

Approximation Methods for Hybrid Diffusion Systems with State-dependent Switching Processes: Numerical Algorithms and Existence and Uniqueness of Solutions

G. Yin,^{*} Xuerong Mao,[†] Chenggui Yuan,[‡] Dingzhou Cao[§]

Abstract

By focusing on hybrid diffusions in which continuous dynamics and discrete events coexist, this work is concerned with approximation of solutions for hybrid stochastic differential equations with a state-dependent switching process. Iterative algorithms are developed. The continuous-state dependent switching process presents added difficulties in analyzing the numerical procedures. Weak convergence of the algorithms is established by a martingale problem formulation first. This weak convergence result is then used as a bridge to obtain strong convergence. In this process, the existence and uniqueness of the solution of the switching diffusions with continuous-state-dependent switching are obtained. Different from the existing results of solutions of stochastic differential equations in which the Picard iterations are utilized, Euler's numerical schemes are considered here. Moreover, decreasing stepsize algorithms together with their weak convergence are given. Numerical experiments are also provided for demonstration.

Key Words. Switching diffusion, numerical algorithm, convergence, existence and uniqueness of strong solution.

AMS Subject Classification. 65C30, 60H35, 65C05.

Running Title. Approximation of Switching Diffusions

^{*}Department of Mathematics, Wayne State University, Detroit, U.S.A. The research of this author was supported in part by the National Science Foundation.

[†]Department of Statistics and Modelling Science, University of Strathclyde, Glasgow G1 1XH, U.K.

[‡]Department of Mathematics, University of Wales Swansea, Swansea SA2 8PP, U.K.

[§]Department of Mathematics, Wayne State University, Detroit, U.S.A.

1 Introduction

Recently, hybrid systems in which continuous dynamics and discrete events coexist have drawn much attention. In particular, resurgent efforts have been devoted to the research on switching diffusion systems (also known as regime-switching diffusions). Much of the study was originated from applications arising from two-time-scale systems, control engineering practice, manufacturing systems, estimation and filtering, and financial engineering; see [5, 7, 15, 17, 19, 20, 21, 24], among others. Recent progress for switching diffusions has been summarized in [14] and references therein. Up to this point, the study has been carried out for such systems whose switching component is a continuous-time Markov chain independent of the continuous state variable [13, 23], whereas lesser is known for the processes with continuous-state-dependent switching [1, 25]. Nevertheless, in many applications, discrete events and continuous dynamics are intertwined, and the independence assumption of the discrete-event process and the continuous component poses restrictions. It would be nice to be able to handle the coupling and dependence of the continuous states and discrete events. Numerical methods for stochastic differential equations have been studied extensively, for example in [4, 9, 16] among others, whereas numerical solutions for stochastic differential equations with Markovian switching have also been well studied (see [14] and the many references therein). Numerical solutions with state-dependent switching diffusions have not been well understood to the best of our knowledge.

Suppose that $\mathcal{M} = \{1, \dots, m\}$ is a finite set. Consider the hybrid diffusion system

$$\begin{aligned} dx(t) &= f(x(t), \alpha(t))dt + \sigma(x(t), \alpha(t))dw(t), \\ x(0) &= x_0, \quad \alpha(0) = \alpha_0, \end{aligned} \tag{1.1}$$

and

$$P(\alpha(t + \Delta t) = j | \alpha(t) = i, x(s), \alpha(s), s \leq t) = q_{ij}(x(t))\Delta t + o(\Delta t), \quad i \neq j, \tag{1.2}$$

where $w(\cdot)$ is an r -dimensional standard Brownian motion, $x(t) \in \mathbb{R}^r$, $f(\cdot, \cdot) : \mathbb{R}^r \times \mathcal{M} \mapsto \mathbb{R}^r$, and $\sigma(\cdot, \cdot) : \mathbb{R}^r \times \mathcal{M} \mapsto \mathbb{R}^{r \times r}$ are appropriate functions satisfying certain regularity conditions, and $Q(x) = (q_{ij}(x)) \in \mathbb{R}^{m \times m}$ satisfies that for each x , $q_{ij}(x) \geq 0$ for $i \neq j$, $\sum_{j=1}^m q_{ij}(x) = 0$ for each $i \in \mathcal{M}$. There is an associated operator for the switching diffusion process defined as follows. Use z' to denote the transpose of z , and use ∇ and ∇^2 as the notation for gradient and Hessian, respectively. For each $i \in \mathcal{M}$ and suitable smooth function $h(\cdot, i)$, define an

operator

$$\mathcal{L}h(x, i) = \nabla h'(x, i)f(x, i) + \frac{1}{2}\text{tr}[\nabla^2 h(x, i)\sigma(x, i)\sigma'(x, i)] + \sum_{j=1}^m q_{ij}(x)h(x, j). \quad (1.3)$$

The state-dependent model is largely motivated by applications in control and optimization. Similar to the literature in Markov decision processes, we may consider a Markov process $(x(t), \alpha(t))$ whose generator has control dependence and the control is in feedback form resulting in the consideration of switching process being state dependent. The consideration of \mathcal{M} being a finite set is mainly from an application point of view. For emerging applications arising in manufacturing, wireless communication, and finance, the set \mathcal{M} is naturally a finite set. In principle, \mathcal{M} can be an infinite set. However, if \mathcal{M} is an infinite set, we will need to deal with a coupled system with infinitely many components. In numerical approximation, we shall approximate the resulting system with finitely many components. For such \mathcal{M} , we often approximate it by a large-dimensional finite set \mathcal{M} . Reduction of computational complexity for large-scale systems can be carried out using a time-scale separation approach [22].

Since the solution of (1.1) can often be obtained only through numerical approximations, constructing numerical solutions is of foremost importance. In this paper, our aim is to construct numerical approximation schemes for solving (1.1) with $\alpha(t)$ taking values in $\mathcal{M} = \{1, \dots, m\}$. The novelty of our work lies in that $Q(x)$, the generator of the switching process $\alpha(t)$, is state dependent, which makes the analysis much more difficult. One of the main difficulties is that due to the continuous-state dependence, $\alpha(t)$ and $x(t)$ are dependent; $\alpha(t)$ is a Markov chain only for a fixed x but is otherwise non-Markovian. Unlike the usual diffusion processes represented by stochastic differential equations, the distribution of the switching diffusion has mixture distribution. The essence in our approach is to treat the pair of processes $(x(t), \alpha(t))$ jointly; the two-component process turns out to be Markovian. Nevertheless, much care needs to be exercised to handle the mixture distributions. To proceed, we will use the following conditions.

(A1) The $Q(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^{m \times m}$ is a bounded and continuous function.

(A2) The functions $f(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ satisfy

- (a) $|f(x, \alpha)| \leq K(1 + |x|)$, $|\sigma(x, \alpha)| \leq K(1 + |x|)$, and
- (b) $|f(x, \alpha) - f(z, \alpha)| \leq K_0|x - z|$ and $|\sigma(x, \alpha) - \sigma(z, \alpha)| \leq K_0|x - z|$ for some $K > 0$ and $K_0 > 0$ and for all $x, z \in \mathbb{R}^r$ and $\alpha \in \mathcal{M}$.

Note that condition (A1) is simply a condition on the function $Q(x)$. If $Q(x) = Q$, we are back to the case of Markovian switching diffusions. Here owing to the x -dependence, the problem becomes more complex. The switching process and the diffusions are intertwined and dependent. Using Poisson random measures (see [18] and also [1]), under conditions (A1) and (A2), it can be shown that (1.1) has a unique solution for each initial condition by following Ikeda and Watanabe [6] with the appropriate use of the stopping times. However, the Picard iteration method does not work. In this paper, we construct Euler's scheme with a constant stepsize for approximating solutions of switching diffusions. As a result, our method differs from the usual approach. To obtain the convergence of the algorithms, we first show the weak convergence of the algorithm by means of martingale problem formulation. Then this result is used as a bridge to obtain strong convergence. In this process, we prove the existence and uniqueness of the solution of (1.1). This result is of independent interest in its own right; it enables us to obtain the strong convergence of the numerical algorithms. As a demonstration, we provide numerical experiments to delineate sample path properties of the approximating solutions. In addition, we present a decreasing stepsize algorithm, and obtain its weak convergence.

To see that the state-dependence switching can contribute to much of the difficulty, we consider the following scenario. Let a switching diffusion $(x(t), \alpha(t))$ be given by (1.1). First let the initial data be $(x(0), \alpha(0)) = (x, \alpha)$ and then consider another initial data $(x(0), \alpha(0)) = (y, \alpha)$ for $y \neq x$. For the state-dependent switching case, since $Q(x)$ depends on x , $\alpha^{x,\alpha}(t) \neq \alpha^{y,\alpha}(t)$ infinitely often even though initially $\alpha^{x,\alpha}(0) = \alpha^{y,\alpha}(0) = \alpha$, where the superscript above denotes the initial data dependence.

In what follows, for notational simplicity, we use the convention that K represents a generic positive constant, whose values may be different for different appearances so that $K + K = K$ and $KK = K$ is understood in an appropriate sense. In addition, for $z \in \mathbb{R}^{\iota_1 \times \iota_2}$ for some $\iota_i \geq 1$, z' denotes the transpose of z .

The rest of the paper is arranged as follows. Section 2 presents a class of constant stepsize algorithms. It needs to be pointed out that in our algorithms, we are not using discretization of the switching diffusion processes, but rather we construct a sequence of iterates to approximate the Brownian motion. In addition, we make another approximation of the transition probability matrix $\exp(\varepsilon Q(x))$ using a truncated Taylor expansion, which is more convenient for numerical solutions and Monte Carlo methods, and simplifies further the computation. Section 3 is devoted to the convergence of the algorithm. Weak convergence methods are used to establish that a sequence of suitable interpolations of the iterates converges weakly

to the solution of a martingale problem with a desired generator. Hence the limit is the solution of the system of switching diffusions. Section 4 presents a couple of computational examples. Also provided in this section is a decreasing stepsize algorithm. Section 5 makes a few more remarks to conclude the paper.

2 Numerical Methods

To approximate the r -dimensional standard Brownian motion $w(\cdot)$, we use a sequence of independent and identically distributed Gaussian random variables $\{\xi_n\}$ for simplicity. To approximate the solution of (1.1), we propose the following algorithm

$$x_{n+1} = x_n + \varepsilon f(x_n, \alpha_n) + \sqrt{\varepsilon} \sigma(x_n, \alpha_n) \xi_n. \quad (2.1)$$

We would like to have α_n be a discrete-time stochastic process that approximates $\alpha(t)$ in an appropriate sense. It is natural that α_n has a transition probability matrix $\exp(Q(x)\varepsilon)$ when $x_{n-1} = x$. It is easily seen that the transition matrix may be approximated further by $I + \varepsilon Q(x) + O(\varepsilon^2)$ by virtue of the boundedness and the continuity of $Q(\cdot)$. Based on this observation, in what follows, we discard the $O(\varepsilon^2)$ term and simply use $I + \varepsilon Q(x)$ for the transition matrix for α_n when $x_{n-1} = x$. We put what we said above into the following assumption.

(A3) In (2.1), for each n , when $x_{n-1} = x$, α_n has the transition matrix $I + \varepsilon Q(x)$, and $\{\xi_n\}$ is a sequence of independent and identically distributed random variables with normal distribution such that ξ_n is independent of the σ -algebra \mathcal{G}_n generated by $\{x_k, \alpha_k : k \leq n\}$, and that $E\xi_n = 0$ and $E\xi_n \xi_n' = I$.

Remark 2.1. One of the features of (2.1) is that it is easily implementable. In lieu of discretizing a Brownian motion, we generate a sequence of independent and identically distributed random variables with normal distribution to approximate the Brownian motion. This facilitates the computational task. In addition, instead of using transition matrix $\exp(\varepsilon Q(x))$ for a fixed x , we use another fold of approximation $I + \varepsilon Q(x)$ based on a truncated Taylor series. All of these stem from consideration of numerical computation and Monte Carlo implementation.

Since a Gaussian random variable is completely determined by its mean and covariance, we know that $E|\xi_n|^p < \infty$ for any $p \geq 2$. Finally, we note that in analyzing the algorithm, we could deal with correlated sequence $\{\xi_n\}$ as long as it has 0 mean and $E|\xi_n|^{2+\Delta} < \infty$ for some $\Delta > 0$. Nevertheless, from a computational point of view, the i.i.d. sequence appears to be most convenient from an implementation point of view.

3 Convergence of the Algorithm

3.1 Preliminary Estimates

We first obtain an estimate on the p th moment of $\{x_n\}$. This is stated as follows.

Lemma 3.1. *Under (A1), (A2) and (A3), for any fixed $p \geq 2$ and $T > 0$,*

$$\sup_{0 \leq n \leq T/\varepsilon} E|x_n|^p \leq (|x_0|^p + KT) \exp(KT) < \infty. \quad (3.1)$$

Remark 3.2. Throughout this paper, we assume the stepsize $0 < \varepsilon < 1$. Note that in Lemma 3.1, by T/ε , we mean the integer part of T/ε , i.e., $\lfloor T/\varepsilon \rfloor$. However, for simplicity, we will not use the floor function notation in what follows.

Note that in the proof of Lemma 3.1, $U(x)$ is a Liapunov type function. The technique used is standard in stochastic approximation [11, Chapter 5]. If correlated random sequences are treated, we can use a perturbed Liapunov function technique [11, Section 4.5, p.112].

Proof of Lemma 3.1. Define $U(x) = |x|^p$ and use E_n to denote the conditional expectation with respect to the σ -algebra \mathcal{G}_n , where \mathcal{G}_n was given in (A3). Note that $E_n \sigma(x_n, \alpha_n) \xi_n = \sigma(x_n, \alpha_n) E_n \xi_n = 0$ and that $E_n |\sigma(x_n, \alpha_n)|^2 |\xi_n|^2 = |\sigma(x_n, \alpha_n)|^2 E_n |\xi_n|^2 \leq K |\sigma(x_n, \alpha_n)|^2$, where K is a generic positive constant. Thus

$$\begin{aligned} E_n U(x_{n+1}) - U(x_n) &= E_n \nabla U'(x_n) [x_{n+1} - x_n] + E_n (x_{n+1} - x_n)' \nabla^2 U(x_n^+) (x_{n+1} - x_n) \\ &\leq \varepsilon \nabla U'(x_n) f(x_n, \alpha_n) + K |x_n|^{p-2} E_n |x_{n+1} - x_n|^2 \\ &\leq \varepsilon K |x_n|^{p-1} (1 + |x_n|) + K \varepsilon |x_n|^{p-2} (1 + |x_n|^2) + K \varepsilon^2 |x_n|^{p-2} (1 + |x_n|^2) \\ &\leq K \varepsilon (1 + |x_n|^p), \end{aligned} \quad (3.2)$$

where ∇U and $\nabla^2 U$ denotes the gradient and Hessian of U w.r.t. to x , and x_n^+ denotes a vector on the line segment joining x_n and x_{n+1} . Note that in the last line of (3.2), we have used the linear growth in x for both $f(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$. Since $U(x_n) = |x_n|^p$, we obtain $E_n |x_{n+1}|^p \leq |x_n|^p + K \varepsilon + K \varepsilon |x_n|^p$. Taking the expectation on both sides and iterating on the resulting recursion, we obtain

$$E|x_{n+1}|^p \leq |x_0|^p + K \varepsilon n + K \varepsilon \sum_{k=0}^n E|x_k|^p.$$

An application of the Gronwall's inequality yields that $E|x_{n+1}|^p \leq (|x_0|^p + KT) \exp(KT)$ as desired. \square

In view of the estimate above, $\{x_n : 0 \leq n \leq T/\varepsilon\}$ is tight in \mathbb{R}^r by means of the well-known Tchebyshev's inequality. That is, for each $\eta > 0$, there is a K_η satisfying $K_\eta > \sqrt{(1/\eta)}$ such that

$$P(|x_n| > K_\eta) \leq \frac{\sup_{0 \leq n \leq T/\varepsilon} E|x_n|^2}{K_\eta^2} \leq K\eta.$$

This indicates that the sequence of iterates is “mass preserving” or no probability is lost. To proceed, take continuous-time interpolations defined by

$$x^\varepsilon(t) = x_n, \quad \alpha^\varepsilon(t) = \alpha_n, \quad \text{for } t \in [n\varepsilon, n\varepsilon + \varepsilon). \quad (3.3)$$

We shall show that $x^\varepsilon(\cdot)$ and $\alpha^\varepsilon(\cdot)$ are tight in suitable function spaces.

Lemma 3.3. *Assume (A1)–(A3). Define*

$$\chi_n = (I_{\{\alpha_n=1\}}, \dots, I_{\{\alpha_n=m\}}) \in \mathbb{R}^{1 \times m} \quad \text{and} \quad \chi^\varepsilon(t) = \chi_n, \quad \text{for } t \in [\varepsilon n, \varepsilon n + \varepsilon). \quad (3.4)$$

Then for any $t, s > 0$,

$$E[\chi^\varepsilon(t+s) - \chi^\varepsilon(t) | \mathcal{F}_t^\varepsilon] = O(s), \quad (3.5)$$

where $\mathcal{F}_t^\varepsilon$ denotes the σ -algebra generated by $\{x^\varepsilon(u), \alpha^\varepsilon(u) : u \leq t\}$.

Proof. First note that by the boundedness and the continuity of $Q(\cdot)$, for each $i \in \mathcal{M}$,

$$\begin{aligned} & \sum_{j=1}^m E[I_{\{\alpha_{k+1}=j\}} - I_{\{\alpha_k=i\}} | \mathcal{G}_k] \\ &= \sum_{j=1}^m [(I + \varepsilon Q(x_k))_{ij} - \delta_{ij}] I_{\{\alpha_k=i\}} \\ &= \sum_{j=1}^m \varepsilon q_{ij}(x_k) I_{\{\alpha_k=i\}} = O(\varepsilon) I_{\{\alpha_k=i\}}, \end{aligned}$$

where $(I + \varepsilon Q(x_k))_{ij}$ denotes the ij th entry of $I + \varepsilon Q(x_k)$ and $\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$ It then follows that there is a random function $\tilde{g}(\cdot)$ such

$$\begin{aligned} & E\left[\sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} [\chi_{k+1} - \chi_k] \middle| \mathcal{F}_t^\varepsilon \right] \\ &= E\left[\sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} E[\chi_{k+1} - \chi_k | \mathcal{G}_k] \middle| \mathcal{F}_t^\varepsilon \right] \\ &= \tilde{g}(t+s-t) = \tilde{g}(s), \end{aligned} \quad (3.6)$$

and that $E\tilde{g}(s) = O(s)$. In the above, we have used the convention that t/ε and $(t+s)/\varepsilon$ denote the integer parts of t/ε and $(t+s)/\varepsilon$, respectively. Since

$$E\left[\chi^\varepsilon(t+s) - \chi^\varepsilon(t) - \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} [\chi_{k+1} - \chi_k] \middle| \mathcal{F}_t^\varepsilon\right] = 0,$$

it follows from (3.6),

$$E[\chi^\varepsilon(t+s) \middle| \mathcal{F}_t^\varepsilon] = \chi^\varepsilon(t) + \tilde{g}(s).$$

The desired result then follows. \square

Lemma 3.4. *Under the conditions of Lemma 3.3, $\{\alpha^\varepsilon(\cdot)\}$ is tight.*

Proof. For any $\eta > 0$, $t \geq 0$, $0 \leq s \leq \eta$, by virtue of Lemma 3.3,

$$\begin{aligned} & E[|\chi^\varepsilon(t+s) - \chi^\varepsilon(t)|^2 \middle| \mathcal{F}_t^\varepsilon] \\ &= E[\chi^\varepsilon(t+s)\chi^{\varepsilon,\prime}(t+s) - 2\chi^\varepsilon(t+s)\chi^{\varepsilon,\prime}(t) + \chi^\varepsilon(t)\chi^{\varepsilon,\prime}(t) \middle| \mathcal{F}_t^\varepsilon] \\ &= \sum_{i=1}^m E[I_{\{\alpha_{(t+s)/\varepsilon}=i\}} - 2I_{\{\alpha_{(t+s)/\varepsilon}=i\}}I_{\{\alpha_{t/\varepsilon}=i\}} + I_{\{\alpha_{t/\varepsilon}=i\}} \middle| \mathcal{F}_t^\varepsilon]. \end{aligned} \tag{3.7}$$

The estimates in Lemma 3.3 then imply that

$$\lim_{\eta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} E[E|\chi^\varepsilon(t+s) - \chi^\varepsilon(t)|^2 \middle| \mathcal{F}_t^\varepsilon] = 0.$$

The tightness criterion in [10, p. 47] yields that $\{\chi^\varepsilon(\cdot)\}$ is tight. Consequently, $\{\alpha^\varepsilon(\cdot)\}$ is tight. \square

Lemma 3.5. *Assume that the conditions of Lemma 3.4 are satisfied. Then $\{x^\varepsilon(\cdot)\}$ is tight in $D^r[0, \infty)$, the space of functions that are right continuous and have left limits, endowed with the Skorohod topology.*

Proof. For any $\eta > 0$, $t \geq 0$, $0 \leq s \leq \eta$, we have

$$\begin{aligned}
E|x^\varepsilon(t+s) - x^\varepsilon(t)|^2 &= E \left| \varepsilon \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} f(x_k, \alpha_k) + \sqrt{\varepsilon} \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} \sigma(x_k, \alpha_k) \xi_k \right|^2 \\
&\leq K\varepsilon^2 \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} (1 + E|x_k|^2) + K\varepsilon \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} E|\sigma(x_k, \alpha_k)|^2 E|\xi_k|^2 \\
&\leq K\varepsilon^2 \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} (1 + \sup_{t/\varepsilon \leq k \leq (t+s)/\varepsilon-1} E|x_k|^2) \\
&\quad + K\varepsilon \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} (1 + \sup_{t/\varepsilon \leq k \leq (t+s)/\varepsilon-1} E|x_k|^2) \\
&\leq O\left(\frac{t+s}{\varepsilon} - \frac{t}{\varepsilon}\right) O(\varepsilon) = O(s).
\end{aligned} \tag{3.8}$$

In the above, we have used Lemma 3.1 to ensure that $\sup_{t/\varepsilon \leq k \leq (t+s)/\varepsilon-1} E|x_k|^2 < \infty$. Therefore, (3.8) leads to

$$\lim_{\eta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} E|x^\varepsilon(t+s) - x^\varepsilon(t)|^2 = 0.$$

The tightness of $\{x^\varepsilon(\cdot)\}$ then follows from [10, p. 47]. \square

With Lemma 3.3, Lemma 3.4, and Lemma 3.5 at our hands, we obtain the following result.

Lemma 3.6 *Under assumptions (A1)–(A3), $\{x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot)\}$ is tight in $D([0, \infty) : \mathbb{R}^r \times \mathcal{M})$.*

3.2 Weak Convergence

Since $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ is tight, by Prohorov's theorem (see [3, 11]), we may select a convergent subsequence. For simplicity, still denote the subsequence by $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ with limit denoted by $(\tilde{x}(\cdot), \tilde{\alpha}(\cdot))$.

Theorem 3.7. *Assume (A1)–(A3). Then $\{x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot)\}$ converges weakly to $(x(\cdot), \alpha(\cdot))$, which is a process with generator given by (1.3).*

Proof. By Skorohod representation (see [3, 11]) without loss of generality and without changing notation, we may assume that $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ converges to $(\tilde{x}(\cdot), \tilde{\alpha}(\cdot))$ w.p.1, and the convergence is uniform on each bounded interval. We proceed to characterize the limit process.

Step 1: We first work with the marginal of the switching component, and characterize the limit of $\alpha^\varepsilon(\cdot)$. The weak convergence of $\alpha^\varepsilon(\cdot)$ to $\tilde{\alpha}(\cdot)$ yields that $\chi^\varepsilon(\cdot)$ converges a limit

process $\chi(\cdot)$ weakly, where $\chi^\varepsilon(\cdot)$ was defined in (3.4). For each $t > 0$ and $s > 0$, each positive integer κ , each $0 \leq t_\iota \leq t$ with $\iota \leq \kappa$, each bounded and continuous function $\rho_0(\cdot, i)$ for each $i \in \mathcal{M}$,

$$E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\chi^\varepsilon(t+s) - \chi^\varepsilon(t) - \sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} (\chi_{k+1} - \chi_k) \right] = 0. \quad (3.9)$$

The weak convergence of $\chi^\varepsilon(\cdot)$ to $\chi(\cdot)$ and the Skorohod representation imply that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) [\chi^\varepsilon(t+s) - \chi^\varepsilon(t)] \\ & = E\rho_0(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota), \iota \leq \kappa) [\chi(t+s) - \chi(t)]. \end{aligned}$$

Pick out a sequence $\{n_\varepsilon\}$ of nonnegative real numbers such that $n_\varepsilon \rightarrow \infty$ as $\varepsilon \rightarrow 0$ but $\delta_\varepsilon = \varepsilon n_\varepsilon \rightarrow 0$.

Let Ξ_l^ε be the set of indices

$$\Xi_l^\varepsilon = \{k : ln_\varepsilon \leq k \leq ln_\varepsilon + n_\varepsilon - 1\}. \quad (3.10)$$

Then the continuity and boundedness of $Q(\cdot)$ imply

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} (\chi_{k+1} - \chi_k) \right] \\ & = \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{k=t/\varepsilon}^{(t+s)/\varepsilon-1} (E(\chi_{k+1} | \mathcal{G}_k) - \chi_k) \right] \\ & = \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \sum_{k \in \Xi_l^\varepsilon} \chi_k (I + \varepsilon Q(x_k) - I) \right] \\ & = \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \delta_\varepsilon \frac{1}{n_\varepsilon} \sum_{k \in \Xi_l^\varepsilon} \chi_k Q(x_{ln_\varepsilon}) \right]. \end{aligned} \quad (3.11)$$

Note that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \delta_\varepsilon \frac{1}{n_\varepsilon} \sum_{k \in \Xi_l^\varepsilon} [\chi_k - \chi_{ln_\varepsilon}] Q(x_{ln_\varepsilon}) \right] \\ & = \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \delta_\varepsilon \frac{1}{n_\varepsilon} \sum_{k \in \Xi_l^\varepsilon} E[\chi_k - \chi_{ln_\varepsilon} | \mathcal{G}_{ln_\varepsilon}] Q(x_{ln_\varepsilon}) \right] \\ & = \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \delta_\varepsilon \frac{1}{n_\varepsilon} \sum_{k \in \Xi_l^\varepsilon} E\chi_{ln_\varepsilon} [(I + \varepsilon Q(x_{ln_\varepsilon}))^{k-ln_\varepsilon} - I] Q(x_{ln_\varepsilon}) \right] \\ & = 0. \end{aligned} \quad (3.12)$$

Therefore,

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \delta_\varepsilon \frac{1}{n_\varepsilon} \sum_{k \in \Xi_\varepsilon^i} \chi_k Q(x_{ln_\varepsilon}) \right] \\
&= \lim_{\varepsilon \rightarrow 0} E\rho_0(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota), \iota \leq \kappa) \left[\sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon} \delta_\varepsilon \chi_{ln_\varepsilon} Q(x_{ln_\varepsilon}) \right] \\
&= E\rho_0(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota), \iota \leq \kappa) \int_t^{t+s} \chi(u) Q(\tilde{x}(u)) du.
\end{aligned} \tag{3.13}$$

Moreover, the limit does not depend on the chosen subsequence. Thus,

$$E\rho_0(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota), \iota \leq \kappa) \left[\chi(t+s) - \chi(t) - \int_t^{t+s} \chi(u) Q(\tilde{x}(u)) du \right] = 0. \tag{3.14}$$

Therefore, the limit process $\tilde{\alpha}(\cdot)$ has a generator $Q(\tilde{x}(\cdot))$.

Step 2: For t, s, κ, t_ι as chosen before, for each bounded and continuous function $\rho(\cdot, i)$, and for each twice continuously differentiable function with compact support $h(\cdot, i)$ with $i \in \mathcal{M}$, we shall show that

$$E\rho(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota); \iota \leq \kappa) [h(\tilde{x}(t+s), \tilde{\alpha}(t+s)) - h(\tilde{x}(t), \tilde{\alpha}(t)) - \int_t^{t+s} \mathcal{L}h(\tilde{x}(u), \tilde{\alpha}(u)) du] = 0. \tag{3.15}$$

This yields that

$$h(\tilde{x}(t), \tilde{\alpha}(t)) - \int_0^t \mathcal{L}h(\tilde{x}(u), \tilde{\alpha}(u)) du \text{ is a continuous-time martingale,}$$

which in turn implies that $(\tilde{x}(\cdot), \tilde{\alpha}(\cdot))$ is a solution of the martingale problem with operator \mathcal{L} defined in (1.3).

To establish the desired result, we work with the sequence $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$. Again, we use the sequence $\{n_\varepsilon\}$ as in Step 1. By virtue of the weak convergence and the Skorohod representation, it is readily seen that

$$\begin{aligned}
& E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) [h(x^\varepsilon(t+s), \alpha^\varepsilon(t+s)) - h(x^\varepsilon(t), \alpha^\varepsilon(t))] \\
& \rightarrow E\rho(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota); \iota \leq \kappa) [h(\tilde{x}(t+s), \tilde{\alpha}(t+s)) - h(\tilde{x}(t), \tilde{\alpha}(t))] \text{ as } \varepsilon \rightarrow 0.
\end{aligned} \tag{3.16}$$

On the other hand, direct calculation shows that

$$\begin{aligned}
& E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) [h(x^\varepsilon(t+s), \alpha^\varepsilon(t+s)) - h(x^\varepsilon(t), \alpha^\varepsilon(t))] \\
&= E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} \left[h(x_{ln_\varepsilon+n_\varepsilon}, \alpha_{ln_\varepsilon+n_\varepsilon}) - h(x_{ln_\varepsilon+n_\varepsilon}, \alpha_{ln_\varepsilon}) \right] \right. \\
& \quad \left. + [h(x_{ln_\varepsilon+n_\varepsilon}, \alpha_{ln_\varepsilon}) - h(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon})] \right\}.
\end{aligned} \tag{3.17}$$

Step 3: Still use the notation Ξ_ℓ^ε defined in (3.10). For the terms on the last line of (3.17), we have

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} [h(x_{ln_\varepsilon+n_\varepsilon}, \alpha_{ln_\varepsilon}) - h(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon})] \\
&= \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} \left[\varepsilon \nabla h'(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) \sum_{k \in \Xi_i^\varepsilon} f(x_k, \alpha_{ln_\varepsilon}) \right. \right. \\
&\quad \left. \left. + \frac{\varepsilon}{2} \sum_{k \in \Xi_i^\varepsilon} \text{tr}[\nabla^2 h(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) \sigma(x_k, \alpha_{ln_\varepsilon}) \sigma'(x_k, \alpha_{ln_\varepsilon})] \right] \right\}.
\end{aligned} \tag{3.18}$$

By the continuity of $f(\cdot, i)$ for each $i \in \mathcal{M}$ and the choice of n_ε ,

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} \delta_\varepsilon \nabla h'(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) \frac{1}{n_\varepsilon} \sum_{k \in \Xi_i^\varepsilon} [f(x_k, \alpha_{ln_\varepsilon}) - f(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon})] \right\} \\
&= 0.
\end{aligned}$$

Thus, in evaluating the limit, $f(x_k, \alpha_{ln_\varepsilon})$ can be replaced by $f(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon})$.

The choice of n_ε implies that $\varepsilon ln_\varepsilon \rightarrow u$ as $\varepsilon \rightarrow 0$ yielding $\varepsilon k \rightarrow u$ for all $ln_\varepsilon \leq k \leq ln_\varepsilon + n_\varepsilon$. Consequently, by the weak convergence and the Skorohod representation,

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} \varepsilon \nabla h'(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) \sum_{k \in \Xi_i^\varepsilon} f(x_k, \alpha_{ln_\varepsilon}) \right\} \\
&= \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} \delta_\varepsilon \nabla h'(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) \frac{1}{n_\varepsilon} \sum_{k \in \Xi_i^\varepsilon} f(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) \right\} \\
&= \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} \delta_\varepsilon \nabla h'(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) f(x^\varepsilon(l\delta_\varepsilon), \alpha^\varepsilon(l\delta_\varepsilon)) \right\} \\
&= E\rho(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota); \iota \leq \kappa) \left\{ \int_t^{t+s} \nabla h'(\tilde{x}(u), \tilde{\alpha}(u)) f(\tilde{x}(u), \tilde{\alpha}(u)) du \right\}.
\end{aligned} \tag{3.19}$$

In the above, treating such terms as $f(x^\varepsilon(l\delta_\varepsilon), \alpha^\varepsilon(l\delta_\varepsilon))$, we can approximate $x^\varepsilon(\cdot)$ by a process taking finitely many values using a standard approximation argument (see for example, [11, p. 169] for more details).

Similar to (3.19), we also obtain

$$\begin{aligned}
& \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \frac{\varepsilon}{2} \sum_{ln_\varepsilon=t/\varepsilon}^{(t+s)/\varepsilon-1} \sum_{k \in \Xi_i^\varepsilon} \text{tr}[\nabla^2 h(x_{ln_\varepsilon}, \alpha_{ln_\varepsilon}) \sigma(x_k, \alpha_{ln_\varepsilon}) \sigma'(x_k, \alpha_{ln_\varepsilon})] \right\} \\
&= E\rho(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota); \iota \leq \kappa) \left\{ \int_t^{t+s} \frac{1}{2} \text{tr}[\nabla^2 h(\tilde{x}(u), \tilde{\alpha}(u)) \sigma(\tilde{x}(u), \tilde{\alpha}(u)) \sigma'(\tilde{x}(u), \tilde{\alpha}(u))] du \right\}.
\end{aligned} \tag{3.20}$$

Step 4: We next examine the terms on the next to the last line of (3.17). First, again using the continuity, the weak convergence, and the Skorohod representation, it can be shown that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\nu), \alpha^\varepsilon(t_\nu); \nu \leq \kappa) \left\{ \sum_{l n_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1} [h(x_{l n_\varepsilon + n_\varepsilon}, \alpha_{l n_\varepsilon + n_\varepsilon}) - h(x_{l n_\varepsilon + n_\varepsilon}, \alpha_{l n_\varepsilon})] \right\} \\ &= \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\nu), \alpha^\varepsilon(t_\nu); \nu \leq \kappa) \left\{ \sum_{l n_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1} [h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon + n_\varepsilon}) - h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon})] \right\}. \end{aligned} \quad (3.21)$$

That is, owing to the choice of $\{n_\varepsilon\}$ and the continuity of $h(\cdot, i)$, the term

$$h(x_{l n_\varepsilon + n_\varepsilon}, \alpha_{l n_\varepsilon + n_\varepsilon}) - h(x_{l n_\varepsilon + n_\varepsilon}, \alpha_{l n_\varepsilon})$$

in the next to the last line of (3.17) can be replaced by

$$h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon + n_\varepsilon}) - h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon})$$

as far as asymptotic analysis is concerned. It follows that

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\nu), \alpha^\varepsilon(t_\nu); \nu \leq \kappa) \left\{ \sum_{l n_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1} [h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon + n_\varepsilon}) - h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon})] \right\} \\ &= \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\nu), \alpha^\varepsilon(t_\nu); \nu \leq \kappa) \left\{ \sum_{l n_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1} \sum_{k \in \Xi_i^\varepsilon} [h(x_{l n_\varepsilon}, \alpha_{k+1}) - h(x_{l n_\varepsilon}, \alpha_k)] \right\} \\ &= \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\nu), \alpha^\varepsilon(t_\nu); \nu \leq \kappa) \left\{ \sum_{l n_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon - 1} \sum_{k \in \Xi_i^\varepsilon} \sum_{i_1=1}^m \sum_{i_1=1}^m E \left[h(x_{l n_\varepsilon}, i) I_{\{\alpha_{k+1}=i\}} \right. \right. \\ & \qquad \qquad \qquad \left. \left. - h(x_{l n_\varepsilon}, i_1) I_{\{\alpha_k=i_1\}} \right] \Big| \mathcal{G}_k \right\}. \end{aligned} \quad (3.22)$$

Note that for $k \geq l n_\varepsilon$,

$$\begin{aligned} & E[[h(x_{l n_\varepsilon}, i) I_{\{\alpha_{k+1}=i\}} - h(x_{l n_\varepsilon}, i_1) I_{\{\alpha_k=i_1\}}] \Big| \mathcal{G}_k] \\ &= [h(x_{l n_\varepsilon}, i) P(\alpha_{k+1} = i \mid \mathcal{G}_k, \alpha_k = i_1) - h(x_{l n_\varepsilon}, i_1)] I_{\{\alpha_k=i_1\}} \\ &= [h(x_{l n_\varepsilon}, i) (\delta_{i_1 i} + \varepsilon q_{i_1 i}(x_k)) - h(x_{l n_\varepsilon}, i_1)] I_{\{\alpha_k=i_1\}} \\ &= \varepsilon h(x_{l n_\varepsilon}, i) q_{i_1 i}(x_k) I_{\{\alpha_k=i_1\}}. \end{aligned} \quad (3.23)$$

Using (3.23) in (3.22) and noting the continuity and boundedness of $Q(\cdot)$, we can replace $q_{i_1 i}(x_k)$ by $q_{i_1 i}(x_{l n_\varepsilon})$ yielding the same limit. Then as in (3.12) and (3.13), replace $I_{\{\alpha_k=i_1\}}$

by $I_{\{\alpha^\varepsilon(\varepsilon l n_\varepsilon)=i_1\}}$, again yielding the same limit. Thus, we have

$$\begin{aligned} & \lim_{\varepsilon \rightarrow 0} E\rho(x^\varepsilon(t_\iota), \alpha^\varepsilon(t_\iota); \iota \leq \kappa) \left\{ \sum_{l n_\varepsilon = t/\varepsilon}^{(t+s)/\varepsilon-1} [h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon + n_\varepsilon}) - h(x_{l n_\varepsilon}, \alpha_{l n_\varepsilon})] \right\} \\ &= E\rho(\tilde{x}(t_\iota), \tilde{\alpha}(t_\iota); \iota \leq \kappa) \left\{ \int_t^{t+s} Q(\tilde{x}(u)) h(\tilde{x}(u), \cdot) (\tilde{\alpha}(u)) du \right\}, \end{aligned} \quad (3.24)$$

where

$$Q(x)h(x, \cdot)(i_1) = \sum_{i=1}^m q_{i_1 i}(x)h(x, i) = \sum_{i \neq i_1} q_{i_1 i}(x)(h(x, i) - h(x, i_1)).$$

Step 5: Combining Steps 1–4, we arrive at $(\tilde{x}(\cdot), \tilde{\alpha}(\cdot))$, the weak limit of $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ is a solution of the martingale problem with operator \mathcal{L} defined in (1.3). Using characteristic functions, we can show as in [22, Lemma 7.18], $(x(\cdot), \alpha(\cdot))$ the solution of the martingale problem with operator \mathcal{L} , is unique in the sense of in distribution. Thus $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ converges to $(x(\cdot), \alpha(\cdot))$ as desired, which concludes the proof of the theorem. \square

3.3 Existence and Uniqueness of the Solutions

We have established convergence of $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ to $(x(\cdot), \alpha(\cdot))$ in the weak sense. In this section, we strengthen the approximation result. We shall show that the weak limit $(x(\cdot), \alpha(\cdot))$ is, in fact, the strong solution of Eq. (1.1). Define the continuous approximation of $x^\varepsilon(t)$, i.e.,

$$\bar{X}^\varepsilon(t) = x_0 + \int_0^t f(x^\varepsilon(s), \alpha^\varepsilon(s)) ds + \int_0^t \sigma(x^\varepsilon(s), \alpha^\varepsilon(s)) dw(s). \quad (3.25)$$

Note that $\bar{X}^\varepsilon(\varepsilon k) = x^\varepsilon(\varepsilon k) = x_k$ and $\alpha^\varepsilon(\varepsilon k) = \alpha_k$. That is, $\bar{X}^\varepsilon(t)$ and $x^\varepsilon(t)$ coincide at the gridpoints $k = t/\varepsilon$ (again, t/ε means $\lfloor t/\varepsilon \rfloor$, the integer part of t/ε).

Theorem 3.8. *Assume (A1)–(A3). Then there exists a unique solution to Eq. (1.1), which is the limit given by the approximation 3.25.*

Proof. Let ε and η be two different stepsizes (both being sufficiently small), by (3.25), we have

$$\begin{aligned} & |\bar{X}^\varepsilon(t) - \bar{X}^\eta(t)|^2 \\ & \leq 2 \left| \int_0^t (f(x^\varepsilon(s), \alpha^\varepsilon(s)) - f(x^\eta(s), \alpha^\eta(s))) ds \right|^2 \\ & \quad + 2 \left| \int_0^t (\sigma(x^\varepsilon(s), \alpha^\varepsilon(s)) - \sigma(x^\eta(s), \alpha^\eta(s))) dw(s) \right|^2. \end{aligned} \quad (3.26)$$

By Hölder's inequality and Doob's martingale inequality, we have

$$\begin{aligned}
& E \left[\sup_{0 \leq t \leq T} |\bar{X}^\varepsilon(t) - \bar{X}^\eta(t)|^2 \right] \\
& \leq 2T \left\{ E \int_0^T |(f(x^\varepsilon(s), \alpha^\varepsilon(s)) - f(x^\eta(s), \alpha^\eta(s)))|^2 ds \right. \\
& \quad \left. + 8E \int_0^T |\sigma(x^\varepsilon(s), \alpha^\varepsilon(s)) - \sigma(x^\eta(s), \alpha^\eta(s))|^2 ds \right\}. \tag{3.27}
\end{aligned}$$

Using condition (A2), the Hölder inequality and Lemma 3.1, we obtain

$$\begin{aligned}
& E \int_0^T |(f(x^\varepsilon(s), \alpha^\varepsilon(s)) - f(x^\eta(s), \alpha^\eta(s)))|^2 ds \\
& \leq E \int_0^T |(f(x^\varepsilon(s), \alpha^\varepsilon(s)) - f(x^\varepsilon(s), \alpha(s)))|^2 ds \\
& \quad + E \int_0^T |(f(x^\eta(s), \alpha^\eta(s)) - f(x^\eta(s), \alpha(s)))|^2 ds \\
& \quad + E \int_0^T |(f(x^\varepsilon(s), \alpha(s)) - f(x^\eta(s), \alpha(s)))|^2 ds \\
& \leq 2E \int_0^T [|f(x^\varepsilon(s), \alpha^\varepsilon(s))|^2 + |f(x^\varepsilon(s), \alpha(s))|^2] I_{\{\alpha^\varepsilon(s) \neq \alpha(s)\}} ds \\
& \quad + 2E \int_0^T [|f(x^\eta(s), \alpha^\eta(s))|^2 + |f(x^\eta(s), \alpha(s))|^2] I_{\{\alpha^\eta(s) \neq \alpha(s)\}} ds \\
& \quad + KE \int_0^T |x^\varepsilon(s) - x^\eta(s)|^2 ds \\
& \leq K \int_0^T E [(1 + |x^\varepsilon(s)|^2) I_{\{\alpha^\varepsilon(s) \neq \alpha(s)\}}] ds + K \int_0^T E [(1 + |x^\eta(s)|^2) I_{\{\alpha^\eta(s) \neq \alpha(s)\}}] ds \\
& \quad + KE \int_0^T |x^\varepsilon(s) - x^\eta(s)|^2 ds \\
& \leq K \int_0^T (E(1 + |x^\varepsilon(s)|^2)^2)^{1/2} (EI_{\{\alpha^\varepsilon(s) \neq \alpha(s)\}})^{1/2} ds \\
& \quad + K \int_0^T (E(1 + |x^\eta(s)|^2)^2)^{1/2} (EI_{\{\alpha^\eta(s) \neq \alpha(s)\}})^{1/2} ds \\
& \quad + KE \int_0^T |x^\varepsilon(s) - x^\eta(s)|^2 ds \\
& \leq K \int_0^T ([EI_{\{\alpha^\varepsilon(s) \neq \alpha(s)\}}]^{1/2} + E[I_{\{\alpha^\eta(s) \neq \alpha(s)\}}]^{1/2}) ds + KE \int_0^T |x^\varepsilon(s) - x^\eta(s)|^2 ds, \tag{3.28}
\end{aligned}$$

where K is still a generic constant and it is independent of ε and η . Obviously, if $f(\cdot, \cdot)$ is replaced by $\sigma(\cdot, \cdot)$, (3.28) still holds. On the other hand, by Lemma 3.1, we have

$$E|\bar{X}^\varepsilon(s) - x^\varepsilon(s)|^2 \leq KE \left[(1 + |x_{s/\varepsilon}|^2)(\varepsilon + |w(s) - w(\varepsilon(s/\varepsilon))|^2) \right] \leq K\varepsilon. \tag{3.29}$$

This together with (3.28) yields

$$E \left[\sup_{0 \leq t \leq T} |\bar{X}^\varepsilon(t) - \bar{X}^\eta(t)|^2 \right] \leq K \int_0^T ([EI_{\{\alpha^\varepsilon(s) \neq \alpha(s)\}}]^{1/2} + E[I_{\{\alpha^\eta(s) \neq \alpha(s)\}}]^{1/2}) ds + K(\varepsilon \vee \eta) + KE \int_0^T |\bar{X}^\varepsilon(s) - \bar{X}^\eta(s)|^2 ds. \quad (3.30)$$

By the well-known Grownwall inequality,

$$E \left[\sup_{0 \leq t \leq T} |\bar{X}^\varepsilon(t) - \bar{X}^\eta(t)|^2 \right] \leq Ke^{KT} \left[\int_0^T ([EI_{\{\alpha^\varepsilon(s) \neq \alpha(s)\}}]^{1/2} + E[I_{\{\alpha^\eta(s) \neq \alpha(s)\}}]^{1/2}) ds + \varepsilon \vee \eta \right]. \quad (3.31)$$

By Theorem 3.7, we have

$$\lim_{\varepsilon \rightarrow 0} E[I_{\{\alpha^\varepsilon(s) \neq \alpha(s)\}}] = 0 \text{ and } \lim_{\eta \rightarrow 0} E[I_{\{\alpha^\eta(s) \neq \alpha(s)\}}] = 0.$$

We hence obtain from (3.31) that

$$E \left[\sup_{0 \leq t \leq T} |\bar{X}^\varepsilon(t) - \bar{X}^\eta(t)|^2 \right] \rightarrow 0, \text{ as } \varepsilon \text{ and } \eta \rightarrow 0. \quad (3.32)$$

Let $\{\varepsilon_l\}_{l=1}^\infty$ be a nonrandom sequence of positive numbers such that ε_l is decreasing and $\lim_{l \rightarrow \infty} \varepsilon_l = 0$. For any positive integer ℓ , by (3.32), we have

$$\lim_{l \rightarrow \infty} E \left[\sup_{0 \leq t \leq T} |\bar{X}^{\varepsilon_l}(t) - \bar{X}^{\varepsilon_{l+\ell}}(t)|^2 \right] = 0. \quad (3.33)$$

This implies that $\{\bar{X}^{\varepsilon_l}(t) : 0 \leq t \leq T\}$ is a Cauchy sequence in $L^2(\Omega; C([0, T], \mathbb{R}^n))$. Recalling that the weak limit $(x(t), \alpha(t))$ of $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ is unique in the sense of in distribution and noting (3.29), we see

$$\lim_{l \rightarrow \infty} E \left[\sup_{0 \leq t \leq T} |\bar{X}^{\varepsilon_l}(t) - x(t)|^2 \right] = 0. \quad (3.34)$$

In the same way as (3.28) was proved, we can show that

$$\lim_{l \rightarrow \infty} E \int_0^T |(f(\bar{X}^{\varepsilon_l}(s), \alpha^{\varepsilon_l}(s)) - f(x(s), \alpha(t)))|^2 ds = 0, \quad (3.35)$$

and

$$\lim_{l \rightarrow \infty} E \left| \int_0^T (\sigma(\bar{X}^{\varepsilon_l}(s), \alpha^{\varepsilon_l}(s)) - \sigma(x(s), \alpha(t))) dw(s) \right|^2 = 0. \quad (3.36)$$

These, together with (3.34), yield that $(x(t), \alpha(t))$ is a strong solution to equation (1.1). The proof is therefore completed. \square

Remark 3.9. The Lipschitz condition may be further relaxed to local Lipschitz condition. Modifying the argument in Theorem 3.7, we can show that replacing the Lipschitz condition in (A2) (a) by local Lipschitz condition of the functions, the weak convergence result still holds. This, together with the standard truncation method of [12], leads to the following theorem.

Theorem 3.10. *Assume that Assumptions (A1) and (A2) (a) hold, but the Lipschitz condition (A2) (b) is replaced by the following local Lipschitz condition: For every integer $j \geq 1$, there exists a positive constant M_j such that for all $t \in [0, T]$, $i \in \mathcal{M}$ and all $x, y \in \mathbb{R}^n$ with $|x| \vee |y| \leq M_j$,*

$$|f(x, t, i) - f(y, t, i)| \vee |\sigma(x, t, i) - \sigma(y, t, i)| \leq M_j |x - y|. \quad (3.37)$$

Then there exists a unique solution $(x(t), \alpha(t))$ to equation (1.1).

In view of Theorem 3.7 and Theorem 3.10, we have obtained the following strong convergence result.

Corollary 3.11. *Under the conditions of either Theorem 3.8 or Theorem 3.10, the sequence $(x^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ converges to $(x(\cdot), \alpha(\cdot))$ in that*

$$E \left[\sup_{0 \leq t \leq T} |x^\varepsilon(t) - x(t)|^2 \right] \rightarrow 0 \quad \text{as } \varepsilon \rightarrow 0. \quad (3.38)$$

4 Ramifications

This section deals with some ramifications. First, decreasing stepsize algorithms are presented. Then a couple of examples are provided to illustrate the use of the numerical approximation schemes. We note that switching diffusions, for example, jump linear systems are used widely in control and systems theory [7, 15, 19]. Such systems also appear in financial engineering for modeling a security asset [24]. So far, the modeling is mostly concerned with Markov modulated switching diffusions in which the switching processing is a Markov chain independent of the Brownian motion. It is clear for more complex situation, one would want to model the switching process by means of state-dependent process. For example, a more realistic model for treating equity asset would be: Not only do the appreciation rate and volatility depend on the switching process (e.g., bull or bear market), but also the switching component depends on the equity asset price. This is surely a better remedy, but it would also make the problem too complex to solve. The proposed numerical algorithm is an effort towards the solution of such problem in this direction.

4.1 Algorithms with Decreasing Stepsizes

So far the development is based on using constant stepsize algorithms. In the literature of numerical solutions of stochastic differential equations, decreasing stepsize algorithms are used most often. Here, we demonstrate that to approximate (1.1), we could also use a decreasing stepsize algorithm of the form

$$x_{n+1} = x_n + \varepsilon_n f(x_n, \alpha_n) + \sqrt{\varepsilon_n} \sigma(x_n, \alpha_n) \xi_n. \quad (4.1)$$

Compared with (2.1), for $x_{n-1} = x$, α_n is a finite state process with transition matrix $I + \varepsilon_n Q(x)$, Instead of (A3), we assume the following condition.

(A4) In (4.1), ε_n is a sequence of decreasing stepsizes satisfying $\varepsilon_n \rightarrow 0$ as $n \rightarrow \infty$ and $\sum_n \varepsilon_n = \infty$. The $\{\xi_n\}$ is a sequence of independent and identically distributed normal random variables such that ξ_n is independent of the σ -algebra \mathcal{G}_n generated by $\{x_k, \alpha_k : k \leq n\}$, and that $E\xi_n = 0$ and $E\xi_n \xi_n' = I$.

Define

$$t_n = \sum_{k=0}^{n-1} \varepsilon_k, \quad m(t) = \max\{n : t_n \leq t\},$$

and continuous-time interpolations

$$x^n(t) = x_n, \quad \alpha^n(t) = \alpha_n, \quad \text{for } t \in [t_n, t_{n+1}).$$

Using essentially the same approach as in the development of Theorem 3.7 together with the ideas from stochastic approximation [11, Chapters 5, 6, and 8], we obtain the following result.

Theorem 4.1. *Under (A1), (A2), and (A4), $(x^n(\cdot), \alpha^n(\cdot))$ converges to $(x(\cdot), \alpha(\cdot))$ weakly, which is a solution of the martingale problem with operator \mathcal{L} defined in (1.3).*

4.2 Examples

Here we present two examples for demonstration. It would be ideal if we could compare the numerical solutions using our algorithms with the analytic solutions. Unfortunately, due to the complexity of the x -dependent switching process, closed-form solutions are not available. We are thus contended with the numerical demonstrations.

Example 4.2. In this example, we display the sample paths of approximated solutions of (1.1) using both a constant stepsizes and a sequence of decreasing stepsizes. Assume that the state-dependent generator $Q(x)$ is given by

$$Q(x) = \begin{pmatrix} -5 \cos^2 x & 5 \cos^2 x \\ 10 \cos^2 x & -10 \cos^2 x \end{pmatrix}. \quad (4.2)$$

Because of the x -dependence in $Q(x)$ given by (4.2), and the different stepsize used, the displays will not be the same. However, it can be seen that with the same random seeds chosen, the sample paths display similar behavior. This suggests that the constant-stepsize approximation is a viable and easily implementable alternative as compared to the decreasing stepsize algorithms. In this example, we assume that the jump process has state space $\mathcal{M} = \{1, 2\}$, and the drift is a nonlinear function $f(\cdot, \cdot) : \mathbb{R} \times \{1, 2\} \mapsto \mathbb{R}$, where

$$f(x, 1) = 2 + \sin x, \quad f(x, 2) = 1 + \sin x \cos x.$$

The diffusion coefficients are given by

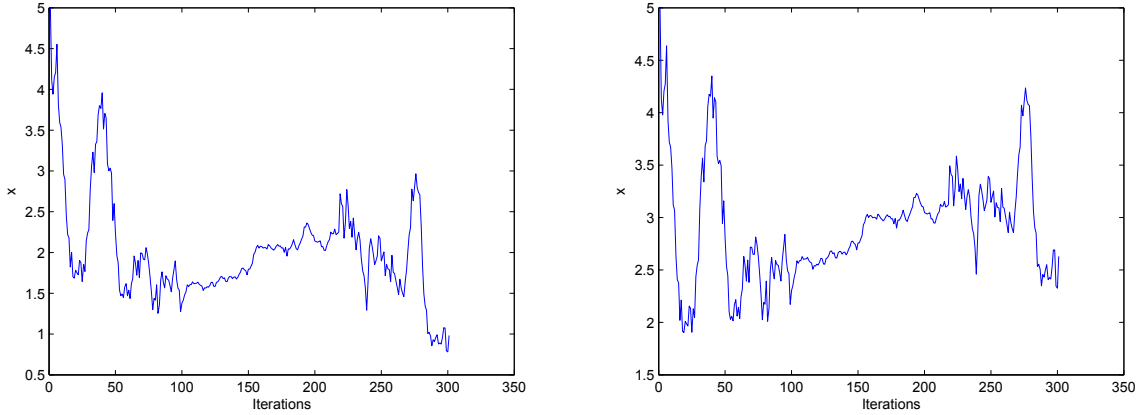
$$\sigma(x, 1) = 0.5x, \quad \sigma(x, 2) = 0.2x.$$

We specify the initial conditions as $x_0 = 5$ and $\alpha_0 = 1$, and use the constant stepsize $\varepsilon = 0.01$ and the decreasing stepsizes $\varepsilon_n = 1/(n+99)$, respectively. The sample paths of the computed iterates are displayed in Figure 1.

Example 4.3. In this example, we demonstrate the computational results of a process whose continuous component is two dimensional. For $x = (x^1, x^2)' \in \mathbb{R}^2$, we use the x -dependent generator $Q(x) = Q(x^1, x^2)$ given by

$$Q(x^1, x^2) = \begin{pmatrix} -5 \cos^2 x^1 - 2 \cos^2 x^2 & 5 \cos^2 x^1 + 2 \cos^2 x^2 \\ 10 \cos^2 x^1 + 2 \sin^2 x^2 & -10 \cos^2 x^1 - 2 \sin^2 x^2 \end{pmatrix}.$$

Use the same constant stepsize as in the Example 4.2 for each dimension and specify the initial data as $(x_0^1, x_0^2)' = (5, 5)'$, $\alpha_0 = 1$. The $\sigma(x^1, x^2, \alpha)$ matrix and $f(x^1, x^2, \alpha)$ are given



(a) A sample path of approximation using constant stepsize algorithm (b) A sample path of approximation using a decreasing stepsize algorithm.

Figure 1: Sample paths for constant-stepsize and decreasing-stepsize algorithms

as follows:

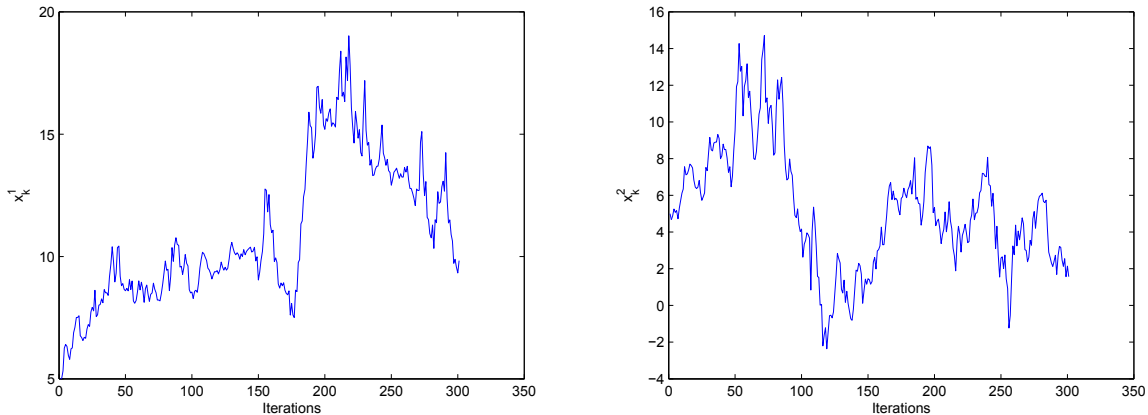
$$\begin{aligned} \sigma(x^1, x^2, 1) &= \begin{pmatrix} 0.5x^1 + 0.2x^2 & 0.003x^1 + 0.001x^2 \\ 0.007x^1 + 0.008x^2 & 0.47x^1 + 0.3x^2 \end{pmatrix}, \\ \sigma(x^1, x^2, 2) &= \begin{pmatrix} 0.2x^1 + 0.3x^2 & 0.001x^1 + 0.002x^2 \\ 0.008x^1 + 0.005x^2 & 1x^1 + 0.5x^2 \end{pmatrix}, \\ f(x^1, x^2, 1) &= \begin{pmatrix} 2 + \sin x^1 + \cos x^2 \\ 1 + \cos x^1 + \cos x^2 \end{pmatrix}. \end{aligned}$$

The sample paths are depicted in Figure 2.

For both of these examples, we have done a number of numerical experiments with different stepsize selections. They all produced similar sample path behavior as displayed above. For numerical experiment purpose, we have also tested different functions $f(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ as well. It is seen that our proposed algorithms are easily implementable and are suited for Monte Carlo studies.

5 Further Remarks

Approximation methods for switching diffusion processes with state-dependent switching have been developed in this paper. Our contributions are: First we presented a class of



(a) A sample path of the first component of the switching diffusion. (b) A sample path of the second component of the switching diffusion.

Figure 2: Sample paths of a 2-dimensional switching diffusion

constant stepsize algorithms, in which the state-dependent modulated switching process is approximated by a discrete-time switching process with state-dependent transition probabilities. Second, we proved the weak convergence of the proposed algorithm using a martingale problem formulation. We used continuity and localized analysis to overcome the difficulty of the state dependence. Third, we obtained existence and uniqueness of the switching diffusions. As a direct consequence, we obtain the strong convergence result in the sense of usual numerical solutions for SDE and established that the approximation sequence convergence strongly to the solution of the switching diffusion. Moreover, decreasing stepsize algorithms were developed as well. Finally, we provided examples to illustrate the utility of the approximation methods.

A number of problems deserve future study and investigation. Pursuing rates of convergence study is a worthwhile under-taking. Effort may also be devoted to numerical methods for regime-switching jump diffusions with state-dependent switching. Considering numerical methods for systems with delays and switching will be both interesting and important.

Acknowledgement. We thank the editors and reviewers for making many detailed comments and suggestions.

References

- [1] G.K. Basak, A. Bisi, and M.K. Ghosh, Stability of a random diffusion with linear drift, *J. Math. Anal. Appl.*, **202** (1996), 604–622.

- [2] K.L. Chung, *Markov Chains with Stationary Transition Probabilities*, 2nd Ed., Springer-Verlag, New York, NY, 1967.
- [3] S.N. Ethier and T.G. Kurtz, *Markov Processes: Characterization and Convergence*, J. Wiley, New York, NY, 1986.
- [4] D.J. Higham, X. Mao, and A.M. Stuart, Strong convergence of Euler-type methods for non-linear stochastic differential equations, *SIAM J. Numerical Anal.* **40** (2002), 1041-1063.
- [5] A.M. Il'in, R.Z. Khasminskii, and G. Yin, Singularly perturbed switching diffusions: rapid switchings and fast diffusions, *J. Optim. Theory Appl.*, **102** (1999), 555-591.
- [6] N. Ikeda and S. Watabe, *Stochastic Differential Equations and Diffusion Processes*, Springer, New York, 1981.
- [7] Y. Ji and H.J. Chizeck, Controllability, stabilizability and continuous-time Markovian jump linear quadratic control, *IEEE Trans. Automat. Control*, **35** (1990), 777-788.
- [8] R.Z. Khasminskii, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Alphen aan den Rijn, Netherlands, 1980.
- [9] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, 1992.
- [10] H.J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, with applications to Stochastic Systems Theory*, MIT Press, Cambridge, MA, 1984.
- [11] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd Ed., Springer, New York, 2003.
- [12] X. Mao, *Stochastic Differential Equations and Applications*, Horwood, 1997.
- [13] X. Mao, Stability of stochastic differential equations with Markovian switching, *Stochastic Process. Appl.*, **79** (1999), 45-67.
- [14] X. Mao and C. Yuan, *Stochastic Differential equations with Markovian Switching*, Imperial College Press, London, 2006.
- [15] M. Mariton, *Jump Linear Systems in Automatic Control*, Marcel Dekker, New York, 1990.
- [16] G.N. Milstein, *Numerical Integration of Stochastic Differential Equations*, Kluwer, New York, 1995.
- [17] S.P. Sethi and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser, Boston, 1994.
- [18] A.V. Skorohod, *Asymptotic Methods in the Theory of Stochastic Differential Equations*, American Mathematical Society, Providence, 1989.
- [19] D.D. Sworner and J.E. Boyd, *Estimation Problems in Hybrid Systems*, Cambridge University Press, 1999.
- [20] G. Yin and S. Dey, Weak convergence of hybrid filtering problems involving nearly completely decomposable hidden Markov chains, *SIAM J. Control Optim.*, **41** (2003), 1820-1842.
- [21] G. Yin, R.H. Liu, and Q. Zhang, Recursive algorithms for stock Liquidation: A stochastic optimization approach, *SIAM J. Optim.*, **13** (2002), 240-263.
- [22] G. Yin and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*, Springer-Verlag, New York, 1998.
- [23] C. Yuan and X. Mao, Asymptotic stability in distribution of stochastic differential equations with Markovian switching, *Stochastic Process Appl.*, **103** (2003), 277-291.
- [24] Q. Zhang, Stock trading: An optimal selling rule, *SIAM J. Control. Optim.*, **40**, (2001), 64-87.
- [25] C. Zhu and G. Yin, Asymptotic properties of hybrid diffusion systems, *SIAM J. Control Optim.*, **46** (2007), 1155-1179.