assumptions as that of Lemma 13, the family of random variables $\{Y^h(t) : t\in[0,1], h > 0\}$ is uniformly integrable, and $Y^h \Rightarrow Y$ as $h\to 0$.*

*Proof.* We conclude first $\{Y^h(t) : 0\le t\le T, h > 0\}$ is uniformly integrable by Proposition 14 together with (3.4). We divide the rest of the proof into several steps.

1. We consider the tightness of $\{\mathbb{P}_h = \mathbb{P}(Y^h)^{-1} : h > 0\}$. It is enough to verify conditions imposed on Theorem 13.3.2 in [3] and its subsequent corollary.
   (a) By Chebyshev's inequality and (3.4)

$$\lim_{a\to\infty}\limsup_{h\to 0}\mathbb{P}\{|Y^h(t)|\ge a\} \le \lim_{a\to\infty}\limsup_{h\to 0}\frac{1}{a^2}\mathbb{E}[|Y^h(t)|^2] = 0.$$

(b) For the purpose of characterization of tightness for discontinuous functions, we need to introduce notions of modulus of continuity $\omega(Y^h,\delta)$ and $\omega'(Y^h,\delta)$ as follows. First, define $\omega(Y^h,\mathcal{I}) = \sup_{r_1,r_2\in\mathcal{I}}|Y^h(r_1) - Y^h(r_2)|$ for any subset $\mathcal{I}\subset[0,1]$. We also use $\mathcal{T}(\delta)$ to denote the collection of all $\delta$-sparse partitions of $[0,1]$. Then, for any $\delta>0$, we can define

$$\omega(Y^h,\delta) = \sup_{0\le t\le 1-\delta}\omega(Y^h,[t,t+\delta])$$

and

$$\omega'(Y^h,\delta) = \inf_{\{t_i\}\in\mathcal{T}(\delta)}\max_i \omega(Y^h,[t_{i-1},t_i)).$$

For the purpose of tightness of discontinuous functions, we also need to introduce a modified version of modulus of continuity. Thanks to (3.5), we have, for an arbitrary $\varepsilon>0$,

$$\begin{aligned}
&\lim_{\delta\to0}\limsup_{h\to0}\mathbb{P}\{\omega'(Y^h,\delta)\ge\varepsilon\}\\
&\quad\le\lim_{\delta\to0}\limsup_{h\to0}\mathbb{P}\{\omega(Y^h,2\delta)\ge\varepsilon\}\\
&\quad\le\lim_{\delta\to0}\limsup_{h\to0}\frac{1}{\varepsilon^2}\mathbb{E}[\omega^2(Y^h,2\delta)]\\
&\quad\le\lim_{\delta\to0}\limsup_{h\to0}\frac{1}{\varepsilon^2}\mathbb{E}\left[\sup_{z^h(t)-1\le n\le z^h(t+2\delta)-1}\left|\sum_{i=0}^n\Delta Y_i^h\right|^2\right]\\
&\quad\le\lim_{\delta\to0}\limsup_{h\to0}\frac{1}{\varepsilon^2}K\delta=0.
\end{aligned}$$

As a result, $\{\mathbb{P}_h=\mathbb{P}(Y^h)^{-1}:h>0\}$ is tight.

2. Since $Y^h$ is tight, for an arbitrary infinite sequence, there exists a subsequence that has a weak limit. For notational convenience, we denote this subsequence again by $\{Y^h\}$ and its limit by $\bar{Y}$. Due to uniqueness of the weak solution, it is enough to show that $\bar{Y}$ is the weak solution of (3.1). Since $Y^h\Rightarrow\bar{Y}$, the uniform integrability and (A5) lead to

$$\lim_{h\to0}\mathbb{E}[Y^h(t)]=\mathbb{E}[\bar{Y}(t)]\quad\text{and}$$
$$\lim_{h\to0}\mathbb{E}[b(Y^h(t),t)]=\mathbb{E}[b(\bar{Y}(t),t)].$$

Therefore, if we set

$$M(t):=\bar{Y}(t)-\bar{Y}(0)-\int_0^t b(\bar{Y}(s),s)ds,$$

we have

$$\begin{aligned}
\mathbb{E}[M(t)]&=\lim_{h\to0}\mathbb{E}\left[Y^h(t)-Y^h(0)-\int_0^t b(Y^h(s),s)ds\right]\\
&=\lim_{h\to0}\mathbb{E}\left[\sum_{i=1}^{z^h(t)}\left((Y_i^h-Y_{i-1}^h)-\int_{t_{i-1}^h}^{t_i^h}b(Y_{i-1}^h,t)dt\right)\right]\\
&\quad-\lim_{h\to0}\mathbb{E}\left[\int_{t^h_{z^h(t)}}^t b(Y^h_{z^h(t)},s)ds\right].
\end{aligned}\tag{3.9}$$

Regarding the last term above, by linear growth of $b$

$$\begin{aligned}\lim_{h\to 0}\mathbb{E}\left[\left|\int_{t^h_{z^h(t)}}^t b(Y^h_{z^h(t)},s)ds\right|\right] &\le \lim_{h\to 0}\mathbb{E}\Big[\int_{t^h_{z^h(t)}}^t K+K|Y^h_{z^h(t)}|ds\Big]\\ &\le K\lim_{h\to 0}\mathbb{E}\big[\Delta t^h_{z^h(t)}(1+|Y^h(t)|)\big].\end{aligned}$$

Therefore, by (QU) and (3.4), we conclude that the second term in (3.9) satisfies

$$\lim_{h\to 0}\mathbb{E}\Big[\int_{t^h_{z^h(t)}}^t b(Y^h_{z^h(t)},s)ds\Big]=0.$$

As for the first term of (3.9), using the Hölder continuity in $t$ and the tower property on local consistency,

$$\begin{aligned}&\lim_{h\to 0}\mathbb{E}\left[\sum_{i=1}^{z^h(t)}\left((Y^h_i-Y^h_{i-1})-\int_{t^h_{i-1}}^{t^h_i} b(Y^h_{i-1},t)dt\right)\right]\\ &=\lim_{h\to 0}\mathbb{E}\left[\sum_{i=1}^{z^h(t)}\big((Y^h_i-Y^h_{i-1})-b(Y^h_{i-1},t)\Delta t^h_{i-1}+O(|\Delta t^h_{i-1}|^{3/2})\big)\right]\\ &=\lim_{h\to 0}\mathbb{E}\left[\sum_{i=1}^{z^h(t)}\big(\Delta t^h_{i-1}O(h)+O(|\Delta t^h_{i-1}|^{3/2})\big)\right]=0.\end{aligned}$$

Therefore, $\mathbb{E}[M(t)]=0$. In fact, one uses exactly the same procedure to show $\mathbb{E}[M(t)|\mathcal{F}_s]=M(s)$ for any $0\le s\le t$, and we conclude that $M(t)$ is a martingale.

Next, we use $\bar Y_l$ and $Y^h_l$ to denote the $l$th component of the vector process $\bar Y$ and $Y^h$ and use $\langle \bar Y_l,\bar Y_m\rangle(t)$ to denote the cross variation of two real processes $Y_l$ and $Y_m$ up to time $t$. Then,

$$\begin{aligned}&\mathbb{E}\left|\langle\bar Y_l,\bar Y_m\rangle(t)-\int_0^t(\sigma_l\sigma_m)(\bar Y(s),s)ds\right|\\ &=\lim_{h\to 0}\mathbb{E}\left|\langle Y^h_l,Y^h_m\rangle(t)-\int_0^t(\sigma_l\sigma_m)(Y^h(s),s)ds\right|\\ &=\lim_{h\to 0}\mathbb{E}\left|\sum_{i=1}^n(Y^h_{l,i}-Y^h_{l,i-1})(Y^h_{m,i}-Y^h_{m,i-1})-\int_0^t(\sigma_l\sigma_m)(Y^h(s),s)ds\right|=0.\end{aligned}$$

The last equality can be obtained similar to the above owing to $\mathbb{E}[M(t)]=0$, using the local consistency, and regularity assumption (A5) on $\sigma$. Therefore, the cross variation $\langle\bar Y_l,\bar Y_m\rangle(t)=\int_0^t(\sigma_l\sigma_m)(\bar Y(s),s)ds$. This again implies that $M(t)$ is a $d$-dimensional martingale process with its quadratic variation $\langle M\rangle(t)=\int_0^t(\sigma\sigma')(\bar Y(s),s)ds$. Applying Levy's martingale characterization on the time-changed Brownian motion, there exists a $d_1$-dimensional Brownian motion $B(t)$ such that $M(t)=\int_0^t\sigma(\bar Y(s),s)dB(s)$. Therefore, $\bar Y$ is the weak solution of (3.1).

Summarizing all the above leads to the conclusion. □

**3.2. Examples of weak convergence of the Markov chain approximation.** The above construction of the Markov chain approximation is based on the local consistency, which is slightly different from the local consistency given by [14, Theorem 10.4.1]. As a result, the convergence result of the Markov chain approximation is generalized in the following sense. $\sigma$ and $b$ may be unbounded but have linear growth. Therefore, the geometric Brownian motion is covered by weak convergence result of Theorem 15 as an important application. In fact, locally consistent Monte Carlo approximation is flexible for its various choices. For illustration, we give several simple Monte Carlo approximations for the one-dimensional process, which is not included in [14].

*Example* 16. Euler approximation can be considered as a special case of Monte Carlo approximations. Let $\{Y_n^h\}$ be a Markov chain generated by the following:

1. $Y_0^h = y, t_0^h = 0$.
2. Let the transition probability be $\Delta t_n^h = h$,

$$Y_{n+1}^h = Y_n^h + b(Y_n^h, nh)h + \sigma(Y_n^h, nh)\sqrt{h}N_n, \tag{3.10}$$

where $\{N_n\}$ is a sequence of independent and identically distributed (i.i.d.) standard normal random variables. One can easily verify all local consistency conditions. Then assuming (A5), Theorem 15 implies that the piecewise constant interpolation $Y^h$ defined in (3.2) converges weakly to $Y$, the solution to (3.1).

*Example* 17. The following MC approximation can be considered as an extension of binomial approximation of Brownian motion. Let $d = d_1 = 1$. Let $\{Y_n^h\}$ be a Markov chain generated by the following:

1. $Y_0^h = y, t_0^h = 0$.
2. Let the transition probability be, with $\Delta t_n^h = h$,

$$\mathbb{P}\big(Y_{n+1}^h = Y_n^h + b(nh, Y_n^h)h \pm \sigma(nh, Y_n^h)\sqrt{h}\big|Y_n^h\big) = 1/2. \tag{3.11}$$

The above Markov chain is locally consistent since direct computation leads to

1. $\mathbb{E}[\Delta Y_n^h | Y_n^h = y, t_n^h = t] = \mathbb{E}[\Delta t_n^h | Y_n^h = y, t_n^h = t] \cdot b(y,t)$,
2. $\mathrm{cov}(\Delta Y_n^h | Y_n^h = y, t_n^h = t) = \mathbb{E}[\Delta t_n^h | Y_n^h = y, t_n^h = t] \cdot \sigma^2(y,t)$.

Therefore, assuming (A5), the piecewise constant interpolation $Y^h$ of the form (3.2) is convergent to $Y$ of (3.1) by Theorem 15.

*Example* 18. Assume $|\sigma| \wedge |1/\sigma| > \epsilon > 0$ in addition to (A5). Then one can use a binomial tree type approximation of the diffusion term $\sigma(Y(t),t)dW(t)$ in (3.1). Let $d = d_1 = 1$. Let $\{Y_n^h\}$ be a Markov chain generated by the following:

1. $Y_0^h = y, t_0^h = 0$.
2. Let $\Delta t_n^h(Y_n^h, t_n^h) = \frac{h}{\sigma^2(Y_n^h, t_n^h)}$ and the transition probability be

$$\mathbb{P}\big(Y_{n+1}^h = Y_n^h + b(Y_n^h, t_n^h)\Delta t_n^h(Y_n^h, t_n^h) \pm \sqrt{h}\big|(Y_n^h, t_n^h)\big) = 1/2. \tag{3.12}$$

Note that this Markov chain satisfies (QU) as well as local consistency since

1. $\mathbb{E}[\Delta Y_n^h | Y_n^h = y, t_n^h = t] = \mathbb{E}[\Delta t_n^h | Y_n^h = y, t_n^h = t] \cdot b(y,t)$,
2. $\mathrm{cov}(\Delta Y_n^h | Y_n^h = y, t_n^h = t) = \mathbb{E}[\Delta t_n^h | Y_n^h = y, t_n^h = t] \cdot (\sigma^2(y,t) + O(h))$.

Therefore, the piecewise constant interpolation $Y^h$ of the form (3.2) is convergent to $Y$ of (3.1) by Theorem 15.

**3.3. Can we expect strong convergence?** Can we expect strong convergence? In general, the answer is no. To illustrate, we consider a special case of Euler approximation of Example 16. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which $\mathcal{F}_t$ is filtration satisfying usual conditions, and $W_t$ of (3.1) is a standard one-dimensional Brownian motion. We construct strong approximation of the EM method by taking on $N_n$ of (3.10) by

$$N_n = \frac{W_{n+1} - W_n}{\sqrt{h}}.$$

Under assumption (A5), the SDE (3.1) has a unique strong solution. Suppose $\hat{Y}^h$ is a continuous interpolation of Euler approximation of $Y^h$ given by

$$\hat{Y}^h(t) = \hat{Y}^h_{nh} + b(nh, \hat{Y}^h_{nh})(t - nh) + \sigma(nh, \hat{Y}^h_{nh})(W(t) - W(nh)) \quad \text{for } t \in [nh, nh + h).$$

Then a classical result (see, e.g., [15, Theorem 2.7.3]) shows that

$$\mathbb{E}\Big[\sup_{0 \le t \le T} |Y(t) - \hat{Y}^h(t)|\Big] \le K h^{1/2}.$$

However, the above inequality fails for the piecewise constant interpolation of EM approximation $\{Y^h\}$. Otherwise, we have the following simple counterexample. Consider the EM approximation of $W_t$ on $[0, 1]$ by equal step size $h = 1/N$. Then, we have

$$\mathbb{E}\Big[\sup_{0 \le t \le 1} |W(t) - W([Nt]/N)|\Big] = \mathbb{E}\left[\sup_{1 \le n \le N} \sup_{(n-1)/N \le t < n/N} \left| W(t) - W\left(\frac{n-1}{N}\right)\right|\right].$$

Note that $\bar{W}(t) = \sqrt{N} W(t/N)$ is a standard Brownian motion w.r.t. a time-scaled filtration. So one can reduce the above equality as

$$\mathbb{E}\Big[\sup_{0 \le t \le 1} |W(t) - W([Nt]/N)|\Big] = \frac{1}{\sqrt{N}} \mathbb{E}\Big[\sup_{1 \le n \le N} \Lambda_n\Big],$$

where $\{\Lambda_n\}$ are i.i.d. random variables defined by

$$\Lambda_n = \sup_{n-1 \le t < n} |\bar{W}(t) - \bar{W}(n-1)|.$$

Since, $\Lambda_n$'s are unbounded i.i.d. random variables, $\mathbb{E}\big[\sup_{1 \le n \le N} \Lambda_n\big]$ goes to infinity as $N \to \infty$. This shows that

$$\mathbb{E}\Big[\sup_{0 \le t \le 1} |W(t) - W([Nt]/N)|\Big] > O(N^{-1/2}).$$

In conclusion, one cannot expect more than weak convergence merely under local consistency.

## 4. Ramification and further remarks.

**4.1. Application to discrete-monitoring-barrier option underlying stochastic volatility.** We begin this section with an application of Theorem 9 to the following stochastic volatility model; see [1]. Let $W$ and $B$ be two standard Brownian motions with correlation $\rho$ in the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathcal{F}_t)$. Suppose that the stock price follows

$$dX(t) = X(t)(rdt + \sigma(Y(t))dW(t)) \tag{4.1}$$

with initial $X(0) = x > 0$ and that the volatility follows

$$dY(t) = Y(t)(\mu(t)dt + b(t)dB(t)) \tag{4.2}$$

with initial $Y(0) = y > 0$. We consider a discrete-monitoring-barrier option price

$$V = e^{-r}\mathbb{E}\left[\left(X(1) - \frac{1}{2}\right)^+ I_{[1,\infty)}\Big(\max_{1\le i\le m} X(i/m)\Big)\right]$$

given by (1.4) of Example 3 by the discrete scheme (1.5):

$$V^h = \mathbb{E}[g(\Pi(X^h, \tau^h\nu^{(1)}), \Pi(X^h, \nu^{(2)}), \Pi(X^{h,*}, \tau^h\nu^{(3)}), \Pi(X^{h,*}, \nu^{(4)}), \tau^h)],$$

where $g : \mathbb{R}^{4m+1} \mapsto \mathbb{R}$ is of the form

$$g(x_1, x_2, \ldots, x_{4m+1}) = e^{-r}\left(x_{2m} - \frac{1}{2}\right)^+ I_{[1,\infty)}\Big(\max_{m+1\le i\le 2m} x_i\Big).$$

One can check that $X$ and $Y$ are a unique nonnegative strong solution of SDEs (4.1) and (4.2) if $\sigma$ satisfies polynomial growth, and $\mu$ and $b$ are Hölder-1/2 continuous. In addition, we assume

$$\text{nondegeneracy of } X, \text{ i.e., } \sigma(y) > 0\ \forall\ y > 0.$$

Note that it is possible to have $\sigma(0) = 0$ under the above assumption. Below, we examine the convergence $\lim_h V^h = V$ when $X^h$ is constructed by the Euler approximation of Example 16.

Note that the regularity condition (A2) is satisfied by [18, Proposition A.1] because $\sigma(\cdot) > 0$. On the other hand, the payoff function $g$ is only discontinuous at the points in the set $\{\max_{m+1\le i\le 2m} x_i = 1\}$. Thanks to the fact

$$\mathbb{P}\Big\{\max_{1\le i\le m} X(i/m) = 1\Big\} = 0,$$

it implies that $g$ is almost surely continuous with respect to $\mathbb{P}$. Another thing yet to be verified is the uniform integrability of $\{X^h(t) : h, t\}$, since $g$ is linear growth. Thanks to Theorem 15 together with (A5), we have $X^h \Rightarrow X$ and the desired uniform integrability holds. Therefore, we reach the affirmative answer $\lim_h V^h = V$.

For the simple demonstration, we present a computational result on the above example with the following data. Let the initial stock price be $X_0 = 0.8$. For simplicity, we assume constant interest rate and volatility $r = .1$ and $\sigma = .3$. Suppose stock is monitored monthly, i.e., $m = 12$. If we compute $k = 5000$ many sample paths, and each sample path is computed by the Euler method with $n = 60000$ subintervals, then the computational result shows that the 95% interval is $[0.2310, 0.2364]$. The total MATLAB running time on a MacBook Air is 194 seconds. The MATLAB code is available for download at http://01law.wordpress.com/2013/07/18/code/.

**4.2. Further remarks.** This work has been devoted to analyzing approximation to path-dependent functionals. Using the methods of weak convergence, we have provided a unified approach for proving the convergence of numerical approximation of path-dependent functionals for a wide range of applications.

A possible alternative approach to study the convergence may be to evaluate the approximating problem by perturbing the boundary $(\alpha(t), \beta(t))$ to $(\alpha(t)-h, \beta(t)+h)$. For instance, [17] has studied the property of the value function using the aforementioned perturbation when the value function is non-path-dependent and the HJB equation is available. It is interesting to check if a similar approach works for the path-dependent case.

This paper focused on diffusion models. For future work, it would be worthwhile to examine systems driven by pure jump processes and jump diffusions and systems with an additional factor process such as the popular regime-switching process. Systems with memory (time delays) form another class of important problems. Much work can also be devoted to numerical solutions of various SDEs, coordination of multiagent systems, and many Monte Carlo optimization problems in which one needs to treat path-dependent functionals.

## REFERENCES

[1] E. Bayraktar, K. Kardaras, and H. Xing, *Valuation equations for stochastic volatility models*, SIAM J. Finan. Math., 3 (2012), pp. 351–373.
[2] E. Bayraktar, Q. S. Song, and J. Yang, *On the continuity of stochastic control problems on bounded domains*, Stochastic Anal. Appl., 29 (2011), pp. 48–60.
[3] P. Billingsley, *Convergence of Probability Measures*, Wiley Ser. Probab. Statist. Probab. Statist., 2nd ed., Wiley-Interscience, New York, 1999.
[4] M. Broadie, P. Glasserman, and S. G. Kou, *Connecting discrete and continuous path-dependent options*, Finance Stoch., 3 (1999), pp. 55–82.
[5] B. de Saporta, F. Dufour, and K. Gonzalez, *Numerical method for optimal stopping of piecewise deterministic Markov processes*, Ann. Appl. Probab., 20 (2010), pp. 1607–1637.
[6] F. Dufour and A. B. Piunovskiy, *Multiobjective stopping problem for discrete-time Markov processes: Convex analytic approach*, J. Appl. Probab., 47 (2010), pp. 947–966.
[7] S. Ethier and T. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley Ser. Probab. Stat., Wiley-Interscience, 1986.
[8] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.
[9] M. Jeanblanc, M. Yor, and M. Chesney, *Mathematical Methods for Financial Markets*, Springer Finance, Springer, New York, 2009.
[10] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Graduate Texts in Math. 113, Springer, New York, 1991.
[11] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Appl. Math. (N.Y.) 23, Springer, New York, 1992.
[12] S. G. Kou, *On pricing of discrete barrier options*, Statist. Sinica, 13 (2003), pp. 955–964.
[13] H. J. Kushner, *Numerical methods for stochastic control problems in continuous time*. SIAM J. Control Optim., 28 (1990), pp. 999–1048.
[14] H. J. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time*, 2nd ed., Appl. Math. (N.Y.) 24, Springer, New York, 2001.
[15] X. R. Mao, *Stochastic Differential Equations and Applications*, Horwood, Chichester, UK, 2007.
[16] R. L. McDonald, *Derivatives Markets*, 2nd ed., Pearson Education, London, 2006.
[17] Q. S. Song and G. Yin, *Rates of convergence of numerical methods for controlled regime-switching diffusions with stopping times in the costs*, SIAM J. Control Optim., 48 (2009), pp. 1831–1857.
[18] Q. S. Song, G. Yin, and C. Zhu, *Optimal switching with constraints and utility maximization of an indivisible market*, SIAM J. Control Optim., 50 (2012), pp. 629–651.
[19] D. Stroock and S. R. S. Varadhan, *On degenerate elliptic-parabolic operators of second order and their associated diffusions*, Comm. Pure Appl. Math., 25 (1972), pp. 651–713.
[20] L. Szpruch and D. J. Higham, *Comparing hitting time behavior of Markov jump processes and their diffusion approximations*, SIAM Multiscale Model. Simul., 8 (2010), pp. 605–621.