Properties of solutions of stochastic differential equations with continuous-state-dependent switching

G. Yin\textsuperscript{a,}* , C. Zhu\textsuperscript{b}

\textsuperscript{a} Department of Mathematics, Wayne State University, Detroit, MI 48202, United States
\textsuperscript{b} Department of Mathematical Sciences, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, United States

\textbf{A R T I C L E   I N F O}

Article history:
Received 6 August 2009
Available online 24 August 2010

MSC:
60J05
60J60

Keywords:
Switching diffusion
Smooth dependence
Continuous dependence

\textbf{A B S T R A C T}

This work is concerned with several properties of solutions of stochastic differential equations arising from hybrid switching diffusions. The word “hybrid” highlights the coexistence of continuous dynamics and discrete events. The underlying process has two components. One component describes the continuous dynamics, whereas the other is a switching process representing discrete events. One of the main features is the switching component depending on the continuous dynamics. In this paper, weak continuity is proved first. Then continuous and smooth dependence on initial data are demonstrated. In addition, it is shown that certain functions of the solutions verify a system of Kolmogorov’s backward differential equations. Moreover, rates of convergence of numerical approximation algorithms are dealt with.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

This work is concerned with solutions of stochastic differential equations of the form

\[ dX(t) = b(X(t), \alpha(t)) \, dt + \sigma(X(t), \alpha(t)) \, dw(t), \]

where \( b(\cdot) \) and \( \sigma(\cdot) \) are appropriate functions, \( w(\cdot) \) is a multi-dimensional standard Brownian motion, and \( \alpha(\cdot) \) is a switching process taking values in \( \mathcal{M} = \{1, \ldots, m_0\} \) whose operator and hence the...
dynamics depend on $X(t)$. The dependence is given in terms of transition probabilities to be specified later. Throughout the rest of the paper, we call such processes $x$-dependent switching-diffusion processes or continuous-state-dependent switching-diffusion processes. One of the distinct features is that in these systems, discrete events and continuous dynamics are highly correlated.

The $(X(t), \alpha(t))$ is a two-component Markov process. Although it is seemingly similar to the well-known diffusion processes, the properties of the solutions of the stochastic differential equation with switching are quite different from those of the usual diffusion processes (e.g., compare [7] and [8, 10, 17]). Moreover, stochastic differential equations with continuous-state-dependent switching are drastically different from stochastic differential equations with Markov switching. For instance, compare the stability of solutions of stochastic differential equations with Markovian switching and state-dependent switching; see [6, 10] and [8]. As another example, for diffusion processes, one of the commonly used numerical procedures is based on Picard iterations. For Markovian switching-diffusion processes, in which the switching process $\alpha(t)$ is a continuous-time Markov chain independent of the Brownian motion, the Picard method can still be applied [11, p. 83]. However, this approach can no longer be used for the $x$-dependent switching diffusion. When Markovian regime-switching processes are treated, with the given generator of the switching process, we can pre-generate the switching process throughout the iterations. In the continuous-state-dependent switching case, since $Q(x)$ varies with each iteration, we can no longer pre-generate the $x$-dependent switching process beforehand and apply the Lipschitz condition directly to obtain the desired estimates. In addition, due to the presence of the continuous-state-dependent switching processes, the analysis for properties of the solutions is usually much more complex than that of the usual diffusion processes without switching. For example, it is a time-honored concept that under suitable conditions, the usual diffusion process is well posed. That is, such a process is well defined, possesses unique solution for each initial condition, and the solution depends on the initial data continuously and smoothly; see for example [3]. When one has a Markovian switching-diffusion process, although more complex notation is needed compared with the diffusion counterpart, the well-posedness can be carried over with no essential difficulty. Nevertheless, to prove the well-posedness property for $x$-dependent switching diffusion is no longer a straightforward matter. One of the salient features and the main difficulties are the continuous dynamics and discrete events are intertwined. Here, we quest under what conditions, the well-posedness will still hold.

In this paper, we take a close scrutiny of solutions of stochastic differential equations with $x$-dependent switching. These properties include continuity of the solutions in the weak sense (to be defined more precisely in the later section), the smoothness of solutions with respect to the initial data $x$, and related error estimates for numerical solutions of the stochastic differential equations. We also derive a system of Kolmogorov-type backward equations.

The rest of the paper is organized as follows. The precise formulation of the stochastic differential equation with $x$-dependent switching is provided in Section 2. Unlike the Markovian switching-diffusion counterpart, in which once the generator of the Markov chain is specified, one could consider the stochastic differential equation alone, here we have also to take into consideration the dynamics of both the continuous state and the switching process. Section 3 is concerned with weak continuity of solutions of stochastic differential equations mentioned above, Section 4 takes up the issue of smooth dependence on the initial data. We also establish Feller property for $x$-dependent switching diffusion in Section 4. We derive a system of Kolmogorov-type backward equations in Section 5. Section 6 proceeds with error estimates of numerical solution for the underlying stochastic differential equations, which ascertains the rates of convergence of the numerical algorithms.

**2. Switching diffusions**

We work with a probability space $(\Omega, \mathcal{F}, P)$. A family of $\sigma$-algebras $\{\mathcal{F}_t\}$, for $t \geq 0$, or simply $\mathcal{F}_t$ is termed a filtration if $\mathcal{F}_t \subset \mathcal{F}_s$ for $s \leq t$. We say that $\mathcal{F}_t$ is complete if it contains all null sets and that the filtration $\{\mathcal{F}_t\}$ satisfies the usual condition if $\mathcal{F}_0$ is complete. As usual, a probability space $(\Omega, \mathcal{F}, P)$ together with a filtration $\{\mathcal{F}_t\}$ is said to be a filtered probability space, denoted by $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$. 
Suppose that \( \alpha(\cdot) \) is a stochastic process with right-continuous sample paths (or a pure jump process), finite-state space \( \mathcal{M} = \{1, \ldots, m_0\} \), and \( x \)-dependent generator \( Q(x) \) so that for a suitable function \( f(\cdot,\cdot) \),

\[
Q(x)f(x,\cdot)(t) = \sum_{j \in \mathcal{M}} q_{ij}(x) \left( f(x, j) - f(x, i) \right), \quad \text{for each } i \in \mathcal{M}. \tag{2.1}
\]

Assume throughout the paper that \( Q(x) \) satisfies the \( q \)-property (see [16]). That is, \( Q(x) = (q_{ij}(x)) \) satisfies

(i) \( q_{ij}(x) \) is Borel measurable and uniformly bounded for all \( i, j \in \mathcal{M} \) and \( x \in \mathbb{R}^r \);

(ii) \( q_{ij}(x) \geq 0 \) for all \( x \in \mathbb{R}^r \) and \( j \neq i \); and

(iii) \( q_{ii}(x) = -\sum_{j \neq i} q_{ij}(x) \) for all \( x \in \mathbb{R}^r \) and \( i \in \mathcal{M} \).

Let \( w(\cdot) \) be an \( \mathbb{R}^d \)-valued standard Brownian motion defined in the filtered probability space \( (\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P}) \). Suppose that \( b(\cdot, \cdot) : \mathbb{R} \times \mathcal{M} \times \mathbb{R}^r \mapsto \mathbb{R}^r \) and that \( \sigma(\cdot, \cdot) : \mathbb{R} \times \mathcal{M} \times \mathbb{R}^r \mapsto \mathbb{R}^d \). Then the two-component process \( (X(\cdot), \alpha(\cdot)) \), satisfying

\[
dX(t) = b(X(t), \alpha(t)) \, dt + \sigma(X(t), \alpha(t)) \, dw(t),
\]

\[
(X(0), \alpha(0)) = (x, \alpha), \tag{2.2}
\]

and for \( i \neq j \),

\[
\mathbb{P}\{\alpha(t+\Delta) = j \mid \alpha(t) = i, X(s), \alpha(s), s \leq t\} = q_{ij}(X(t)) \Delta + o(\Delta). \tag{2.3}
\]

is called an \( x \)-dependent switching diffusion.

The evolution of the discrete component or the switching process \( \alpha(\cdot) \) can be represented by a stochastic integral with respect to a Poisson random measure (see, e.g., [11,13]). To do so, for \( x \in \mathbb{R}^r \) and \( i, j \in \mathcal{M} \) with \( j \neq i \), let \( \Delta_{ij}(x) \) be the consecutive left-closed, right-open intervals of the real line, each having length \( q_{ij}(x) \). Define a function \( h : \mathbb{R} \times \mathcal{M} \times \mathbb{R} \mapsto \mathbb{R} \) by

\[
h(x, i, z) = \sum_{j=1}^{m_0} (j - i) I_{\{z \in \Delta_{ij}(x)\}}. \tag{2.4}
\]

Then we may write the switching process as a stochastic integral

\[
d\alpha(t) = \int h(X(t), \alpha(t-), z) \, p(dt, dz), \tag{2.5}
\]

where \( p(dt, dz) \) is a Poisson random measure with intensity \( dt \times m(dz) \), and \( m(\cdot) \) is the Lebesgue measure on \( \mathbb{R} \). The Poisson random measure \( p(\cdot, \cdot) \) is independent of the Brownian motion \( w(\cdot) \). Then the compensated or centered Poisson measure

\[
\mu(ds, dz) = p(ds, dz) - ds \times m(dz)
\]

is a martingale measure.

Similar to diffusions, for each \( i \in \mathcal{M} \) and each \( f(\cdot, i) \in C^2 \), a result known as generalized Itô’s lemma (see [11,13]) reads
\[ f(X(t), \alpha(t)) - f(X(0), \alpha(0)) = \int_0^t \mathcal{L} f(X(s), \alpha(s)) \, ds + M_1(t) + M_2(t), \tag{2.6} \]

where \( \mathcal{L} \) is the operator associated with the process \((X(t), \alpha(t))\), that is, for any \( f(\cdot, i) \in C^2, i \in \mathcal{M} \), we define

\[ \mathcal{L} f(x, i) = \frac{1}{2} \sum_{j,k=1}^{r} a_{jk}(x, i) \frac{\partial^2 f(x, i)}{\partial x_j \partial x_k} + \sum_{j=1}^{r} b_j(x, i) \frac{\partial f(x, i)}{\partial x_j} + Q(x)h(x, \cdot)(i), \tag{2.7} \]

and

\[ M_1(t) = \int_0^t [\nabla f(X(s), \alpha(s)), \sigma(X(s), \alpha(s))] dw(s), \]

\[ M_2(t) = \int_0^t \int_{\mathbb{R}} \left[ f(X(s), \alpha(0) + h(X(s), \alpha(s), z)) - f(X(s), \alpha(s)) \right] \mu(ds, dz), \]

where \((z, y)\) denotes the usual inner product on \(\mathbb{R}^r\).

In view of the generalized Itô formula,

\[ M_f(t) = M_1(t) + M_2(t) \]

\[ = f(X(t), \alpha(t)) - f(X(0), \alpha(0)) - \int_0^t \mathcal{L} f(X(s), \alpha(s)) \, ds \tag{2.8} \]

is a local martingale. If for each \( i \in \mathcal{M}, f(x, i) \in C^2_0 \) (class of functions possessing bounded and continuous derivatives with respect to \( x \) of orders up to 2) or \( f(x, i) \in C^2_0 \) (\( C^2 \) functions with compact support), then \( M_f(t) \) defined in (2.8) becomes a martingale. Similar to the case of diffusion processes, we can define the corresponding notion of solution of martingale problem accordingly.

**Proposition 2.1.** Let \( x \in \mathbb{R}^r, \mathcal{M} = \{1, \ldots, m_0\} \), and \( Q(x) = (q_{ij}(x)) \) an \( m_0 \times m_0 \) matrix depending on \( x \) satisfying the \( q \)-property. Consider the two-component process \( Y(t) = (X(t), \alpha(t)) \) given by (2.2)–(2.3) with initial data \((x, \alpha)\). Suppose that \( Q(\cdot) : \mathbb{R}^r \mapsto \mathbb{R}^{m_0 \times m_0} \) is a continuous function, that the functions \( b(\cdot, \cdot) \) and \( \sigma(\cdot, \cdot) \) satisfy

\[ |b(x, \alpha)| + |\sigma(x, \alpha)| \leq K_0(1 + |x|), \quad \alpha \in \mathcal{M}, \tag{2.9} \]

for some \( K_0 > 0 \) and that for every integer \( N \geq 1 \), there exists a positive constant \( M_N \) such that for all \( t \in [0, T], i \in \mathcal{M} \) and all \( x, y \in \mathbb{R}^r \) with \(|x| \vee |y| \leq M_N \),

\[ |b(x, i) - b(y, i)| \vee |\sigma(x, i) - \sigma(y, i)| \leq M_N |x - y|. \tag{2.10} \]

Then there exists a unique solution \((X(t), \alpha(t))\) to Eq. (2.2) with given initial data in which the evolution of the jump process is specified by (2.3).
The proof of the above proposition is deemed to be well known. For brevity, the detailed proof is omitted. Instead, we make the following remarks. There are a number of possible proofs. For example, the existence can be obtained as in [13, pp. 103–104]. Viewing the switching diffusion as a special case of a jump-diffusion process (see the stochastic integral representation of $\alpha(t)$ in (2.5)), one may prove the existence and uniqueness using [5, Section III.2]. Another possibility is to use a martingale problem formulation together with utilization of truncations and stopping times as in [4, Chapter IV]. In [15], we proposed and analyzed a couple of numerical approximation algorithms for approximating solutions of switching diffusions. We showed that the interpolations of the iterates converge weakly to the switching diffusion by a martingale problem formulation. Then using Lipschitz continuity and the weak convergence, we further obtain the strong convergence of the approximations. As a byproduct, we also obtained the existence and uniqueness of the solution.

3. Weak continuity

We begin this section with a definition of weak continuity. We then proceed with obtaining such a property for the solutions of stochastic differential equations of $x$-dependent switching-diffusion processes.

**Definition 3.1.** Recall that a stochastic process $Y(t)$ with right-continuous sample paths is said to be weakly continuous or continuous in probability at $t$ if for any $\eta > 0$,

$$\lim_{\Delta \to 0} P\left( |Y(t + \Delta) - Y(t)| \geq \eta \right) = 0. \quad (3.1)$$

It is mean square continuous at $t$ if

$$\lim_{\Delta \to 0} E|Y(t + \Delta) - Y(t)|^2 = 0. \quad (3.2)$$

The process $Y(t)$ is said to be continuous in probability in the interval $[0, T]$ (or in short continuous in probability if the interval $[0, T]$ is clearly understood), if it is continuous in probability at every $t \in [0, T]$. Likewise it is continuous in mean square if it is continuous in mean square at every $t \in [0, T]$.

As a preparation, we first state a lemma, whose proof can be found in [18].

**Lemma 3.2.** Assume the conditions of Proposition 2.1. Let $T > 0$ be fixed. Then for any positive constant $\gamma$, we have

$$E_{x,i} \left[ \sup_{t \in [0, T]} |X(t)|^\gamma \right] \leq K < \infty, \quad (x, i) \in \mathbb{R}^r \times \mathcal{M}, \quad (3.3)$$

where $K = K(x, T, \gamma)$ is a constant depending on $x$, $T$, and $\gamma$.

We proceed to obtain the weak continuity for the two-component switching-diffusion processes $Y(t) = (X(t), \alpha(t))$. The results are to be presented in two parts. The first part concentrates on Markovian switching diffusions, whereas the second one is concerned with continuous-state-dependent switching diffusions.

**Theorem 3.3.** Suppose that the conditions of Proposition 2.1 are satisfied with the modification $Q(x) = Q$ that generates a Markov chain independent of the Brownian motion. Then the process $Y(t) = (X(t), \alpha(t))$ is continuous in probability and also continuous in mean square.
Proof. We will show that for any $\eta > 0$,
\[
P\left(\left| Y(t + \Delta) - Y(t) \right| \geq \eta \right) \to 0 \quad \text{as } \Delta \to 0.
\] (3.4)

Note that
\[
Y(t + \Delta) - Y(t) = \left( X(t + \Delta), \alpha(t + \Delta) \right) - \left( X(t + \Delta), \alpha(t) \right)
+ \left( X(t + \Delta), \alpha(t) \right) - \left( X(t), \alpha(t) \right).
\] (3.5)

We divide the rest of the proof into several steps.

Step 1: First we recognize that in view of (3.5),
\[
E \left[ \left| Y(t + \Delta) - Y(t) \right|^2 \right] \leq 2 \left[ E \left[ \left| \alpha(t + \Delta) - \alpha(t) \right|^2 \right] + E \left[ \left| X(t + \Delta) - X(t) \right|^2 \right] \right].
\] (3.6)

Thus to estimate the difference of the second moment, it suffices to consider the two marginal estimates separately. We do this in the next two steps.

Step 2: We claim that for any $t \geq 0$ and $\Delta \geq 0$,
\[
E \left[ \left| X(t + \Delta) - X(t) \right|^2 \right] \leq K \Delta.
\] (3.7)

In the above and hereafter, $K$ is taken to be generic positive constant whose values may be different for different appearances.

Estimate (3.7), in fact, is a modification of the standard estimates for solutions of stochastic differential equations without switching. It mainly uses the linear growth and Lipschitz conditions of the drift and diffusion coefficients and Lemma 3.2. We thus omit the details.

Step 3: Note that for any $t \geq 0$,
\[
\alpha(t) = \sum_{i=1}^{m_0} i I_{\{\alpha(t) = i\}} = \chi(t)(1, \ldots, m_0)',
\]
where
\[
\chi(t) = (\chi_1(t), \ldots, \chi_{m_0}(t)) = (I_{\{\alpha(t) = 1\}}, \ldots, I_{\{\alpha(t) = m_0\}}) \in \mathbb{R}^{1 \times m_0},
\] (3.8)
and $(1, \ldots, m_0)' \in \mathbb{R}^{m_0}$ is a column vector. Since the Markov chain $\alpha(t)$ is independent of the Brownian motion $w(\cdot)$ ($Q$ is a constant matrix), it is well known that
\[
\chi(t + \Delta) - \chi(t) - \int_{t}^{t+\Delta} \chi(s) Q \, ds \quad \text{is a martingale;}
\]
see, for instance, [16, Lemma 2.4]. It follows that
\[
E_t \left[ \chi(t + \Delta) - \chi(t) - \int_{t}^{t+\Delta} \chi(s) Q \, ds \right] = 0,
\]
where $E_t$ denotes the conditional expectation on the $\sigma$-algebra.
\[ \mathcal{F}_t = \{(X(u), \alpha(u)) : u \leq t\}. \]

The boundedness of \( \chi(s) \) then implies that

\[
\left| \int_t^{t+\Delta} \chi(s) Q \, ds \right| = O(\Delta) \quad \text{a.s.} \tag{3.9}
\]

Thus, we obtain

\[
\mathbb{E}_t \chi(t+\Delta) = \chi(t) + O(\Delta) \quad \text{a.s.} \tag{3.10}
\]

Note the structure of \( \chi(t) \),

\[
\chi(t+\Delta) - \chi(t) = (\chi_1(t+\Delta) - \chi_1(t)), \ldots, \chi_{m_0}(t+\Delta) - \chi_{m_0}(t)
\]

with \( \chi_i(\cdot) \) given by (3.8). Using (3.10), we have

\[
\mathbb{E}_t \left[ \chi_i(t+\Delta) - \chi_i(t) \right]^2 \\
= \mathbb{E}_t \left[ I_{[\alpha(t+\Delta)=i]} - I_{[\alpha(t)=i]} \right]^2 \\
= \left[ \mathbb{E}_t I_{[\alpha(t+\Delta)=i]} - 2I_{[\alpha(t)=i]} \mathbb{E}_t I_{[\alpha(t+\Delta)=i]} + I_{[\alpha(t)=i]} \right] \\
= O(\Delta) \quad \text{a.s.} \tag{3.11}
\]

Step 4: Next we consider

\[
\mathbb{E} \left[ \alpha(t+\Delta) - \alpha(t) \right]^2 = \mathbb{E} \left[ (\chi(t+\Delta) - \chi(t))(1, \ldots, m_0) \right]^2 \\
\leq K \mathbb{E} \left[ \chi(t+\Delta) - \chi(t) \right]^2 \\
\leq K \sum_{i=1}^{m_0} \mathbb{E} \mathbb{E}_t \left[ \chi_i(t+\Delta) - \chi_i(t) \right]^2 \\
\leq K \Delta \to 0 \quad \text{as } \Delta \to 0. \tag{3.12}
\]

From the next to the last line above, we have used (3.11). By combining (3.7) and (3.12), we obtain that (3.6) leads to

\[
\mathbb{E} \left[ Y(t+\Delta) - Y(t) \right]^2 \to 0 \quad \text{as } \Delta \to 0.
\]

The mean square continuity has been established. Then the desired continuity in probability follows from Tchebyshev's inequality. \( \square \)

We next generalize the above result and allow the switching process to be \( x \)-dependent. The assertion is presented next.

**Theorem 3.4.** Suppose that the conditions of Proposition 2.1 are satisfied. Then the process \( Y(t) = (X(t), \alpha(t)) \) is continuous in probability and continuous in mean square.
Proof. Steps 1 and 2 are the same as before. We shall only point out the main difference as compared to Theorem 3.3.

Consider the function $H(x, \alpha) = I_{\{\alpha = i\}}$ for each $i \in \mathcal{M}$. Since $H$ is independent of $x$, it is readily seen that

$$\mathcal{L}H(x, \alpha) = Q(x)H(x, \cdot)(\alpha).$$

Consequently,

$$0 = \mathbb{E}_t\left[ H(X(t + \delta), \alpha(t + \Delta)) - H(X(t), \alpha(t)) - \int_t^{t+\Delta} \mathcal{L}H(X(s), \cdot)(\alpha) ds \right]$$

$$= \mathbb{E}_t\left[ H(X(t + \Delta), \alpha(t + \Delta)) - H(X(t), \alpha(t)) - \int_t^{t+\Delta} Q(X(s))H(X(s), \cdot)(\alpha(s)) ds \right].$$

Then the definition of $H(x, \alpha)$ yields that

$$\mathbb{E}_t\left[ H(X(t + \Delta), \alpha(t + \Delta)) - H(X(t), \alpha(t)) - \int_t^{t+\Delta} \mathcal{L}H(X(s), \alpha(s)) ds \right]$$

$$= \mathbb{E}_t\left[ I_{\{\alpha(t + \Delta) = i\}} - I_{\{\alpha(t) = i\}} - \int_t^{t+\Delta} \sum_{j=1}^{m_q} q_{ij}(X(s))I_{\{\alpha(s) = i\}} ds \right].$$

Since $Q(x)$ is bounded, similar to (3.9), we obtain

$$\left| \int_t^{t+\Delta} \chi(s)Q(X(s)) ds \right| = O(\Delta) \quad \text{a.s.} \quad (3.13)$$

With (3.13) at our hands, we proceed as the rest of Steps 3 and 4 in the proof of Theorem 3.3. The desired result follows. □

4. Continuous and smooth dependence on the initial data $x$

When one deals with a continuous-time dynamic system modeled by a differential equation together with appropriate initial data, the well-posedness is crucial. The well-posedness appears in ordinary differential equations as well as in partial differential equations together with initial and/or boundary data. They are time-honored phenomena, which naturally carry over to stochastic differential equations as well as stochastic differential equations with Markovian switching. A problem for the associated switching diffusion is well posed if there is a unique solution for the initial value problem and the solution continuously depends on the initial data.

In this section, we devote our attention to the continuous and smoothness dependence on initial data. Since it is more difficult to obtain the smoothness property with respect to the initial data in the mean square sense, we shall first treat this problem.
4.1. Continuity and differentiability

Let us first recall the notion of multi-index. A vector \( \beta = (\beta_1, \ldots, \beta_r) \) with nonnegative integer components is referred to as a multi-index. Put \(|\beta| = \beta_1 + \cdots + \beta_r\), and define \( D_\beta^x \) as

\[
D_\beta^x = \frac{\partial^{\beta}}{\partial x^{\beta}} = \frac{\partial|\beta|}{\partial x_1^{\beta_1} \cdots x_r^{\beta_r}}.
\]

**Definition 4.1.** Suppose that \( \Psi(x_1, \ldots, x_r, t) \) is a random function. Its partial derivative in mean square with respect to \( x_i \) for some \( 1 \leq i \leq r \) is defined as the random variable \( \tilde{\Psi}(x_1, \ldots, x_r, t) \) such that

\[
E \left| \frac{1}{\Delta x_i} \left[ \Psi(x_1, \ldots, x_i + \Delta x_i, \ldots, x_r, t) - \Psi(x_1, \ldots, x_i, \ldots, x_r, t) \right] - \tilde{\Psi}(x_1, \ldots, x_r, t) \right|^2 \to 0 \quad \text{as} \ \Delta x_i \to 0.
\]

When the mean square partial derivative exists, we normally write it as

\[
\tilde{\Psi}(x_1, \ldots, x_r) = \frac{\partial \Psi(x_1, \ldots, x_r)}{\partial x_i} = \Psi_{x_i}(x_1, \ldots, x_r).
\]  

(4.1)

**Theorem 4.2.** Assume the conditions of Proposition 2.1 with the modification of the local Lipschitz condition replaced by a global Lipschitz condition. Let \((X^x(t), \alpha^x(t))\) be the solution to the system given by (2.2) and (2.3). Assume that for each \( i \in \mathcal{M} \), \( b(\cdot, i) \) and \( \sigma(\cdot, i) \) have continuous partial derivatives with respect to the variable \( x \) up to the second order and that

\[
|D_\beta^x b(x, i)| + |D_\beta^x \sigma(x, i)| \leq K_0 \left( 1 + |x|^\gamma \right),
\]  

(4.2)

where \( K_0 \) and \( \gamma \) are positive constants and \( \beta \) is a multi-index with \(|\beta| \leq 2\). Then \( X^{x, \alpha}(t) \) is twice continuously differentiable in mean square with respect to \( x \).

For ease of presentation, we will prove Theorem 4.2 for the case when \( X(t) \) is 1-dimensional. Multi-dimensional case can be handled similarly; only notation is more involved. It seems to be more instructive to consider a case with simpler notation so as to gain the main insight without much notational complication. To proceed, we need to introduce a few more notations. Let \( \Delta \neq 0 \) be small and denote \( \tilde{x} = x + \Delta \). Let \((X(t), \alpha(t))\) be the switching-diffusion process satisfying (2.2) and (2.3) with initial condition \((x, \alpha)\) and \((\tilde{X}(t), \tilde{\alpha}(t))\) be the process starting from \((\tilde{x}, \tilde{\alpha})\) (i.e., \((X(0), \alpha(0)) = (x, \alpha)\) and \((\tilde{X}(0), \tilde{\alpha}(0)) = (\tilde{x}, \tilde{\alpha})\) respectively).

Fix any \( T > 0 \) and let \( 0 < t < T \). Put

\[
Z(t) = Z^\Delta(t) = Z^{x, \Delta, \alpha}(t) := \frac{\tilde{X}(t) - X(t)}{\Delta}.
\]  

(4.3)

Then we have

\[
Z^\Delta(t) = 1 + \frac{1}{\Delta} \int_0^t \left[ b(\tilde{X}(s), \tilde{\alpha}(s)) - b(X(s), \alpha(s)) \right] ds
\]
\[
+ \frac{1}{\Delta} \int_0^t \left[ \sigma(\tilde{X}(s), \tilde{\alpha}(s)) - b(X(s), \alpha(s)) \right] dw(s)
\]

\[
= 1 + \phi^\Delta(t) + \frac{1}{\Delta} \int_0^t \left[ b(\tilde{X}(s), \alpha(s)) - b(\tilde{X}(s), \alpha(s)) \right] ds
\]

\[
+ \frac{1}{\Delta} \int_0^t \left[ \sigma(\tilde{X}(s), \alpha(s)) - \sigma(\tilde{X}(s), \alpha(s)) \right] dw(s),
\]

(4.4)

where

\[
\phi^\Delta(t) = \frac{1}{\Delta} \int_0^t \left[ b(\tilde{X}(s), \tilde{\alpha}(s)) - b(\tilde{X}(s), \alpha(s)) \right] ds
\]

\[
+ \frac{1}{\Delta} \int_0^t \left[ \sigma(\tilde{X}(s), \tilde{\alpha}(s)) - \sigma(\tilde{X}(s), \alpha(s)) \right] dw(s).
\]

(4.5)

**Lemma 4.3.** Under the conditions of Theorem 4.2,

\[
\lim_{\Delta \to 0} E \left[ \sup_{0 \leq t \leq T} \left| \phi^\Delta(t) \right|^2 \right] = 0.
\]

**Proof.** By virtue of Hölder’s inequality and Doob’s martingale inequality (see, for example, [5, p. 11] or [11, p. 18]),

\[
E \sup_{0 \leq t \leq T} \left| \phi^\Delta(t) \right|^2 \leq \frac{2T}{\Delta^2} E \int_0^T \left| b(\tilde{X}(s), \tilde{\alpha}(s)) - b(\tilde{X}(s), \alpha(s)) \right|^2 ds
\]

\[
+ \frac{8}{\Delta^2} E \left| \int_0^T \left[ \sigma(\tilde{X}(s), \tilde{\alpha}(s)) - \sigma(\tilde{X}(s), \alpha(s)) \right] dw(s) \right|^2.
\]

We treat each of the terms above separately. Choose \( \eta = \Delta^{\eta_0} \) with \( \eta_0 > 2 \) and partition the interval \([0, T]\) by \( \eta \). We obtain

\[
E \int_0^T \left| b(\tilde{X}(s), \tilde{\alpha}(s)) - b(\tilde{X}(s), \alpha(s)) \right|^2 ds
\]

\[
= E \sum_{k=0}^{\lfloor T/\eta \rfloor - 1} \int_{k\eta}^{(k+1)\eta} \left| b(\tilde{X}(s), \tilde{\alpha}(s)) - b(\tilde{X}(s), \alpha(s)) \right|^2 ds
\]

\[
\leq \sum_{k=0}^{\lfloor T/\eta \rfloor - 1} KE \int_{k\eta}^{(k+1)\eta} \left| b(\tilde{X}(s), \tilde{\alpha}(s)) - b(\tilde{X}(\eta k), \tilde{\alpha}(s)) \right|^2 ds
\]
Note that the constant $K$ in (4.6) does not depend on $k = 0, 1, \ldots, \lfloor T/\eta \rfloor$ or $\eta$. The exact value of $K$ may be different in each occurrence (i.e., we view $K$ as a generic positive constant and use this convention throughout).

By the Lipschitz continuity and the tightness type of estimate (3.7), we obtain

\begin{align*}
\mathbb{E} & \int_{k\eta}^{k\eta+\eta} \left| b(\tilde{X}(s), \tilde{\alpha}(s)) - b(\tilde{X}(\eta k), \tilde{\alpha}(s)) \right|^2 ds \\
& \leq K \int_{k\eta}^{k\eta+\eta} \mathbb{E} \left| \tilde{X}(s) - \tilde{X}(\eta k) \right|^2 ds \\
& \leq K \int_{k\eta}^{k\eta+\eta} (s - \eta k) ds \leq K \eta^2. \quad (4.7)
\end{align*}

Likewise, we can deal with the term on the last line of (4.6), and obtain

\begin{align*}
\mathbb{E} & \int_{k\eta}^{k\eta+\eta} \left| b(\tilde{X}(\eta k), \alpha(\eta k)) - b(\tilde{X}(s), \alpha(s)) \right|^2 ds \leq K \eta^2. \quad (4.8)
\end{align*}

To treat the term on the next to the last line of (4.6), note that for $k = 0, 1, \ldots, \lfloor T/\eta \rfloor - 1$,

\begin{align*}
\mathbb{E} & \int_{k\eta}^{k\eta+\eta} \left| b(\tilde{X}(\eta k), \tilde{\alpha}(s)) - b(\tilde{X}(\eta k), \tilde{\alpha}(\eta k)) \right|^2 ds \\
& \leq K \mathbb{E} \int_{k\eta}^{k\eta+\eta} \left| b(\tilde{X}(\eta k), \tilde{\alpha}(s)) - b(\tilde{X}(\eta k), \tilde{\alpha}(\eta k)) \right|^2 ds \\
& + K \mathbb{E} \int_{k\eta}^{k\eta+\eta} \left| b(\tilde{X}(\eta k), \tilde{\alpha}(\eta k)) - b(\tilde{X}(\eta k), \alpha(s)) \right|^2 ds. \quad (4.9)
\end{align*}

For the term on the second line of (4.9) and $k = 0, 1, \ldots, \lfloor T/\eta \rfloor - 1$,
Theorem 3.2 and the boundedness of $Q(x)$. Next, we show that for $k = 1, \ldots, \lceil T/\eta \rceil - 1$,

$$
\mathbb{E} \int_{k\eta}^{k\eta+\eta} \left| b\left(\tilde{X}(\eta k), \tilde{\alpha}(\eta k)\right) - b\left(\tilde{X}(\eta k), \tilde{\alpha}(\eta k)\right) \right|^2 ds \leq K \eta.
$$

(4.10)

To do so, we use the technique of basic coupling of Markov processes (see, for example, the book of Chen [1, p. 11]). For $x, \tilde{x} \in \mathbb{R}^r$, and $i, j \in \mathcal{M}$, consider the measure

$$
\Lambda((x, j), (\tilde{x}, i)) = |x - \tilde{x}| + d(j, i),
$$

where

$$
d(j, i) = \begin{cases} 
0 & \text{if } j = i, \\
1 & \text{if } j \neq i. 
\end{cases}
$$

That is, $\Lambda(\cdot, \cdot)$ is a measure obtained by piecing the usual Euclidean length of two vectors and the discrete measure together. Let $(\alpha(t), \tilde{\alpha}(t))$ be a discrete random process with a finite-state space $\mathcal{M} \times \mathcal{M}$ such that

$$
\mathbb{P}\left[ (\alpha(t+h), \tilde{\alpha}(t+h)) = (j, i) \mid (\alpha(t), \tilde{\alpha}(t)) = (k, l), \ (X(t), \tilde{X}(t)) = (x, \tilde{x}) \right] = \begin{cases} 
\hat{q}(k, l)(j, i)(x, \tilde{x})h + o(h) & \text{if } (k, l) \neq (j, i), \\
1 + \hat{q}(k, l)(k, l)(x, \tilde{x})h + o(h) & \text{if } (k, l) = (j, i), 
\end{cases}
$$

(4.11)
where $h \to 0$, and the matrix $(\tilde{q}_{(k,l)(j,i)}(x, \tilde{x}))$ is the basic coupling of matrices $Q(x) = (q_{kl}(x))$ and $Q(\tilde{x}) = (q_{kl}(\tilde{x}))$ satisfying

$$
\bar{Q}(x, \tilde{x}) \hat{f}(k, l) = \sum_{(j,i) \in \mathcal{M} \times \mathcal{M}} q_{(k,l)(j,i)}(x, \tilde{x})(\hat{f}(j, i) - \hat{f}(k, l))
$$

$$
= \sum_{j} (q_{kj}(x) - q_{lj}(\tilde{x}))^+ (\hat{f}(j, l) - \hat{f}(k, l))
$$

$$
+ \sum_{j} (q_{lj}(\tilde{x}) - q_{kj}(x))^+ (\hat{f}(k, j) - \hat{f}(k, l))
$$

$$
+ \sum_{j} (q_{kj}(x) \wedge q_{lj}(\tilde{x})) (\hat{f}(j, j) - \hat{f}(k, l)).
$$

(4.12)

for any function $\hat{f}(\cdot, \cdot)$ defined on $\mathcal{M} \times \mathcal{M}$. Note that for $s \in [\eta k, \eta k + \eta)$, $\tilde{a}(s)$ can be written as $
\tilde{a}(s) = \sum_{l \in \mathcal{M}} I_{[\tilde{a}(s) = l]}$. Owing to the coupling defined above and noting the transition probabilities (4.11), for $i_1, i, j, l \in \mathcal{M}$ with $j \neq i$ and $s \in [\eta k, \eta k + \eta)$, we have

$$
\mathbb{E}[I_{[\tilde{a}(s) = j]}|a(\eta k) = i_1, \tilde{a}(\eta k) = i, X(\eta k) = x, \tilde{X}(\eta k) = \tilde{x}]
$$

$$
= \sum_{l \in \mathcal{M}} \mathbb{E}[I_{[\tilde{a}(s) = j]}|a(\eta k) = i_1, \tilde{a}(\eta k) = i, X(\eta k) = x, \tilde{X}(\eta k) = \tilde{x}]
$$

$$
= \sum_{l \in \mathcal{M}} \bar{q}_{(i_1,i)(j,l)}(x, \tilde{x})(s - \eta k) + o(s - \eta k) = O(\eta).
$$

(4.13)

By virtue of (4.13), we obtain

$$
\mathbb{E} \int_{k\eta}^{k\eta + \eta} \left| b(\tilde{X}(\eta k), \tilde{a}(\eta k)) - b(\tilde{X}(\eta k), a(s)) \right|^2 ds
$$

$$
= \mathbb{E} \int_{k\eta}^{k\eta + \eta} \left| b(\tilde{X}(\eta k), a(s)) - b(\tilde{X}(\eta k), \tilde{a}(\eta k)) \right|^2 I_{[a(s) \neq \tilde{a}(\eta k)]} ds
$$

$$
\leq \mathbb{E} \sum_{i, i_1 \in \mathcal{M}} \sum_{j \neq i} \int_{k\eta}^{k\eta + \eta} \left| b(\tilde{X}(\eta k), i) - b(\tilde{X}(\eta k), j) \right|^2 I_{[a(s) = j]} I_{[\tilde{a}(\eta k) = i]} ds
$$

$$
\times \mathbb{E} \left[ a(s) = j | a(\eta k) = i_1, \tilde{a}(\eta k) = i, X(\eta k) = x, \tilde{X}(\eta k) = \tilde{x} \right] ds
$$

$$
= O(\eta^2).
$$

Using the assumption $\tilde{a}(0) = a(0) = a$ and noting $\tilde{X}(0) = \tilde{x}$, we obtain
\[ E \int_0^\eta \left| b(\tilde{X}(0), \tilde{\alpha}(0)) - b(\tilde{X}(0), \alpha(s)) \right|^2 ds = E \int_0^\eta \left| b(\tilde{x}, \alpha(0)) - b(\tilde{x}, \alpha(s)) \right|^2 ds = E \int_0^\eta \sum_{j \neq \alpha} \left| b(\tilde{x}, \alpha) - b(\tilde{x}, j) \right|^2 I_{\{\alpha(s) = j\}} ds \leq K \eta^2. \] (4.14)

Thus, it follows that for \( k = 0, 1, \ldots, \lfloor T/\eta \rfloor - 1, \)
\[ E \int_{k \eta}^{k \eta + \eta} \left| b(\tilde{X}(\eta k), \tilde{\alpha}(s)) - b(\tilde{X}(\eta k), \alpha(s)) \right|^2 ds \leq K \eta^2. \] (4.15)

Using the estimates (4.7), (4.8), and (4.15) in (4.6), we obtain
\[ E \int_0^T \left| b(\tilde{X}(s), \tilde{\alpha}(s)) - b(\tilde{X}(s), \alpha(s)) \right|^2 ds \leq \sum_{k=0}^{\lfloor T/\eta \rfloor - 1} K \eta^2 \leq K \eta. \] (4.16)

Likewise, we obtain
\[ E \left\| \int_0^T \left[ \sigma(\tilde{X}(s), \tilde{\alpha}(s)) - \sigma(\tilde{X}(s), \alpha(s)) \right] dw(s) \right\|^2 \leq K \eta. \] (4.17)

Recall the definition of \( \phi^\Delta(t) \) given in (4.5). Putting (4.16) and (4.17) into \( \phi^\Delta(t) \), we obtain
\[ E \sup_{0 \leq t \leq T} \left| \phi^\Delta(t) \right|^2 \leq K \frac{\eta}{\Delta^2} = K \Delta^{\gamma_0 - 2} \to 0 \quad \text{as} \quad \Delta \to 0, \] (4.18)

since \( \gamma_0 > 2 \). The lemma is proved. \( \square \)

**Remark 4.4.** In deriving (4.14), we used \( \alpha(0) = \tilde{\alpha}(0) = \alpha \). This condition is crucial. If the initial data of the switching processes are not the same, there will be a nonzero contribution resulting in difficulties in obtaining the differentiability.

**Proposition 4.5.** Assume the conditions of Proposition 2.1 hold with the modification of the local Lipschitz condition replaced by a global Lipschitz condition. Then for any fixed \( T > 0 \), we have
\[ E \left[ \sup_{0 \leq t \leq T} \left| X^x,\alpha(t) - X^x,\alpha(t) \right|^2 \right] \leq K |\tilde{x} - x|^2, \] (4.19)

where \( K \) is a constant depending only on \( T \) and the global Lipschitz and the linear growth constant \( K_0 \).
Proof. As before, let \((X(t), \alpha(t))\) denote the switching-diffusion process satisfying (2.2) and (2.3) with initial condition \((x, \alpha)\) and \((\tilde{X}(t), \tilde{\alpha}(t))\) be the process starting from \((\tilde{x}, \tilde{\alpha})\) (i.e., \((X(0), \alpha(0)) = (x, \alpha)\) and \((\tilde{X}(0), \tilde{\alpha}(0)) = (\tilde{x}, \tilde{\alpha})\) respectively). Let \(T > 0\) be fixed and denote \(\Delta = \tilde{x} - x\). Then we have \(X(t) - \tilde{X}(t) = \Delta + A(t) + B(t)\), and hence for any \(0 < T_1 \leq T\), we have

\[
\sup_{t \in [0, T_1]} |X(t) - \tilde{X}(t)|^2 \leq 3 \Delta^2 + 3 \sup_{t \in [0, T_1]} |A(t)|^2 + 3 \sup_{t \in [0, T_1]} |B(t)|^2,
\]

where

\[
A(t) := \int_0^t \left[ b(\tilde{X}(s), \tilde{\alpha}(s)) - b(X(s), \alpha(s)) \right] ds
\]
\[
+ \int_0^t \left[ \sigma(\tilde{X}(s), \tilde{\alpha}(s)) - \sigma(X(s), \alpha(s)) \right] dw(s) = \Delta \phi^\Delta(t),
\]

and

\[
B(t) := \int_0^t \left[ b(\tilde{X}(s), \alpha(s)) - b(X(s), \alpha(s)) \right] ds
\]
\[
+ \int_0^t \left[ \sigma(\tilde{X}(s), \alpha(s)) - \sigma(X(s), \alpha(s)) \right] dw(s).
\]

Because \(\gamma_0 > 2\), (4.18) yields that

\[
E \sup_{t \in [0, T_1]} |A(t)|^2 \leq K \Delta^2 \Delta^{\gamma_0 - 2} = K \Delta^{\gamma_0 - 2} = O(\Delta^{\gamma_0}) = o(\Delta^2).
\]

Meanwhile, by virtue of Hölder inequality and Doob’s martingale inequality, we obtain

\[
E \sup_{t \in [0, T_1]} |B(t)|^2 \leq 2E \left[ \int_0^{T_1} \left| b(\tilde{X}(s), \alpha(s)) - b(\tilde{X}(s), \alpha(s)) \right|^2 ds \right]^2
\]
\[
+ 8E \left[ \int_0^{T_1} \left| \sigma(\tilde{X}(s), \alpha(s)) - \sigma(\tilde{X}(s), \alpha(s)) \right|^2 dw(s) \right]^2
\]
\[
\leq 2TE \left[ \int_0^{T_1} \left| b(\tilde{X}(s), \alpha(s)) - b(\tilde{X}(s), \alpha(s)) \right|^2 ds \right]
\]
\[
+ 8E \left[ \int_0^{T_1} \left| \sigma(\tilde{X}(s), \alpha(s)) - \sigma(\tilde{X}(s), \alpha(s)) \right|^2 dw(s) \right]
\]

Furthermore, using Lipschitz continuity, we obtain
\[
\mathbb{E} \sup_{t \in [0, T_1]} \left| B(t) \right|^2 \leq K(T + 1) \int_0^{T_1} \mathbb{E} \left[ \tilde{X}(s) - X(s) \right]^2 ds \\
\leq K(T + 1) \int_0^{T_1} \mathbb{E} \left[ \sup_{0 \leq u \leq s} \tilde{X}(u) - X(u) \right]^2 ds.
\]

Therefore, we have
\[
\mathbb{E} \sup_{t \in [0, T_1]} \left| \tilde{X}(t) - X(t) \right|^2 \leq K \Delta^2 + o(\Delta^2) + K(T + 1) \int_0^{T_1} \mathbb{E} \left[ \sup_{0 \leq u \leq s} \tilde{X}(u) - X(u) \right]^2 ds.
\]

Then it follows from Gronwall’s inequality that
\[
\mathbb{E} \sup_{t \in [0, T_1]} \left| \tilde{X}(t) - X(t) \right|^2 \leq K \Delta^2 \exp \{K(T + 1)T_1\} \leq K \Delta^2 \exp \{K(T + 1)T\}.
\]

Since \( T_1 \leq T \) is arbitrary, we have
\[
\mathbb{E} \left[ \sup_{t \in [0, T]} \left| \tilde{X}(t) - X(t) \right|^2 \right] \leq K \Delta^2 = K |\tilde{x} - x|^2.
\]

This finishes the proof of the proposition. \( \square \)

**Remark 4.6.** While most proofs of the uniqueness of the solutions of stochastic differential equations take two different solutions with the same initial data and show their difference should be 0 by using Lipschitz continuity and Gronwall’s inequality, it is possible to consider the difference of the two solutions with different initial data whose difference is arbitrarily small. In this regard, the uniqueness of the solution of (2.2) with (2.3) can be derived from Proposition 4.5. Earlier work using such an approach may be found in [12].

A direct consequence of Proposition 4.5 is the mean square continuity of the solution of the switching diffusion with respect to \( x \), that is, for any \( T > 0 \),
\[
\lim_{y \to x} \mathbb{E} \left| X^{y, \alpha}(t) - X^{x, \alpha}(t) \right|^2 = 0, \quad \text{for each } \alpha \in \mathcal{M} \text{ and } t \in [0, T].
\]

That is, the continuous dependence on the initial data \( x \) is obtained. We state this fact below.

**Corollary 4.7.** Assume the conditions of Theorem 4.2. Then \( X^{x, \alpha}(t) \) is continuous in mean square with respect to \( x \).

**Proof of Theorem 4.2.** With Lemma 4.3 and Proposition 4.5 at our hands, we proceed to prove Theorem 4.2. Since \( b(\cdot, j) \) is twice continuously differentiable with respect to \( x \), we can write
\[
\frac{1}{\Delta} \int_0^t \left[ b(\tilde{X}(s), \alpha(s)) - b(X(s), \alpha(s)) \right] ds
\]
\[=rac{1}{\Delta} \int_0^t \int_0^1 \frac{d}{dv} b(X(s) + v(\tilde{X}(s) - X(s)), \alpha(s)) \, dv \, ds\]
\[= \int_0^t \left[ \int_0^1 b_x(X(s) + v(\tilde{X}(s) - X(s)), \alpha(s)) \, dv \right] Z^\Delta(s) \, ds,\]

where \( b_x(\cdot) \) denotes the partial derivative of \( b(\cdot) \) with respect to \( x \) (i.e., \( b_x = (\partial/\partial x)b \)). It follows from Proposition 4.5 that for any \( s \in [0, T] \)

\[\tilde{X}(s) - X(s) \to 0 \quad \text{in probability as } \Delta \to 0.\]

This implies that
\[\int_0^1 b_x(X(s) + v(\tilde{X}(s) - X(s)), \alpha(s)) \, dv \to b_x(X(s), \alpha(s)) \quad (4.21)\]

in probability as \( \Delta \to 0 \). Similarly, we have
\[\frac{1}{\Delta} \int_0^t \left[ \sigma(\tilde{X}(s), \alpha(s)) - \sigma(X(s), \alpha(s)) \right] dw(s)\]
\[= \int_0^t \left[ \int_0^1 \sigma_x(X(s) + v(\tilde{X}(s) - X(s)), \alpha(s)) \, dv \right] Z^\Delta(s) \, dw(s)\]

and
\[\int_0^1 \sigma_x(X(s) + v(\tilde{X}(s) - X(s)), \alpha(s)) \, dv \to \sigma_x(X(s), \alpha(s)) \quad (4.22)\]

in probability as \( \Delta \to 0 \). Let \( \zeta(t) := \zeta^{x,\alpha}(t) \) be the solution of
\[\zeta(t) = 1 + \int_0^t b_x(X(s), \alpha(s)) \zeta(s) \, ds + \int_0^t \sigma_x(X(s), \alpha(s)) \zeta(s) \, dw(s), \quad (4.23)\]

where \( b_x \) and \( \sigma_x \) denote the partial derivatives of \( b \) and \( \sigma \) with respect to \( x \), respectively. Then (4.4), (4.18), (4.21), (4.22), and [2, Theorem 5.5.2] imply that
\[E \left| Z^\Delta(t) - \zeta(t) \right|^2 \to 0 \quad \text{as } \Delta \to 0 \quad (4.24)\]

and \( \zeta(t) = \zeta^{x,\alpha}(t) \) is mean square continuous with respect to \( x \). Therefore, \( \frac{\partial}{\partial x} X(t) \) exists in the mean square sense and \( \frac{\partial}{\partial x} X(t) = \zeta(t). \)

Likewise, we can show that \( (\partial^2/\partial x^2)X^{x,\alpha}(t) \) exists in the mean square sense and is mean square continuous with respect to \( x \). The proof of the theorem is thus concluded. \( \square \)
Corollary 4.8. Under the assumptions of Theorem 4.2, the mean square derivatives \( \frac{\partial X_{x,\alpha}(t)}{\partial x} \) and \( \frac{\partial^2 X_{x,\alpha}(t)}{\partial x_j \partial x_k} \), for \( j, k = 1, \ldots, r \), are mean square continuous with respect to \( t \).

Proof. As in the proof of Theorem 4.2, we consider only the case when \( X(t) \) is real valued. Also, we use the same notations as those in the proof of Theorem 4.2. To see that \( \zeta(t) \) is continuous in the mean square sense, we first observe that for any \( t \in [0, T] \),

\[
\mathbb{E} |\zeta(t)|^2 \leq 2\mathbb{E} |\zeta(t) - Z^\Delta(t)|^2 + 2\mathbb{E} |Z^\Delta(t)|^2.
\]

It follows from (4.18), the Lipschitz condition, and Proposition 4.5 that

\[
\mathbb{E} |Z^\Delta(t)|^2 \leq 3\mathbb{E} |\phi^\Delta(t)|^2 + 3\mathbb{E} \left| \frac{1}{\Delta} \int_0^t [b(\bar{X}(u), \alpha(u)) - b(X(u), \alpha(u))] \, du \right|^2
\]

\[
+ 3\mathbb{E} \left| \frac{1}{\Delta} \int_0^t [\sigma(\bar{X}(u), \alpha(u)) - \sigma(X(u), \alpha(u))] \, d\mathbb{W}(u) \right|^2
\]

\[
\leq K + 3t \frac{1}{\Delta^2} \mathbb{E} \int_0^t |b(\bar{X}(u), \alpha(u)) - b(X(u), \alpha(u))|^2 \, du
\]

\[
+ 3\mathbb{E} \frac{1}{\Delta^2} \int_0^t |\sigma(\bar{X}(u), \alpha(u)) - \sigma(X(u), \alpha(u))|^2 \, du
\]

\[
\leq K + 3K(T + 1) \frac{1}{\Delta^2} \mathbb{E} \int_0^t |\bar{X}(u) - X(u)|^2 \, du
\]

\[
\leq K = K(x, T, K_0).
\]

That is, \( K(x, T, K_0) \) is a constant depending on \( x, T \), and \( K_0 \). Hence we have from (4.24) that

\[
\sup_{t \in [0, T]} \mathbb{E} |\zeta(t)|^2 \leq K = K(x, T, K_0) < \infty. \tag{4.25}
\]

Thus \( \zeta(t) \) is mean square continuous if we can show that

\[
\mathbb{E} |\zeta(t) - \zeta(s)|^2 \to 0 \quad \text{as } |s - t| \to 0.
\]

To this end, we note that for any \( s, t \in [0, T] \),

\[
\mathbb{E} |\zeta(t) - \zeta(s)|^2 \leq 3\mathbb{E} \left[ |\zeta(t) - Z^\Delta(t)|^2 + |\zeta(s) - Z^\Delta(s)|^2 \right.
\]

\[
+ \left. |Z^\Delta(t) - Z^\Delta(s)|^2 \right].
\]

In view of (4.24), we need only prove that

\[
\mathbb{E} |Z^\Delta(t) - Z^\Delta(s)|^2 \to 0 \quad \text{as } |s - t| \to 0.
\]
Without loss of generality, assume that $s < t$. Then by (4.4), we have

$$
E \left| Z^\Delta (t) - Z^\Delta (s) \right|^2 \leq 3 E \left| \phi^\Delta (t) - \phi^\Delta (s) \right|^2 \\
+ 3 E \left| \frac{1}{\Delta} \int_s^t b(\tilde{X}(u), \alpha(u)) - b(X(u), \alpha(u)) \, du \right|^2 \\
+ 3 E \left| \frac{1}{\Delta} \int_s^t \sigma(\tilde{X}(u), \alpha(u)) - \sigma(X(u), \alpha(u)) \, dw(u) \right|^2.
$$

(4.26)

It follows from the Cauchy–Schwartz inequality, the Lipschitz condition, and Proposition 4.5 that

$$
E \left| \frac{1}{\Delta} \int_s^t b(\tilde{X}(u), \alpha(u)) - b(X(u), \alpha(u)) \, du \right|^2 \\
\leq (t - s) \frac{1}{|\Delta|^2} E \int_s^t \left| b(\tilde{X}(u), \alpha(u)) - b(X(u), \alpha(u)) \right|^2 du \\
\leq (t - s) \frac{1}{|\Delta|^2} \mathbb{E} \int_s^t |\tilde{X}(u) - X(u)|^2 \, du \\
\leq K_0 K (t - s)^2,
$$

(4.27)

where $K_0$ is the Lipschitz constant and $K$ is a constant independent of $t$, $s$, or $\Delta$. Similarly, we can show that

$$
E \left| \frac{1}{\Delta} \int_s^t \sigma(\tilde{X}(u), \alpha(u)) - \sigma(X(u), \alpha(u)) \, dw(u) \right|^2 \\
= E \frac{1}{|\Delta|^2} \int_s^t \left| \sigma(\tilde{X}(u), \alpha(u)) - \sigma(X(u), \alpha(u)) \right|^2 du \\
\leq K_0 K (t - s).
$$

(4.28)

Next, using the same argument as that of Lemma 4.3, we can show that

$$
E |\phi^\Delta (t) - \phi^\Delta (s)|^2 \leq K (t - s).
$$

(4.29)

Thus it follows from (4.26)–(4.29) that $E|Z^\Delta (t) - Z^\Delta (s)|^2 = O(|t - s|) \rightarrow 0$ as $|t - s| \rightarrow 0$ and hence $\zeta (t)$ is mean square continuous with respect to $t$.

Likewise, we can show that $(\partial^2 / \partial x^2)^{X^\alpha(t)}\zeta^\alpha (t)$ is mean square continuous with respect to $t$. This concludes the proof. \qed
4.2. Feller continuity

Let us recall the definition of Feller continuity. Suppose that \( \xi^x(t) \) is a Markov process satisfying \( \xi(0) = x \). Denote \( u(t, x) = \mathbb{E}_x f(\xi(t)) = \mathbb{E}_0 f(\xi^x(t)) \) for some appropriate function \( f \). The process is said to be Feller if for any bounded and continuous function \( f(\cdot) \), (i) \( u(t, \cdot) \), \( t > 0 \), is continuous, and (ii) \( \lim_{t \to 0} u(t, x) = f(x) \). We show that such property carries over to the switching diffusions. That is, we need to prove that the function

\[
u(t, x, \alpha) = \mathbb{E}_{x, \alpha} f(X(t), \alpha(t)) = \mathbb{E}_x f(X^{x,\alpha}(t), \alpha^{x,\alpha}(t))
\]
is continuous with respect to the initial data \((x, \alpha)\) for any \( t \geq 0 \) and

\[
\lim_{t \downarrow 0} u(t, x, \alpha) = f(x, \alpha)
\]

for any bounded and continuous function \( f \). By virtue of [18, Lemma 3.6], the process \((X(t), \alpha(t))\) is càdlàg, that is, the sample paths of \((X(t), \alpha(t))\) are right continuous and have left limits. Therefore, it follows from the boundedness and continuity of \( f \) that \( \lim_{t \to 0} u(t, x, \alpha) = \mathbb{E}_{x, \alpha} f(X(0), \alpha(0)) = f(x, \alpha) \). Note also that \( u(0, x, \alpha) \) is automatically continuous by the continuity of \( f \). Since \( \mathcal{M} = \{1, \ldots, m_0\} \) is a finite set, it is enough to show that \( u(t, x, \alpha) \) is continuous with respect to \( x \) for any \( t > 0 \).

We show in this subsection that the Feller property of \((X(t), \alpha(t))\) follows directly from the estimate (4.20) or Corollary 4.7. The details of the proof are similar to that of, for example, [12]. Since we have done the preparation in the ground work, we will be brief, and refer to, for example [12], if the arguments of related results for diffusions can be used. Note that the recent work [14] establishes Feller property for switching diffusions using an alternative approach.

**Proposition 4.9.** Assume the conditions of Theorem 4.2. Then the switching diffusion \((X(t), \alpha(t))\) is a Feller process.

**Proof.** For any bounded and continuous function \( f(\cdot) \) and each fixed \( t \geq 0 \), let \( u(x, \alpha) = \mathbb{E}_x f(X^{x,\alpha}(t), \alpha^{x,\alpha}(t)) \). As we argued earlier, we need only show that \( u(\cdot, \alpha) \) is continuous for each \( \alpha \in \mathcal{M} \). To this end, let \( \{x_n\} \) be any sequence in \( \mathbb{R}^d \) converging to \( x \). Then by virtue of Corollary 4.7, \( (X^{x_n,\alpha}(t), \alpha^{x_n,\alpha}(t)) \to (X^{x,\alpha}(t), \alpha^{x,\alpha}(t)) \) as \( n \to \infty \) in the mean square sense. Hence we can choose a subsequence \( \{x_{n_k}\} \) of \( \{x_n\} \) converging to \( x \) such that \( (X^{x_{n_k},\alpha}(t), \alpha^{x_{n_k},\alpha}(t)) \to (X^{x,\alpha}(t), \alpha^{x,\alpha}(t)) \) a.s. as \( k \to \infty \). Then the dominated convergence theorem implies that

\[
u(x_{n_k}, \alpha) = \mathbb{E}_x f(X^{x_{n_k},\alpha}(t), \alpha^{x_{n_k},\alpha}(t)) \\
\to \mathbb{E}_x f(X^{x,\alpha}(t), \alpha^{x,\alpha}(t)) = u(x, \alpha) \quad \text{as} \ k \to \infty.
\]

Therefore, every sequence \( \{x_n\} \) converging to \( x \) has a subsequence \( \{x_{n_k}\} \) such that \( u(x, \alpha) \leq \liminf_{k \to \infty} u(x_{n_k}, \alpha) \). Hence \( u(\cdot, \alpha) \) is lower semi-continuous for each \( \alpha \in \mathcal{M} \).

Applying the above argument to \( -u(x, \alpha) = \mathbb{E}[-f(X^{x,\alpha}(t), \alpha^{x,\alpha}(t))] \), we obtain that

\[
u(x, \alpha) \geq \limsup_{k \to \infty} u(x_{n_k}, \alpha),
\]

and hence \( u(\cdot, \alpha) \) is upper semi-continuous for each \( \alpha \in \mathcal{M} \). Therefore, \( u(x, \alpha) \) is continuous with respect to \( x \). This establishes the Feller property for \((X(t), \alpha(t))\) as desired.

**Corollary 4.10.** Assume the conditions of Proposition 4.9. Then the process \((X(t), \alpha(t))\) is strong Markov.
Proof. First, note by virtue of [18, Lemma 3.6], the sample paths of \((X(t), \alpha(t))\) are right continuous and have left limits. The desired assertion then follows from [2, Theorem 2.2.4] and Proposition 4.9. □

5. System of Kolmogorov backward equations

In this section, we derive a system of the Kolmogorov backward equations. One remarkable point is that we obtain the equations without assuming nondegeneracy of the diffusion part. These results are interesting in their own right.

Theorem 5.1. Suppose that for each \(i \in \mathcal{M}\), the coefficients of (2.2) and (2.3) satisfy \(b(\cdot, i), \sigma(\cdot, i) \in C^2\), that conditions of Theorem 4.2 hold, that \(\phi(\cdot, i) \in C^2\), that \(D_x^\beta (x, i)\) is Lipschitz continuous for each \(i \in \mathcal{M}\) and \(|\theta| = 2\), and that

\[
|D_x^\beta b(x, i)| + |D_x^\beta \sigma(x, i)| + |D_x^\beta \phi(x, i)| \leq K_0 (1 + |x|^\gamma), \quad i \in \mathcal{M},
\]

where \(K_0\) and \(\gamma\) are positive constants and \(\beta\) is a multi-index with \(|\beta| \leq 2\). Then for any \(T > 0\), the function

\[
u(t, x, i) := E_{x,i} [\phi(X(t), \alpha(t))] = E[\phi(X^{x,i}(t), \alpha^{x,i}(t))]
\]

is twice continuously differentiable with respect to the variable \(x\) and satisfies

\[
|D_x^\beta \nu(t, x, i)| \leq K (1 + |x|^\gamma),
\]

for each \(i \in \mathcal{M}\), \(|\beta| \leq 2\), \(t \in [0, T]\), and \(x \in \mathbb{R}^r\).

Proof. Again, for notational simplicity, we will prove the theorem when \(X(t)\) is one-dimensional. Multi-dimensional case can be handled in a similar manner. Fix \((t, x, i) \in [0, T] \times \mathbb{R}^r \times \mathcal{M}\). For any \(0 < |\Delta| < 1\), let \(\bar{x} = x + \Delta\). As in the proof of Theorem 4.2, denote

\[
(X(t), \alpha(t)) = (X^{x,i}(t), \alpha^{x,i}(t)) \quad \text{and} \quad (\bar{X}(t), \bar{\alpha}(t)) = (X^{\bar{x},i}(t), \alpha^{\bar{x},i}(t)).
\]

By virtue of Theorem 4.2, the mean square derivative \(\xi(t) = \frac{\partial X^{x,i}(t)}{\partial x}\) exists and is mean square continuous with respect to \(x\) and \(t\). Note that

\[
\frac{u(t, \bar{x}, i) - u(t, x, i)}{\Delta} = \frac{1}{\Delta} E[\phi(\bar{X}(t), \bar{\alpha}(t)) - \phi(X(t), \alpha(t))]
\]

\[
= \frac{1}{\Delta} E[\phi(\bar{X}(t), \bar{\alpha}(t)) - \phi(\bar{X}(t), \alpha(t))] + \frac{1}{\Delta} E[\phi(\bar{X}(t), \alpha(t)) - \phi(X(t), \alpha(t))].
\]

Similar to the proof of Lemma 4.3, we can show that

\[
\frac{1}{\Delta^2} E \left[ \sup_{0 \leq s \leq T} \left| \phi(\bar{X}(s), \bar{\alpha}(s)) - \phi(\bar{X}(s), \alpha(s)) \right|^2 \right] \to 0 \quad \text{as} \quad \Delta \to 0.
\]

To proceed, for each \(i \in \mathcal{M}\), we shall use \(\phi_x(\cdot, i)\) and \(\phi_{xx}(\cdot, i)\) to denote the first and the second derivatives of \(\phi(\cdot, i)\) with respect to \(x\), respectively. We obtain
\[
\frac{1}{\Delta} \mathbb{E} \left[ \phi(\tilde{X}(t), \alpha(t)) - \phi(X(t), \alpha(t)) \right] \\
= \frac{1}{\Delta} \mathbb{E} \int_0^1 \frac{d}{dv} \phi(X(t) + v(\tilde{X}(t) - X(t)), \alpha(t)) \, dv \\
= \mathbb{E} \left[ Z(t) \int_0^1 \phi_x(X(t) + v(\tilde{X}(t) - X(t)), \alpha(t)) \, dv \right],
\]
where \( Z(t) = \frac{\tilde{X}(t) - X(t)}{\Delta} \) as defined in (4.3). Thus it follows that
\[
\begin{align*}
\left| \frac{1}{\Delta} \mathbb{E} \left[ \phi(\tilde{X}(t), \alpha(t)) - \phi(X(t), \alpha(t)) \right] - \mathbb{E} \left[ \phi_x(X(t), \alpha(t)) \right] \right| \\
\leq \mathbb{E} \left[ Z(t) \int_0^1 \phi_x(X(t) + v(\tilde{X}(t) - X(t)), \alpha(t)) \, dv - \phi_x(X(t), \alpha(t)) \, \tilde{\zeta}(t) \right] \\
\leq e_1 + e_2,
\end{align*}
\]
where
\[
e_1 := \mathbb{E} \left[ \int_0^1 \phi_x(X(t) + v(\tilde{X}(t) - X(t)), \alpha(t)) \, dv - \phi_x(X(t), \alpha(t)) \right] Z(t),
\]
\[
e_2 := \mathbb{E} \left[ \phi_x(X(t), \alpha(t)) \left[ Z(t) - \tilde{\zeta}(t) \right] \right],
\]
with \( \zeta(t) \) given in (4.23). It follows from (5.1), Lemma 3.2, and (4.24) that
\[
e_2 = \mathbb{E} \left[ \phi_x(X(t), \alpha(t)) \left[ Z(t) - \tilde{\zeta}(t) \right] \right] \\
\leq \mathbb{E}^{1/2} \left| \phi_x(X(t), \alpha(t)) \right|^2 \mathbb{E}^{1/2} \left[ Z(t) - \tilde{\zeta}(t) \right]^2 \\
\leq K \mathbb{E}^{1/2} \left| Z(t) - \tilde{\zeta}(t) \right|^2 \rightarrow 0 \quad \text{as} \ \Delta \rightarrow 0.
\]
To estimate the term \( e_1 \), we note that (5.1) and Lemma 3.2 imply that
\[
\mathbb{E} \left[ \phi_x(X(t) + v(\tilde{X}(t) - X(t)), \alpha(t)) - \phi_x(X(t), \alpha(t)) \right]^2 \leq K
\]
for all \( 0 < |\Delta| < 1 \). By virtue of the argument in the proof of Theorem 4.2, \( \tilde{X}(t) \rightarrow X(t) \) in probability as \( \Delta \rightarrow 0 \) for any \( t \in [0, T] \). Thus it follows that
\[
\mathbb{E} \left[ \phi_x(X(t) + v(\tilde{X}(t) - X(t)), \alpha(t)) - \phi_x(X(t), \alpha(t)) \right]^2 \rightarrow 0 \quad \text{as} \ \Delta \rightarrow 0.
\]
Note that we proved in Corollary 4.8 that \( \mathbb{E}[Z(t)]^2 \leq K \). Then we have from the Cauchy–Schwartz inequality that
where

It follows that

Moreover, (5.1), Lemma 3.2, and (4.25) imply that for some

Therefore it follows from (5.3), (5.4), and (5.5) that

Hence we have shown that as \( \Delta \to 0 \),

Therefore it follows from (5.3), (5.4), and (5.5) that

This reveals that \( u(t, \cdot, i) \) is differentiable with respect to the variable \( x \) and

Moreover, (5.1), Lemma 3.2, and (4.25) imply that for some \( K > 0 \), we have

where \( \gamma_0 > 2 \) is as given in the proof of Lemma 4.3. Next, we verify that \( \frac{\partial u(t, x, i)}{\partial x} \) is continuous with respect to \( x \). To this end, we consider

where \( \tilde{\zeta}(t) = \tilde{x}_i(t) = \frac{\partial X_i(t)}{\partial x} \). By virtue of Theorem 4.2, \( \zeta(t) = \frac{\partial X(t)}{\partial x} \) is mean square continuous. Hence it follows that

Meanwhile, detailed calculations similar to those in deriving (5.4) lead to
differentiable with respect to the variable \( t \). Move over, \( u \) satisfies the system of Kolmogorov backward equations with initial condition \( (5.7) \).

Consequently, we can verify that

\[
\frac{\partial^2 u(t, x, i)}{\partial x^2} = E_{x, i} \left[ \phi(x) X(t, \alpha(t)) \left( \frac{\partial X(t)}{\partial x} \right)^2 + \phi(x) X(t, \alpha(t)) \frac{\partial^2 X(t)}{\partial x^2} \right].
\]

Consequently, we can verify that

\[
\left| \frac{\partial^2 u(t, x, i)}{\partial x^2} \right| \leq K \left( 1 + |x|^{\gamma} \right).
\]

This completes the proof of the theorem. \( \square \)

**Theorem 5.2.** Assume the conditions of Theorem 5.1. Then the function \( u \) defined in (5.2) is continuously differentiable with respect to the variable \( t \). Moreover, \( u \) satisfies the system of Kolmogorov backward equations

\[
\frac{\partial u(t, x, i)}{\partial t} = \mathcal{L} u(t, x, i), \quad (t, x, i) \in (0, T] \times \mathbb{R}^r \times \mathcal{M},
\]

with initial condition

\[
\lim_{\tau \downarrow 0} u(t, x, i) = \phi(x, i), \quad (x, i) \in \mathbb{R}^r \times \mathcal{M},
\]

where \( \mathcal{L} u(t, x, i) \) in (5.6) is to be interpreted as \( \mathcal{L} \) applied to the function \((x, i) \mapsto u(t, x, i)\).

**Proof.** First note that by virtue of [18, Lemma 3.6], the process \((X(t), \alpha(t))\) is càdlàg. Hence the initial condition (5.7) follows from the continuity of \( \phi \). We will divide the rest of the proof into several steps.

**Step 1.** For fixed \((x, i) \in \mathbb{R}^r \times \mathcal{M}, u(t, x, i)\) is absolutely continuous with respect to \( t \in [0, T] \). In fact, for any \( 0 \leq s \leq t \leq T \), by virtue of Dynkin’s formula,

\[
u(t, x, i) - u(s, x, i) = E_{x, i} \phi(X(t), \alpha(t)) - E_{x, i} \phi(X(s), \alpha(s))
\]

\[
= E_{x, i} \left[ E_{x, i} \left[ (\phi(X(t), \alpha(t)) - \phi(X(s), \alpha(s))) \right] \bigg| \mathcal{F}_s \right]
\]

\[
= E_{x, i} \left[ \int_s^t \mathcal{L} \phi(X(v), \alpha(v)) \, dv \bigg| \mathcal{F}_s \right]
\]

\[
\leq \int_s^t E_{x, i} \left[ |\mathcal{L} \phi(X(v), \alpha(v))| \bigg| \mathcal{F}_s \right] \, dv.
\]
Using (5.1), we can verify that for some positive constants $K$ and $\gamma_0$,

$$|\mathcal{L}\phi(x, i)| \leq K (1 + |x|^{\gamma_0}) \text{ for all } (x, i) \in \mathbb{R}^r \times M.$$ 

Hence it follows from Lemma 3.2 that

$$E_{x, i} \left[ |\mathcal{L}\phi(X(v), \alpha(v))| \right| \mathcal{F}_s ] \leq K E_{x, i} \left[ (1 + |X(v)|^{\gamma_0}) \right| \mathcal{F}_s ] \leq K,$$

where $K$ is independent of $t$, $s$, or $v$. Thus we have

$$|u(t, x, i) - u(s, x, i)| \leq K |t - s|.$$ 

This shows that $u$ is absolutely continuous with respect to $t$. Hence $\frac{\partial u(t, x, i)}{\partial t}$ exists a.s. on $[0, T]$ and we have

$$u(t, x, i) = u(0, x, i) + \int_0^t \frac{\partial u(v, x, i)}{\partial v} dv.$$ (5.8)

**Step 2.** For any $h > 0$, we have from the strong Markov property that

$$u(t + h, x, i) = E_{x, i} \phi(X(t + h), \alpha(t + h))$$

$$= E_{x, i} \left[ E_{x, i} \left[ \phi(X(t + h), \alpha(t + h)) \right| \mathcal{F}_h ] \right]$$

$$= E_{x, i} \left[ E_{x(h), \alpha(h)} \phi(X(t + h), \alpha(t + h)) \right]$$

$$= E_{x, i} u(t, X(h), \alpha(h)).$$ (5.9)

Now let $g(x, i) := u(t, x, i)$. Then Theorem 5.1 implies that $g(\cdot, i) \in C^2$ for each $i \in M$ and for some $K > 0$ and $\gamma_0 > 0$,

$$|D^\beta g(x, i)| \leq K (1 + |x|^{\gamma_0}), \quad i \in M.$$ 

Thus it follows from Dynkin's formula that

$$E_{x, i} g(X(h), \alpha(h)) - g(x, i) = E_{x, i} \int_0^h \mathcal{L}g(X(v), \alpha(v)) dv.$$ 

Using the same argument as in the proof of [2, Theorem 5.6.1], we can show that

$$\frac{1}{h} E_{x, i} \int_0^h \mathcal{L}g(X(v), \alpha(v)) dv \rightarrow \mathcal{L}g(x, i) \quad \text{as } h \downarrow 0.$$ (5.10)

Therefore it follows that

$$\lim_{h \downarrow 0} \frac{E_{x, i} g(X(h), \alpha(h)) - g(x, i)}{h} = \mathcal{L}g(x, i).$$
But by the definition of \( g \), we have from (5.9) that

\[
\lim_{h \to 0} \frac{u(t + h, x, i) - u(t, x, i)}{h} = \mathcal{L}g(x, i) = \mathcal{L}u(t, x, i). \tag{5.11}
\]

Thus a combination of (5.8) and (5.11) leads to

\[
u(t, x, i) = u(0, x, i) + \int_0^t \mathcal{L}u(v, x, i) \, dv. \tag{5.12}
\]

**Step 3.** We claim that \( \mathcal{L}u(t, x, i) \) is continuous with respect to the variable \( t \). Note that

\[
\mathcal{L}u(t, x, i) = b(x, i) \frac{\partial u(t, x, i)}{\partial x} + \frac{1}{2}\sigma^2(x, i) \frac{\partial^2 u(t, x, i)}{\partial x^2} + \sum_{j=1}^{m_0} q_{ij}(x)u(t, x, j).
\]

Thus the claim will be true if we can show that \( \frac{\partial u(t, x, i)}{\partial x} \) and \( \frac{\partial^2 u(t, x, i)}{\partial x^2} \) are continuous with respect to \( t \), since Step 1 above shows that \( u(t, x, i) \) is continuous with respect to \( t \). To this end, let \( t, s \in [0, T] \). Then we have

\[
\left| \frac{\partial u(t, x, i)}{\partial x} - \frac{\partial u(s, x, i)}{\partial x} \right| = \left| \mathbb{E}_{x,i} \left[ \phi_x(X(t), \alpha(t)) \xi(t) \right] - \mathbb{E}_{x,i} \left[ \phi_x(X(s), \alpha(s)) \xi(s) \right] \right|
\]

\[
\leq \mathbb{E}_{x,i} \left| \phi_x(X(t), \alpha(t)) \xi(t) - \phi_x(X(s), \alpha(s)) \xi(s) \right| \leq \mathbb{E}_{x,i} \left| \phi_x(X(t), \alpha(t)) \right| \left| \xi(t) - \xi(s) \right|
\]

\[
= \mathbb{E}_{x,i}^{1/2} \left| \phi_x(X(t), \alpha(t)) \right| \left| \phi_x(X(s), \alpha(s)) \right| \mathbb{E}_{x,i}^{1/2} \left| \xi(t) \right|^2 + \mathbb{E}_{x,i}^{1/2} \left| \phi_x(X(s), \alpha(s)) \right| \mathbb{E}_{x,i}^{1/2} \left| \xi(t) - \xi(s) \right|^2.
\]

As demonstrated before, \( \mathbb{E}_{x,i}^{1/2} \left| \phi_x(X(s), \alpha(s)) \right|^2 \leq K \). Corollary 4.8 implies that \( \xi(t) \) is mean square continuous with respect to \( t \). Hence it follows that

\[
\mathbb{E}_{x,i}^{1/2} \left| \xi(t) - \xi(s) \right|^2 \to 0 \quad \text{as} \quad |t - s| \to 0.
\]

Meanwhile,

\[
\mathbb{E}_{x,i} \left| \phi_x(X(t), \alpha(t)) - \phi_x(X(s), \alpha(s)) \right|^2 \leq e_1 + e_2,
\]

where

\[
e_1 := K \mathbb{E}_{x,i} \left| \phi_x(X(t), \alpha(t)) \right|^2,
\]

\[
e_2 := K \mathbb{E}_{x,i} \left| \phi_x(X(s), \alpha(t)) \right|^2.
\]
Using (3.7) and (4.25), detailed computations show that
\[
e_1 \leq KE_{x,i} \left| \frac{1}{0} \phi_{x\alpha}(X(s) + v(X(t) - X(s)), \alpha(t)) \right| dv(X(t) - X(s)) \right| ^2 \\
\rightarrow 0 \text{ as } |t - s| \rightarrow 0.
\]

To treat the term \( e_2 \), we assume without loss of generality that \( t > s \). We obtain

\[
e_2 = KE_{x,i} |\phi_x(X(s), \alpha(t)) - \phi_x(X(s), \alpha(s))|^2 \\
= \sum_{i=1}^{m_0} \sum_{j \neq i} E_{x,i} |\phi_x(X(s), j) - \phi_x(X(s), i)|^2 \mathbb{I}_{\{\alpha(t) = j\}} \mathbb{I}_{\{\alpha(s) = i\}} \\
= \sum_{i=1}^{m_0} \sum_{j \neq i} E_{x,i} \left[ |\phi_x(X(s), j) - \phi_x(X(s), i)|^2 \mathbb{I}_{\{\alpha(s) = i\}} \mathbb{I}_{\{\alpha(t) = j\}} \mathbb{F}_s \right] \\
= \sum_{i=1}^{m_0} \sum_{j \neq i} E_{x,i} \left[ |\phi_x(X(s), j) - \phi_x(X(s), i)|^2 \mathbb{I}_{\{\alpha(s) = i\}} q_{ij}(X(s))(t - s) + o(t - s) \right] \\
\leq K(t - s).
\]

Thus it follows that \( e_2 \rightarrow 0 \text{ as } |t - s| \rightarrow 0 \). Hence we have shown that

\[
\left| \frac{\partial u(t, x, i)}{\partial x} - \frac{\partial u(s, x, i)}{\partial x} \right| \rightarrow 0 \text{ as } |t - s| \rightarrow 0
\]

and so \( \frac{\partial u(t, x, i)}{\partial x} \) is continuous with respect to the variable \( t \). Similarly, we can show that \( \frac{\partial^2 u(t, x, i)}{\partial x^2} \) is also continuous with respect to the variable \( t \). Therefore \( Lu(t, x, i) \) is continuous with respect to the variable \( t \).

Step 4. Finally, by virtue of (5.12) and Step 3 above, we conclude that \( \frac{\partial u(t, x, i)}{\partial t} \) exists everywhere for \( t \in (0, T) \) and that

\[
\frac{\partial u(t, x, i)}{\partial t} = Lu(t, x, i).
\]

This finishes the proof of the theorem. \( \square \)

6. Remarks on rates of convergence of numerical algorithms

Since stochastic differential equations with switching are more complex than stochastic differential equations without switching, closed-form solutions are difficult to obtain. As a consequence, numerical approximation is a viable alternative. In a recent paper, we proposed an algorithm for solving (2.2) together with (2.3). One of the main difficulties is that due to the continuous-state dependence, \( \alpha(t) \) and \( X(t) \) are dependent; \( \alpha(t) \) is non-Markovian. The essence in our approach is to treat the pair of processes \( (X(t), \alpha(t)) \) jointly by noting the two-component process being Markovian. The algorithm proposed is of the form

\[
X_{n+1} = X_n + \epsilon b(X_n, \alpha_n) + \sqrt{\epsilon} \sigma(X_n, \alpha_n) \xi_n.
\]

(6.1)
Here $\alpha_n$ is a discrete-time stochastic process so that when $X_{n-1} = x$, $\alpha_n$ has a transition probability matrix $\exp(Q(x)e)$, and $\{\xi_n\}$ is a sequence of random variables satisfying
\[
\sqrt{\varepsilon} \xi_n = w(\varepsilon(n + 1)) - w(\varepsilon n),
\]
where $w(\cdot)$ is the standard Brownian motion. Clearly, $\xi_n$ is independent of the $\sigma$-algebra $G_n$ generated by $\{X_k, \alpha_k: k \leq n\}$, $E[\xi_n] = 0$, $E[\xi_n^p] < \infty$ for $p \geq 2$, and $E[\xi_n^p] = 1$. Define a piecewise constant interpolation $X^\varepsilon(t) = X_k$ and $\alpha^\varepsilon(t) = \alpha_k$ for $t \in [sk, sk + \varepsilon)$. In [15], weak convergence of the algorithm was established using a martingale problem formulation and $L^2$ convergence was also obtained. It was shown that $(X^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$ converges weakly to $(X(\cdot), \alpha(\cdot))$ that is the solution of (2.2) and (2.3). Moreover,

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

Since we are dealing with a numerical algorithm, it is desirable to have estimation error bounds and the rates of convergence. This section takes up this issue. In the original definition of Kloeden and Platen [9, p. 323], the rate of convergence was defined as follows. For a finite time $T > 0$, if there exists a positive constant $K$ that does not depend on $\varepsilon$ such that $E[|X^\varepsilon(T) - X(T)|] \leq K\varepsilon^\gamma$ for some $\gamma > 0$ then the discrete approximation $X^\varepsilon$ is said to converge strongly to $X$ with order $\gamma$. Here we adopt the more recent approach as in [11] to require the rate be defined uniformly for $t \in [0, T]$. We thus concentrate on error bounds for $E\sup_{0 \leq t \leq T} |X^\varepsilon(t) - X(t)|^2$.

We assume the conditions of Proposition 2.1 and the conditions for $\{\xi_n\}$ and $\alpha_n$ are satisfied. It is straightforward that the piecewise constant interpolation of (6.1) leads to

\[
X^\varepsilon(t) = X_0 + \int_0^t b(X^\varepsilon(s), \alpha^\varepsilon(s)) \, ds + \int_0^t \sigma(X^\varepsilon(s), \alpha^\varepsilon(s)) \, dw(s). \tag{6.2}
\]

The representation (6.2) enables us to compare the solution (2.2) with that of the discrete iterations. Comparing the interpolation of the iterates and the solution of (2.2) and (2.3), we obtain

\[
E \sup_{0 \leq t \leq T} \left| X^\varepsilon(t) - X(t) \right|^2 \\
\leq 2E \sup_{0 \leq t \leq T} \left| \int_0^t [b(X(s), \alpha(s)) - b(X^\varepsilon(s), \alpha^\varepsilon(s))] \, ds \right|^2 \\
+ 2E \sup_{0 \leq t \leq T} \left| \int_0^t [\sigma(X(s), \alpha(s)) - \sigma(X^\varepsilon(s), \alpha^\varepsilon(s))] \, dw \right|^2 \\
\leq 2TE \int_0^T |b(X(s), \alpha(s)) - b(X^\varepsilon(s), \alpha^\varepsilon(s))|^2 \, ds \\
+ 8E \int_0^T |\sigma(X(s), \alpha(s)) - \sigma(X^\varepsilon(s), \alpha^\varepsilon(s))|^2 \, ds. \tag{6.3}
\]

Note that in (6.3), the first inequality is obtained from the familiar inequality $(a + b)^2 \leq 2(a^2 + b^2)$ for two real numbers $a$ and $b$. The first term on the right side of the second inequality follows
from Hölder’s inequality, and the second term is a consequence of the well-known Doob’s martingale inequality [11, p. 18]. To proceed, we treat the drift and diffusion terms separately. Note that

\[
\mathbb{E} \int_0^T \left| b(X(s), \alpha(s)) - b(X^\varepsilon(s), \alpha(s)) \right|^2 ds 
\]

\[
\leq \mathbb{E} \int_0^T \left| b(X(s), \alpha(s)) - b(X^\varepsilon(s), \alpha(s)) \right|^2 ds
\]

\[
+ \mathbb{E} \int_0^T \left| b(X^\varepsilon(s), \alpha(s)) - b(X^\varepsilon(s), \alpha^\varepsilon(s)) \right|^2 ds
\]

\[
\leq K \int_0^T \mathbb{E} \left| X^\varepsilon(s) - X(s) \right|^2 ds
\]

\[
+ \mathbb{E} \int_0^T \left[ 1 + \left| X^\varepsilon(s) \right| \right] I_{\{\alpha(s) \neq \alpha^\varepsilon(s)\}} ds.
\] (6.4)

The first inequality in (6.4) follows from the familiar triangle inequality, and the second inequality is a consequence of the Lipschitz continuity, the Cauchy inequality, and the linear growth condition. We now concentrate on the last term in (6.4). Noting the discrete iteration, we have

\[
\mathbb{E} \int_0^T \left[ 1 + \left| X^\varepsilon(s) \right| \right] I_{\{\alpha(s) \neq \alpha^\varepsilon(s)\}} ds = \sum_{k=0}^{\lfloor t/\varepsilon \rfloor - 1} \mathbb{E} \int_{\varepsilon k}^{\varepsilon k + \varepsilon} \left[ 1 + \left| X^\varepsilon(s) \right| \right] I_{\{\alpha(s) \neq \alpha^\varepsilon(s)\}} ds,
\]

where \( \lfloor t/\varepsilon \rfloor \) denotes the integer part of \( t/\varepsilon \). Using nested conditioning, we further obtain

\[
\mathbb{E} \int_{\varepsilon k}^{\varepsilon k + \varepsilon} \left[ 1 + \left| X^\varepsilon(s) \right| \right] I_{\{\alpha(s) \neq \alpha^\varepsilon(s)\}} ds
\]

\[
= \mathbb{E} \int_{\varepsilon k}^{\varepsilon k + \varepsilon} \left[ 1 + |X_k|^2\right] \mathbb{E}[I_{\{\alpha(s) \neq \alpha^\varepsilon(s)\}} | \mathcal{F}_{\varepsilon k}] ds
\]

\[
= \mathbb{E} \left( \left[ 1 + |X_k|^2\right] \int_{\varepsilon k}^{\varepsilon k + \varepsilon} \sum_{i \in \mathcal{M}} I_{\{\alpha^\varepsilon = i\}} \sum_{j \neq i} q_{ij}(X(\varepsilon k))(s - \varepsilon k) + o(s - \varepsilon k) ds \right)
\]

\[
\leq K \varepsilon \int_{\varepsilon k}^{\varepsilon k + \varepsilon} ds \leq K \varepsilon^2.
\]

Thus the moment estimate of \( \mathbb{E}|X(t)|^2 \) yields that
\[
E \int_0^T |b(X(s), \alpha(s)) - b(X^\varepsilon(s), \alpha(s))|^2 \, ds \\
\leq K \varepsilon + K \int_0^T E |X^\varepsilon(s) - X(s)|^2 \, ds.
\]  
(6.5)

Likewise, for the term involving diffusion, we also obtain
\[
E \int_0^T |\sigma(X(s), \alpha(s)) - \sigma(X^\varepsilon(s), \alpha(s))|^2 \, dw \\
\leq K \varepsilon + K \int_0^T E |X^\varepsilon(s) - X(s)|^2 \, ds.
\]  
(6.6)

Using (6.3)–(6.6), we obtain
\[
E \sup_{0 \leq t \leq T} |X^\varepsilon(t) - X(t)|^2 \\
\leq K \varepsilon + \int_0^T E \sup_{0 \leq v \leq s} |X^\varepsilon(v) - X(v)|^2 \, ds.
\]  
(6.7)

An application of Gronwall’s inequality leads to
\[
E \sup_{0 \leq t \leq T} |X^\varepsilon(t) - X(t)|^2 \leq K \varepsilon.
\]

Thus, we conclude that the sequence of discrete iterates converges strongly in the $L^2$ sense with an error bound $O(\varepsilon)$. We state it as a result below.

**Theorem 6.1.** Assume the conditions of Theorem 4.2. Then $(X^\varepsilon(\cdot), \alpha^\varepsilon(\cdot))$, the interpolation of the approximation sequence, converges to $(X(\cdot), \alpha(\cdot))$ in the sense
\[
E \sup_{0 \leq t \leq T} |X^\varepsilon(t) - X(t)|^2 \to 0 \quad \text{as } \varepsilon \to 0.
\]  
(6.8)

Moreover, we have the following rate of convergence estimate
\[
E \sup_{0 \leq t \leq T} |X^\varepsilon(t) - X(t)|^2 = O(\varepsilon).
\]  
(6.9)
References