SMOOTHNESS OF DENSITY FOR STOCHASTIC DIFFERENTIAL EQUATIONS WITH MARKOVIAN SWITCHING

YAOZHONG HU
Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, Alberta, Canada T6G 2G1

DAVID NUALART∗
Department of Mathematics
University of Kansas
Lawrence, Kansas, 66045, USA

XIAOBIN SUN AND YINGCHAO XIE
School of Mathematics and Statistics
Jiangsu Normal University
Xuzhou 221116, China

(Communicated by Tomas Caraballo)

Abstract. This paper is concerned with a class of stochastic differential equations with Markovian switching. The Malliavin calculus is used to study the smoothness of the density of the solution under a Hörmander type condition. Furthermore, we obtain a Bismut type formula which is used to establish the strong Feller property.

1. Introduction. This paper considers the following stochastic differential equation with Markovian switching on $\mathbb{R}^n$:

$$dX_t = b(X_t, \alpha_t)dt + \sigma(X_t, \alpha_t)dW_t, \quad (X_0, \alpha_0) = (x, \alpha) \in \mathbb{R}^n \times S,$$

where $S = \{1, 2, \ldots, m_0\}$ and $\{\alpha_t, t \geq 0\}$ is a right-continuous $S$-valued Markov chain described by

$$\mathbb{P}\{\alpha_{t+\Delta} = j|\alpha_t = i\} = \begin{cases} q_{ij}\Delta + o(\Delta), & i \neq j, \\ 1 + q_{ii}\Delta + o(\Delta), & i = j, \end{cases}$$

and $Q = (q_{ij})_{1 \leq i, j \leq m_0}$ is a $Q$-matrix.

Stochastic differential equations (SDEs) of this type have been extensively studied (see, for instance [1, 4, 6, 7]). Existence and uniqueness of a solution, existence of an invariant measure, stability and other important properties have been analyzed.
In this paper we first study the smoothness of the density of the solution, then establish a Bismut type formula. Finally, as an application, we prove the strong Feller property.

Our first purpose is to show the differentiability in the sense of Malliavin calculus of the solution $X_t$ to equation (1). The appearance of the Markovian switching term $\alpha_t$, which is a jump process, creates some difficulties in the study of the Malliavin derivative of $X_t$. We will perform perturbations of the underlying Brownian motion, keeping the Markovian switching process $\alpha_t$ unperturbed. The technique for this analysis is inspired in the partial Malliavin calculus, which can be regarded as a stochastic calculus of variation for random variables with values in a Hilbert space.

After developing the Malliavin calculus for the solution $X_t$, we investigate the smoothness of the density of the law of the random vector $X_t$, for a fixed $t > 0$, with respect to the Lebesgue measure. For this we need to show that the determinant of the Malliavin matrix of $X_t$ has all negative finite moments. In the classical diffusion case, this is guaranteed by a Hörmander type nondegeneracy condition. To follow this classical approach we immediately encounter a difficulty: the process $X_t$ depends on the discrete process $\alpha_t$, and the application of Itô’s formula yields some jump terms. To overcome this difficulty we shall use the following strategy inspired by [2]. First we notice that the jump times form a subset of the jump times of some Poisson process inspired by [2]. Then conditioning on $N_t = k$, there exists a random interval $[T_1, T_2]$ such that $T_2 - T_1 \geq \frac{1}{k+t}$. On this random time interval, the Itô’s formula for $(X_t, \alpha_t)$ will not produce a jump term, and we can apply the classical procedure. This requires a version of the classical Norris lemma on time intervals.

We also use Malliavin calculus to establish a Bismut type formula for the transition semigroup of the Markov process $(X_t, \alpha_t)$. As an application of Bismut type formula, we derive the strong Feller property of the process $(X_t, \alpha_t)$.

The paper is organized as follows. In the next section, we introduce some notation and assumptions that we use throughout the paper. We develop the Malliavin calculus for SDEs with Markovian switching in Section 3. In Section 4, we show that the determinant of the Malliavin covariance matrix has all negative finite moments under a suitable uniform Hörmander’s condition. Finally, we establish the Bismut type formula and use it to prove the strong Feller property in Section 5.

2. Preliminaries. Let $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ be the $d$-dimensional canonical Wiener space with the natural filtration $\mathcal{F}_1 = \{\mathcal{F}_1(t), t \geq 0\}$. That is, $\Omega_1$ is the set of all continuous maps $\omega_1 : \mathbb{R}_+ \to \mathbb{R}^d$ such that $\omega_1(0) = 0$ and $\mathcal{F}_1$ is the completion of the Borel $\sigma$ field of $\Omega_1$ with respect to $\mathbb{P}_1$, where $\mathbb{P}_1$ is the canonical Wiener measure. Then, $W = \{W_t(\omega_1) := \omega_1(t), t \geq 0\}$ is a $d$-dimensional Brownian motion.

Let $\mathbb{S} = \{1, 2, \ldots, m_0\}$, where $m_0$ is a given positive integer which will be fixed throughout the paper. Let $Q = (q_{ij})_{1 \leq i,j \leq m_0}$ be a $Q$-matrix satisfying the following assumption:

(i) $q_{ij} \geq 0$ for $i \neq j$,
(ii) $q_{ii} = -\sum_{j \neq i} q_{ij}$ for $i \in \mathbb{S}$,
(iii) $\sup_{i,j \in \mathbb{S}} |q_{ij}| := K$.

Let $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ be another complete probability space with a filtration $\mathcal{F}_2 = \{\mathcal{F}_2(t), t \geq 0\}$ satisfying the usual conditions, on which there exists a right-continuous $\mathbb{S}$-valued Markov chain $\alpha = \{\alpha_t, t \geq 0\}$ satisfying (2).
Denote the product probability space by \((\Omega, \mathcal{F}, \mathbb{P}) := (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)\) with the product filtration \(\mathcal{F} = \{\mathcal{F}_t, t \geq 0\}\), where \(\mathcal{F}_t = \mathcal{F}_1(t) \times \mathcal{F}_2(t)\). We extend \(W_t\) and \(\alpha_t\) to random variables defined on \(\Omega\) by letting \(W_t(\omega) = \omega(1)\) and \(\alpha_t(\omega) = \alpha_t(\omega_2)\), respectively, if \(\omega = (\omega_1, \omega_2)\). Notice that on the probability space \((\Omega, \mathcal{F}, \mathbb{P})\), the processes \(W\) and \(\alpha\) are independent.

It is well known (see [1]) that the process \(\alpha\) can be described in the following manner. Introduce the function \(g : S \times [0, m_0(m_0 - 1)K] \rightarrow \mathbb{R}\) defined by

\[
g(i, z) = \sum_{j \in S \setminus i} (j - i) 1_{x \in \triangle_{ij}}, \quad i \in S,
\]

where \(\triangle_{ij}\) are the consecutive (with respect to the lexicographic ordering on \(S \times S\)) left-closed, right-open intervals of \(\mathbb{R}_+\), each having length \(q_{ij}\), with \(\triangle_{12} = [0, q_{12}]\). Then, equation (2) can also be written as

\[
da(t) = \int_{[0, m_0(m_0 - 1)K]} g(\alpha_{t-}, z) N(dt, dz),
\]

where \(N(dt, dz)\) is a Poisson random measure defined on \(\Omega \times \mathcal{B}(\mathbb{R}_+) \times \mathcal{B}(\mathbb{R}_+)\), whose intensity measure is Lebesgue measure, and \(N(dt, dz)\) is independent of \(W\).

For \(k \in \mathbb{N}\) we denote by \(C^k(\mathbb{R}^n \times S; \mathbb{R}^n)\) the family of all \(\mathbb{R}^n\)-valued functions \(f(x, \alpha)\) on \(\mathbb{R}^n \times S\) which are \(k\)-times continuously differentiable in \(x\) for any \(\alpha \in S\). The \(k\)-th derivative tensor of \(f\) with respect to \(x\) is denoted by \(\nabla^k f(x, \alpha)\).

We denote by \(|\cdot|\) the Euclidean norm and consider the metric \(\Lambda\) on \(\mathbb{R}^n \times S\) given by \(\Lambda((x, i), (y, j)) = |x - y| + d(i, j)\), for \(x, y \in \mathbb{R}^n, i, j \in S\), where \(d(i, j) = 0\) if \(i = j\) and \(d(i, j) = 1\) if \(i \neq j\). Let \(\mathcal{B}_0(\mathbb{R}^n \times S)\) be the family of all bounded Borel measurable functions on \(\mathbb{R}^n \times S\).

Suppose that \(b : \mathbb{R}^n \times S \rightarrow \mathbb{R}^n\) and \(\sigma : \mathbb{R}^n \times S \rightarrow \mathbb{R}^{nd}\) are functions satisfying the following assumptions:

\(\text{(A}_1\text{)}\) There is a positive constant \(C_1\) such that

\[|b(x, \alpha) - b(y, \alpha)| \vee |\sigma(x, \alpha) - \sigma(y, \alpha)| \leq C_1|x - y|\]

for any \(x, y \in \mathbb{R}^n, \alpha \in S\).

\(\text{(H}_1\text{)}\) Fix an integer \(k \geq 1\). For each \(l = 1, \ldots, d\), the functions \(b\) and \(\sigma_l\) belong to \(C^k(\mathbb{R}^n \times S; \mathbb{R}^n)\), and have bounded partial derivatives up to the order \(k\), where \(\sigma_l\) is the \(l\)-th column of \(\sigma\).

\(\text{(H}_k\text{)}\) For each \(l = 1, \ldots, d\) and for any \(k \geq 1\), the functions \(b\) and \(\sigma_l\) belong to \(C^k(\mathbb{R}^n \times S; \mathbb{R}^n)\), and have bounded partial derivatives of all orders.

It is clear that \(\text{(H}_k\text{)}\) implies \(\text{(A}_1\text{)}\) for any \(k \geq 1\).

Under condition \(\text{(A}_1\text{)}\), equation (1) possesses a unique strong solution \(X = \{X_t, t \geq 0\}\). Moreover, for any \(p \geq 2\) and \(T > 0\), \(\mathbb{E}[\sup_{0 \leq t \leq T} |X_t|^p] \leq C_2\), where \(C_2\) is a positive constant depending only on \(p, T\) and \(x\). On the other hand, \((X, \alpha) = \{(X_t, \alpha_t), t \geq 0\}\) is an homogeneous Markov process and the associated Markov semigroup \(P_t\) satisfies

\[P_tf(x, \alpha) = \mathbb{E}f(X_t(x, \alpha, \alpha_t(x, \alpha)), \quad t \geq 0, f \in \mathcal{B}_0(\mathbb{R}^n \times S).
\]

We refer, for instance, to [4] and [7] for a detailed presentation and proofs of the above results.

Along the paper \(C\) will denote a generic constant which may vary from line to line and it might depend on \(T\), the exponent \(p \geq 2\), the initial condition \(x\) and a fixed element \(h \in H\) (the precise definition of \(H\) is in the next section).
3. The Malliavin calculus. In this section we analyze the regularity, in the sense of Malliavin calculus, of the solution \( X_t \) to equation (1). The procedure is to perform perturbations of the underlying Brownian motion, keeping the Markovian switching process \( \alpha_t \) invariant. The technique for this analysis is inspired in the partial Malliavin calculus which can be regarded as a stochastic calculus of variation for random variables with values on the Hilbert space \( L^2(\Omega_2) \).

We follow the notation introduced in the Preliminaries. Denote by \( H \) the Hilbert space \( H = L^2(\mathbb{R}_+; \mathbb{R}^d) \), equipped with the inner product \( \langle h_1, h_2 \rangle_H = \int_0^\infty (h_1(s), h_2(s))_{\mathbb{R}^d} ds \).

For a Hilbert space \( U \) and a real number \( p \geq 1 \), we denote by \( L^p(\Omega_1; U) \) the space of \( U \)-valued random variables \( \xi \) such that \( \mathbb{E}_1 \| \xi \|_U^p < \infty \), where \( \mathbb{E}_1 \) is the mathematical expectation on the probability space \((\Omega_1, \mathcal{F}_1, \mathbb{P}_1)\). We also set \( L^\infty(\Omega_1; U) := \cap_{p<\infty} L^p(\Omega_1; U) \).

We introduce the derivative operator for a random variable \( F \) in the space \( L^\infty(\Omega_1; U) \) following the approach of Malliavin in [3]. We say that \( F \) belongs to \( \mathbb{D}^{1,\infty}(U) \) if there exists \( DF \in L^\infty(\Omega_1; H \otimes U) \) such that for any \( h \in H \),

\[
\lim_{\varepsilon \to 0} \mathbb{E}_1 \left| \frac{F(\omega_1 + \varepsilon \int_0^t h_s ds)}{\varepsilon} - \langle DF, h \rangle_H \right|_U^p = 0
\]

holds for every \( p \geq 1 \). In this case, we define the Malliavin derivative of \( F \) in the direction \( h \) by \( D^hF := \langle DF, h \rangle_H \). Then, for any \( p \geq 1 \) we define the Sobolev space \( \mathbb{D}^{1,p}(U) \) as the completion of \( \mathbb{D}^{1,\infty}(U) \) under the following norm

\[
\| F \|_{1,p,U} = \left[ \mathbb{E}_1 \| F \|_U^p \right]^{1/p} + \left[ \mathbb{E}_1 \| DF \|_{H \otimes U}^p \right]^{1/p}.
\]

By induction we define the \( k \)-th derivative by \( D^kF = D(D^{k-1}F) \), which is a random element with values in \( H^\otimes k \otimes U \). For any integer \( k \geq 1 \), the Sobolev space \( \mathbb{D}^{k,p}(U) \) is the completion of \( \mathbb{D}^{k,\infty}(U) \) under the norm

\[
\| F \|_{k,p,U} = \| F \|_{k-1,p,U} + \| D^kF \|_{1,p,H^\otimes k \otimes U}.
\]

It turns out that \( D \) is a closed operator from \( L^p(\Omega_1; U) \) to \( L^p(\Omega_1; H \otimes U) \). Its adjoint \( \delta \) is called the divergence operator, and is continuous form \( \mathbb{D}^{1,p}(H \otimes U) \) to \( L^p(\Omega_1; U) \) for any \( p > 1 \). The duality relationship reads

\[
\mathbb{E}_1((DF, u)_{H \otimes U}) = \mathbb{E}_1((F, \delta(u))_U),
\]

for any \( F \in \mathbb{D}^{1,p}(U) \) and \( u \in \mathbb{D}^{1,q}(H \otimes U) \), with \( \frac{1}{p} + \frac{1}{q} = 1 \).

A square integrable random variable \( F \in L^2(\Omega_1; V) \) can be identified with an element of \( L^2(\Omega_1; V) \), where \( V = L^2(\Omega_2) \).

The following is the main result of this section.

**Theorem 3.1.** Under (H2), equation (1) has a unique solution \( X_t \) and, moreover, for any \( t \geq 0 \) and any \( h \in H \), \( X_t \in \mathbb{D}^{1,\infty}(\mathbb{R}^n \otimes V) \) and the derivative \( D^hX_t \) is the unique solution to the linear SDE

\[
d\psi^h_t = \nabla b(X_t, \alpha_t)\psi^h_t dt + \sum_{l=1}^d \nabla \sigma_l(X_t, \alpha_t)\psi^h_t dW^l_t + \sigma(X_t, \alpha_t)h_t dt
\]

with initial value \( \psi^h_0 = 0 \).
To prove the theorem, let $X_t^{\varepsilon h}$ be the solution of equation (1) with $W_t$ replaced by $W_t + \varepsilon \int_0^t h_s ds$, where $\varepsilon \in (0, 1)$, that is,

$$
\begin{align*}
\begin{cases}
    dX_t^{\varepsilon h} = b(X_t^{\varepsilon h}, \alpha_t)dt + \sigma(X_t^{\varepsilon h}, \alpha_t)dW_t + \varepsilon \sigma(X_t^{\varepsilon h}, \alpha_t)h_t dt, \\
    (X_0^{\varepsilon h}, \alpha_0) = (x, \alpha) \in \mathbb{R}^n \times \mathcal{S},
\end{cases}
\end{align*}
$$

(5)

where $\alpha_t$ is defined in (2). Then, we can write

$$
\frac{X_t^{\varepsilon h} - X_t}{\varepsilon} = \frac{1}{\varepsilon} \int_0^t [b(X_s^{\varepsilon h}, \alpha_s) - b(X_s, \alpha_s)]ds + \frac{1}{\varepsilon} \int_0^t [\sigma(X_s^{\varepsilon h}, \alpha_s) - \sigma(X_s, \alpha_s)]dW_s
$$

$$
+ \int_0^t \sigma(X_s^{\varepsilon h}, \alpha_s)h_s ds.
$$

In order to prove Theorem 3.1, we first give some preliminary lemmas.

**Lemma 3.2.** Suppose that Hypothesis (A₁) holds. Then for any $h \in H$, $T > 0$ and $p \geq 2$, we have

$$
\mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t^{\varepsilon h}|^p \right] \leq C.
$$

**Proof.** From equation (5) it is easy to see that

$$
|X_t^{\varepsilon h}|^p \leq C \left[ |x|^p + \left| \int_0^t b(X_s^{\varepsilon h}, \alpha_s)ds \right|^p + \left| \int_0^t \sigma(X_s^{\varepsilon h}, \alpha_s)dW_s \right|^p 
$$

$$
+ \varepsilon \left| \int_0^t \sigma(X_s^{\varepsilon h}, \alpha_s)h_s ds \right|^p \right] = C \left[ |x|^p + I_1(t) + I_2(t) + I_3(t) \right].
$$

By Hölder’s and Burkholder-Davis-Gundy’s inequalities, we obtain

$$
\mathbb{E} \left[ \sup_{0 \leq t \leq T} (I_1(t) + I_2(t) + I_3(t)) \right] \leq C \int_0^T (\mathbb{E}|X_s^{\varepsilon h}|^p + 1)ds.
$$

Then the desired estimate follows from Gronwall’s lemma. \qed

**Lemma 3.3.** Suppose that Hypothesis (A₁) holds. Then for any $h \in H$, $T > 0$ and $p \geq 2$, we have

$$
\mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t^{\varepsilon h} - X_t|^p \right] \leq C \varepsilon^p.
$$

**Proof.** We write

$$
X_t^{\varepsilon h} - X_t = \int_0^t [b(X_s^{\varepsilon h}, \alpha_s) - b(X_s, \alpha_s)]ds + \varepsilon \int_0^t \sigma(X_s^{\varepsilon h}, \alpha_s)h_s ds
$$

$$
+ \int_0^t \left[ \sigma(X_s^{\varepsilon h}, \alpha_s) - \sigma(X_s, \alpha_s) \right]dW_s.
$$

Applying Hölder’s inequality, Burkholder-Davis-Gundy’s inequality and Lemma 3.2, we have

$$
\mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t^{\varepsilon h} - X_t|^p \right] \leq C \int_0^T \mathbb{E}|X_s^{\varepsilon h} - X_s|^p ds + C \varepsilon^p.
$$

Hence, Gronwall’s inequality yields

$$
\mathbb{E} \left[ \sup_{0 \leq t \leq T} |X_t^{\varepsilon h} - X_t|^p \right] \leq C \varepsilon^p.
$$

\qed
Proof of Theorem 3.1. Let $\psi^h_t$ be the solution of equation (4). It is easy to show that $\mathbb{E} \left[ \sup_{0 \leq s \leq t} \left| \psi^h_t \right|^p \right] \leq C$, where $C$ is a constant depending on $T, x, h$ and $p$. We have

$$
\frac{X^{\varepsilon h}_t - X_t}{\varepsilon} - \psi^h_t = \frac{1}{\varepsilon} \int_0^t \left[ \int_0^1 b(X_s^\varepsilon, \alpha_s) - b(X_s, \alpha_s) - \varepsilon \nabla b(X_s, \alpha_s) \psi^h_s \right] ds 
+ \frac{1}{\varepsilon} \int_0^t \sum_{l=1}^d \left[ \sigma_l(X_s^\varepsilon, \alpha_s) - \sigma_l(X_s, \alpha_s) - \varepsilon \sigma_l(X_s, \alpha_s) \psi^h_s \right] dW^l_s 
+ \int_0^t \left[ \sigma(X_s^\varepsilon, \alpha_s) - \sigma(X_s, \alpha_s) \right] h_s ds.
$$

Using twice the mean valued theorem we have

$$
\frac{X^{\varepsilon h}_t - X_t}{\varepsilon} = \frac{1}{\varepsilon} \left[ \int_0^1 \nabla b(X_s + \nu(X_s^\varepsilon - X_s), \alpha_s) d\nu \right] \frac{X^{\varepsilon h}_t - X_s}{\varepsilon} - \nabla b(X_s, \alpha_s) \psi^h_s ds 
+ \int_0^t \sum_{l=1}^d \left[ \int_0^1 \nabla \sigma_l(X_s + \nu(X_s^\varepsilon - X_s), \alpha_s) d\nu \right] \frac{X^{\varepsilon h}_s - X_s}{\varepsilon} - \nabla \sigma_l(X_s, \alpha_s) \psi^h_s ds 
+ \int_0^t \left[ \sigma(X_s^\varepsilon, \alpha_s) - \sigma(X_s, \alpha_s) \right] h_s ds.
$$

where $\varphi^\varepsilon_t$ defined by

$$
\varphi^\varepsilon_t = \int_0^t \left[ \sigma(X_s^\varepsilon, \alpha_s) - \sigma(X_s, \alpha_s) \right] h_s ds 
+ \int_0^t \left( \int_0^1 \nabla b(X_s + \nu(X_s^\varepsilon - X_s), \alpha_s) d\nu - \nabla b(X_s, \alpha_s) \right) \psi^h_s ds 
+ \int_0^t \sum_{l=1}^d \left( \int_0^1 \nabla \sigma_l(X_s + \nu(X_s^\varepsilon - X_s), \alpha_s) d\nu - \nabla \sigma_l(X_s, \alpha_s) \right) \psi^h_s dW^l_s.
$$

By Hypothesis (H2), we obtain

$$
\mathbb{E} \left[ \sup_{0 \leq s \leq t} \left| \frac{X^{\varepsilon h}_s - X_s}{\varepsilon} - \psi^h_s \right|^p \right] \leq C \int_0^t \mathbb{E} \left[ \left| \frac{X^{\varepsilon h}_s - X_s}{\varepsilon} - \psi^h_s \right|^p \right] ds 
+ C \mathbb{E} \left[ \sup_{0 \leq s \leq t} \left| X^{\varepsilon h}_s - X_s \right|^p \right] 
+ C \left( \mathbb{E} \sup_{0 \leq s \leq t} \left| X^{\varepsilon h}_s - X_s \right|^{2p} \right)^{1/2} \left( \mathbb{E} \sup_{0 \leq s \leq t} \left| \psi^h_s \right|^{2p} \right)^{1/2}.
$$

Using Gronwall’s inequality and Lemma 3.3, we obtain

$$
\lim_{\varepsilon \to 0} \mathbb{E} \left[ \sup_{0 \leq s \leq t} \left| \frac{X^{\varepsilon h}_s - X_s}{\varepsilon} - \psi^h_s \right|^p \right] = 0.
$$
This implies that for any \( p \geq 2 \),
\[
\lim_{\varepsilon \to 0} \mathbb{E}_1 \left[ \sup_{0 \leq s \leq t} \frac{\| X_{s\varepsilon} - X_s \|}{\varepsilon} = 0. \right.
\]

Now, let \( D_sX_t \) be the solution of the following equation:
\[
D_sX_t = \sigma(X_s, \alpha_s) + \int_s^t \nabla h(X_r, \alpha_r)D_sX_r dr + \int_s^t \sum_{i=1}^d \nabla \sigma_i(X_r, \alpha_r)D_sX_r dW_t^i
\]
for \( s \leq t \) and \( D_sX_t = 0 \) for \( s > t \). Then we can easily obtain that \( D^hX_t = \psi_t^h \) and \( DX_t \in L^\infty - (\Omega, H \otimes \mathbb{R}^n \otimes V) \). The proof is complete.

As a consequence, we can show the following version of the chain rule.

**Theorem 3.4. (Chain rule)** Assume that condition (H2) holds. Then for any \( h \in H \), \( t \geq 0 \) and \( p \geq 2 \), if \( f \in C_b^2(\mathbb{R}^n \times \mathbb{S}) \), we have
\[
\lim_{\varepsilon \to 0} \mathbb{E} \left[ \left| \frac{f(X_{t\varepsilon}^h, \alpha_t)}{\varepsilon} - \nabla f(X_t, \alpha_t)D^hX_t \right| ^p \right] = 0.
\]
Moreover, \( f(X_t, \alpha_t) \in \mathbb{D}^{1, \infty}(V) \) and \( Df(X_t, \alpha_t) = \nabla f(X_t, \alpha_t)DX_t \).

**Proof.** We can write
\[
\mathbb{E} \left[ \left| \frac{f(X_{t\varepsilon}^h, \alpha_t)}{\varepsilon} - \nabla f(X_t, \alpha_t)D^hX_t \right| ^p \right] \\
\leq C\mathbb{E} \left[ \left| \frac{f(X_{t\varepsilon}^h, \alpha_t)}{\varepsilon} - \nabla f(X_t, \alpha_t) \frac{X_{t\varepsilon}^h - X_t}{\varepsilon} \right| ^p \right. \\
\left. + C\mathbb{E} \left| \nabla f(X_t, \alpha_t) \frac{X_{t\varepsilon}^h - X_t}{\varepsilon} - \nabla f(X_t, \alpha_t)D^hX_t \right| ^p \right]. \quad (6)
\]
For the second term in (6), by Theorem 3.1 we have
\[
\lim_{\varepsilon \to 0} \mathbb{E} \left[ \left| \nabla f(X_t, \alpha_t) \frac{X_{t\varepsilon}^h - X_t}{\varepsilon} - \nabla f(X_t, \alpha_t)D^hX_t \right| ^p \right] = 0. \quad (7)
\]
Consider now the first term in (6), since \( f \in C_b^2(\mathbb{R}^n \times \mathbb{S}) \) we have
\[
\mathbb{E} \left[ \left| \frac{f(X_{t\varepsilon}^h, \alpha_t)}{\varepsilon} - \nabla f(X_t, \alpha_t) \frac{X_{t\varepsilon}^h - X_t}{\varepsilon} \right| ^p \right] \\
\leq C\mathbb{E} \left| X_{t\varepsilon}^h - X_t \right| ^{2p} \leq C\varepsilon^p. \quad (8)
\]
From (6)-(8) the theorem follows.

4. **Smoothness of the density.** In this section we show that under suitable non degeneracy assumptions on the coefficients, for any \( t > 0 \) the random vector \( X_t \) has a smooth density. To this end we first study the stochastic flow associated with equation (1) and then we show that the determinant of the Malliavin covariance matrix of \( X_t \) has finite negative moments of all orders for any \( t > 0 \), under a uniform Hörmander’s condition.
**Definition 4.1.** Suppose that $F(x, \alpha) : \Omega \to \mathbb{R}^n$ is a measurable function for all $x \in \mathbb{R}^n$ and $\alpha \in \mathcal{S}$. We say that its gradient with respect to $x$ exists (in mean square sense) if there is $A(x, \alpha) : \Omega \to \mathbb{R}^{n^2}$ such that for any $\xi \in \mathbb{R}^n$ we have

$$
\lim_{\varepsilon \to 0} \mathbb{E} \left| \frac{F(x + \varepsilon \xi, \alpha) - F(x, \alpha) - A(x, \alpha)\xi}{\varepsilon} \right|^2 = 0.
$$

We denote the gradient matrix $A(x, \alpha)$ by $\nabla F(x, \alpha)$.

By arguments similar to those used in the proof of Theorem 3.1, we can obtain the following results.

**Theorem 4.2.** Assume the hypothesis $(H_2)$ holds. Let $\{(X_{s,t}(x, \alpha), \alpha_{s,t}(\alpha)), t \geq s\}$ be the solution of equations (1) and (2), which starts from $(x, \alpha)$ at time $s$. Then the gradient of $X_{s,t}(x, \alpha)$ with respect to $x$ (in mean square) exists. If we denote

$$J_{s,t} := \nabla X_{s,t}(x, \alpha),$$

then

$$
\begin{align*}
\left\{ 
\begin{array}{l}
dJ_{s,t} = \nabla b(X_t, \alpha_t)J_{s,t}dt + \sum_{l=1}^{d} \nabla \sigma_l(X_t, \alpha_t)J_{s,t}dW^l_t, \\
J_{s,s} = I,
\end{array}
\right. 
\end{align*}
$$

(9)

where $I$ is the $n$-dimensional identity matrix. Moreover, $J_{s,t}$ is invertible and its inverse $J_{s,t}^{-1}$ satisfies

$$
\begin{align*}
\left\{ 
\begin{array}{l}
dJ_{s,t}^{-1} = -J_{s,t}^{-1} \left( \nabla b(X_t, \alpha_t) - \sum_{l=1}^{d} \nabla \sigma_l(X_t, \alpha_t)\nabla \sigma_l(X_t, \alpha_t) \right) dt \\
&\quad - \sum_{l=1}^{d} J_{s,t}^{-1}\nabla \sigma_l(X_t, \alpha_t)dW^l_t,
\end{array}
\right. 
\end{align*}
$$

(10)

The following lemma provides estimates on the $L^p$-norm of the gradient of the solution and its inverse.

**Lemma 4.3.** Assume that Hypothesis $(H_2)$ holds. Then for any $p \geq 2$, there exists a positive constant $C$ depending only on $T$ and $p$ such that

$$
\mathbb{E} \left[ \sup_{s \leq t \leq T} \left( |J_{s,t}|^p + |J_{s,t}^{-1}|^p \right) \right] \leq C.
$$

**Proof.** From

$$J_{s,t} = I + \int_s^t \nabla b(X_r, \alpha_r)J_{s,r}dr + \int_s^t \nabla \sigma(X_r, \alpha_r)J_{s,r}dW_r,$$

we obtain

$$|J_{s,t}|^p \leq C \left[ 1 + \left( \int_s^t |\nabla b(X_r, \alpha_r)J_{s,r}|^p dr \right)^{\frac{1}{p}} + \left( \int_s^t |\nabla \sigma(X_r, \alpha_r)J_{s,r}|^p dW_r \right)^{\frac{1}{p}} \right]$$

for any $p \geq 2$. Then by Hölder’s inequality, we can write

$$
\mathbb{E} \left[ \sup_{s \leq t \leq T} \left( \int_s^t |\nabla b(X_r, \alpha_r)J_{s,r}|^p dr \right)^{\frac{1}{p}} \right] \leq C \int_s^T \mathbb{E} |J_{s,r}|^p dr,
$$

(10)
and by Burkholder-Davis-Gundy’s inequality we have
\[ E \left[ \sup_{s \leq t \leq T} \left| \int_s^t \nabla \sigma(X_r, \alpha_r) J_s \, dW_r \right|^p \right] \leq C \int_s^T E |J_s|^p \, ds. \]
Hence, we have
\[ E \left[ \sup_{s \leq t \leq T} |J_{s,t}|^p \right] \leq C + C \int_s^T E |J_s|^p \, ds. \]
Gronwall’s inequality yields
\[ E \left[ \sup_{s \leq t \leq T} |J_{s,t}|^p \right] \leq C. \]
Similarly,
\[ E \left[ \sup_{s \leq t \leq T} |J_{s,t}^{-1}|^p \right] \leq C. \]
The proof of the lemma is now complete. \( \square \)

Using the gradient of the flow \( J_{s,t} \) we can represent the Malliavin derivative \( DX_t \) as follows:
\[ D_s X_t = J_{s,t} \sigma(X_s, \alpha_s), \quad 0 \leq s \leq t \leq T; \quad D_s X_t = 0, s > t. \]

Next, we shall study the Malliavin differentiability of \( J_{s,t} \). Denote by \( D^l_r \) the Malliavin derivative with respect to the \( l \)-th component of the Brownian motion \( W \) at time \( r \).

**Lemma 4.4.** Suppose that Hypothesis (H3) holds. Then the following two statements hold:

1. For all \( 0 \leq s \leq t \leq T \), \( J_{s,t} \in D^{1,\infty}(\mathbb{R}^n \otimes \mathbb{R}^n \otimes V) \) and for any \( p \geq 2 \), there exists a positive constant \( C \) depending on \( T, p \) and \( x \), such that for all \( l = 1, \ldots, d \) and \( r \in [0, T] \)
\[ E \left[ \sup_{s \leq t \leq T} |D^l_r J_{s,t}|^p \right] \leq C. \]
2. For any \( t \leq T \), \( X_t \in D^{2,\infty}(\mathbb{R}^n \otimes V) \) and for any \( p \geq 2 \), there exists a positive constant \( C \) depending on \( T, p \) and \( x \), such that for all \( l, k = 1, \ldots, d \) and \( r, s \leq t \)
\[ \mathbb{E}[D^l_r(D^k_s X_t)]^p \leq C. \]

**Proof.** First, we prove statement 1. By equation (9), the definition of Malliavin derivative and the arguments used in the proof of Theorem 3.1, we obtain
\[ \lim_{\varepsilon \to 0} \mathbb{E} \left[ \frac{J_{s,t}(\omega_1 + \varepsilon \int_0^h h_s \, ds, \omega_2) - J_{s,t}(\omega_1, \omega_2)}{\varepsilon} - \langle DJ_{s,t}, h \rangle_H \right]^p_{\mathbb{R}^n \otimes \mathbb{R}^n \otimes V} = 0 \]
for any \( h \in H \) and any \( p \geq 2 \), where \( DJ_{s,t} \) satisfies the following equation for all \( r \leq t \)
\[ D^l_r J_{s,t} = \nabla \sigma_t(X_r, \alpha_r) J_{s,r} \mathbf{1}_{\{s \leq r \leq t\}}. \]
An application of chain rule, which can be easily established, yields

\[ \mathbb{E} \left[ \sup_{r \leq s \leq t} |D^k_r J_{s,t}| \right] \leq C |J_{s,t}| + C \int_r^T \mathbb{E}(|D^k_r X_{u}^p| \cdot |J_{s,u}|^p) du 
+ C \int_r^T \mathbb{E}|D^k_{s,u}|^p du 
\leq C |J_{s,t}| + C \int_r^T (\mathbb{E}|D^k_r X_{u}^{2p}|^{1/2} \cdot (\mathbb{E}|J_{s,u}|^{2p})^{1/2}) du 
+ C \int_r^T \mathbb{E}|D^k_{s,u}|^p du. \]

Hence, Lemma 4.3 and Gronwall’s inequality yield

\[ \mathbb{E} \left[ \sup_{r \leq s \leq t} |D^k_r J_{s,t}| \right] \leq C e^{C(T-r)}. \]

The case \( r < s \) can be handled in a similar way and hence the statement 1 is proved.

We prove statement 2. Note that for any \( k = 1, \ldots, d \),

\[ D^k_s X_t = J_{s,t} \sigma_k(X_s, \alpha_s) \quad \text{when} \quad s \leq t \quad \text{and} \quad D^k_s X_t = 0 \quad \text{when} \quad s > t. \]

An application of chain rule, which can be easily established, yields

\[ D^k_t (D^k_s X_t) = (D^k_t J_{s,t}) \sigma_k(X_s, \alpha_s) + J_{s,t} \nabla \sigma_k(X_s, \alpha_s) D^k_t X_s. \]

Then, Hölder’s inequality gives

\[ \mathbb{E}|D^k_t (D^k_s X_t)|^p \leq C \mathbb{E}|(D^k_t J_{s,t}) \sigma_k(X_s, \alpha_s)|^p + C \mathbb{E}|J_{s,t} \nabla \sigma_k(X_s, \alpha_s) D^k_t X_s|^p \]
\[ \leq C (\mathbb{E}|D^k_t J_{s,t}|^{2p})^{1/2} (\mathbb{E}|1 + |X_s|^{2p}|)^{1/2} \]
\[ + C(\mathbb{E}|J_{s,t}|^{2p})^{1/2} (\mathbb{E}|D^k_t X_s|^{2p})^{1/2} \leq C, \]

which implies statement 2.

\[ \square \]

\textbf{Remark 1.} Following the same procedure as above we can prove that if Hypothesis \((H_{\infty})\) holds, then \( J_{s,t} \in \mathbb{D}^{\infty}(\mathbb{R}^n \otimes \mathbb{R}^n \otimes V) \) and \( X_t \in \mathbb{D}^{\infty}(\mathbb{R}^n \otimes V). \)
We denote by \((DX_t)^*\) the transpose of the random matrix \(DX_t\). From the relation between \(DX_t\) and \(J_{s,t}\), we have \((DX_t)^*(r) = \sigma(X_r, \alpha_r)^*J_{r,t}^\ast\). Then, the Malliavin matrix \(M_t\) of the random vector \(X_t\) is defined by:

\[
M_t = \langle DX_t, (DX_t)^\ast \rangle_H = \int_0^t J_{s,t}\sigma(X_s, \alpha_s)\sigma(X_s, \alpha_s)^*J_{s,t}^\ast ds
\]

\[
= J_{0,t} \int_0^t J_{0,s}^{-1}\sigma(X_s, \alpha_s)\sigma(X_s, \alpha_s)^*(J_{0,s}^{-1})^* ds J_{0,t}^\ast
\]

where

\[
C_t = \int_0^t J_{0,s}^{-1}\sigma(X_s, \alpha_s)\sigma(X_s, \alpha_s)^*(J_{0,s}^{-1})^* ds,
\]

is the so-called reduced Malliavin matrix of \(X_t\).

Our aim is to show that, under a suitable nondegeneracy condition on the coefficients, the Malliavin matrix \(M_t\) is invertible \(\mathbb{P}\)-a.s. and the determinant of its inverse has negative moments of all orders. The difficulty in our current situation is that the vector fields \(b\) and \(\sigma_1, \ldots, \sigma_d\) depend on the Markovian switching process \(\alpha_t\). To overcome this difficulty we follow the following procedure inspired by [2].

For \(t \geq 0\) we define \(N_t := N([0, t], m_0(m_0 - 1)K)\), so \(\{N_t, t \geq 0\}\) is a Poisson process with parameter \(m_0(m_0 - 1)K\). Conditioned on the number of jumps of the Poisson process up to time \(t\), that is, \(N_t = k\), there exists a random interval \([T_1, T_2]\) with \(0 \leq T_1 < T_2 \leq t\), such that \(T_2 - T_1 \geq \frac{t}{k+1}\). This implies that \(\alpha_t = \alpha_{T_1}\) for all \(t \in [T_1, T_2]\) (because that the jump times of \(\alpha_t\) are a subset of the jump times of \(N_t\)). On this random time interval, we will apply the classical techniques of Malliavin calculus.

To this end we need the following version of Norris lemma on time intervals.

**Lemma 4.5.** Let \(t_1 \geq 0\) and let \(\xi_1, \xi_2\) be two \(\mathcal{F}_{t_1}\)-measurable random variables. Suppose that \(\beta(t), \gamma(t) = (\gamma_1(t), \ldots, \gamma_d(t))\) and \(u(t) = (u_1(t), \ldots, u_d(t))\) are \(\mathcal{F}_t\)-adapted processes. For any \(t \geq t_1\), set

\[
a(t) = \xi_1 + \int_{t_1}^t \beta(s)ds + \sum_{l=1}^d \int_{t_1}^t \gamma_l(s) dW^l_s
\]

\[
Y(t) = \xi_2 + \int_{t_1}^t a(s)ds + \sum_{l=1}^d \int_{t_1}^t u_l(s) dW^l_s
\]

and assume that for some \(p \geq 2\) and \(T > 0\)

\[
\mathbb{E} \left( \sup_{t_1 \leq t \leq T} (|\beta(t)| + |\gamma(t)| + |a(t)| + |u(t)|)^p \right) < \infty. \tag{13}
\]

Consider \(t_2 \in [0, T]\) satisfying \(t_2 - t_1 \geq \epsilon\) for some constant \(c > 0\). Then, for any \(q > 8\) and \(r > 0\) such that \(18r < q - 8\), there exists \(\epsilon_0 = \delta_0 c^{\gamma_0}\), where the positive constants \(\delta_0\) and \(\gamma_0\) depend on \(p, q, r\) and \(T\), such that for all \(\epsilon \in (0, \epsilon_0)\)

\[
\mathbb{P} \left( \int_{t_1}^{t_2} Y^2_t dt < \epsilon^{q}, \int_{t_1}^{t_2} (|a(t)|^2 + |u(t)|^2) dt \geq \epsilon \right) \leq \epsilon^{\epsilon p}.
\]

The proof is similar to the proof of [5, Lemma 2.3.2] and we omit the details. Just remark that the lower bound \(c\) on the length of the interval \([t_1, t_2]\) is crucial in the proof.
We are going to impose a uniform Hörmander’s condition on the coefficients. To formulate this condition we need some notation. Consider the following sets of vector fields:

\[
\Sigma_0 = \{\sigma_1, \ldots, \sigma_d\}, \\
\Sigma_n = \{[\sigma_k, V], k = 0, \ldots, d, V \in \Sigma_{n-1}\}, \quad n \geq 1, \\
\Sigma = \bigcup_{n=0}^{\infty} \Sigma_n,
\]

where \(\sigma_0 = b - \frac{1}{2} \sum_{l=1}^{d} (\nabla \sigma_l) \sigma_l\) and \([V, G] = (\nabla G)V - (\nabla V)G\) denotes the Lie bracket between two vector fields \(V\) and \(G\).

The following uniform Hörmander’s condition requires that the vector space spanned by \(\{V(x, \alpha), V \in \Sigma\}\) is \(\mathbb{R}^n\) for all \((x, \alpha) \in \mathbb{R}^n \times \mathbb{S}\) in a uniform way, where \(V_k(x, \alpha)\) denotes the vector obtained by freezing the variables \(x\) and \(\alpha\) in the vector field \(V_k\).

(UHC) (Uniform Hörmander’s condition) Condition \((H_{\infty})\) holds and there exists an integer \(l_0 \geq 0\) and a constant \(c > 0\) such that

\[
\sum_{l=0}^{l_0} \sum_{V \in \Sigma_l} (v^* V(x, \alpha))^2 \geq c, \quad (14)
\]

for all \(x \in \mathbb{R}^n, \alpha \in \mathbb{S}\) and \(v \in \mathbb{R}^n\) with \(|v| = 1\).

**Theorem 4.6.** Assume that the uniform Hörmander’s condition \((UHC)\) holds. Then for all \(t > 0\) the Malliavin matrix \(M_t\) of the random vector \(X_t\) is invertible \(\mathbb{P}\text{-a.s.}\) and \(\det(M_t^{-1}) \in L^p(\Omega)\) for all \(p \geq 2\). As a consequence, for any \(t > 0\), the law of \(X_t\) is absolutely continuous with respect to Lebesgue’s measure and the density is smooth.

**Proof.** We recall that \(M_t = J_{0,t} C_t J_{0,t}^*\). By Lemma 4.3 it suffices to prove that \(\det(C_t^{-1}) \in L^p(\Omega)\) for all \(p \geq 2\).

Recall that \(\{N_t = N([0, t], m_0(m_0 - 1)K), t \geq 0\}\) is a Poisson process with parameter \(\lambda := m_0(m_0 - 1)K\). For a fixed \(t > 0\), conditioned on \(N_t = k\), there exists a random interval \([T_1, T_2]\) such that \(T_2 - T_1 \geq \frac{t}{k+1}\) and \(\alpha_t = \alpha_{T_1}\) for all \(t \in [T_1, T_2]\).

We introduce the following sets of vector fields:

\[
\Sigma'_0 = \Sigma_0; \\
\Sigma'_n = \left\{[\sigma_k, V], k = 1, \ldots, d, V \in \Sigma'_{n-1}; \right. \\
\left. [\sigma_0, V] + \frac{1}{2} \sum_{l=1}^{d} [\sigma_l, [\sigma_l, V]], V \in \Sigma'_{n-1} \right\}, \quad n \geq 1; \\
\Sigma' = \bigcup_{n=0}^{\infty} \Sigma'_n.
\]

We denote by \(\Sigma_n(x, \alpha)\) (resp. \(\Sigma'_n(x, \alpha)\)) the subset of \(\mathbb{R}^n\) obtained by freezing the variable \(x, \alpha\) in the vector fields of \(\Sigma_n\) (resp. \(\Sigma'_n\)). Clearly, the vector spaces spanned by \(\Sigma(x, \alpha)\) or by \(\Sigma'(x, \alpha)\) coincide. By condition \((UHC)\), there exists an integer
$k_0 \geq 0$ and a $c > 0$ such that
\[
\inf_{x \in \mathbb{R}^n} \inf_{\alpha \in \mathcal{S}} \sum_{k=0}^{k_0} \sum_{V \in \mathcal{V}_k'} (v^* V(x, \alpha))^2 \geq c,
\]
for all $|v| = 1$.

For all $l = 0, 1, \ldots, l_0$, denote $m(l) = 2^{-4l}$ and define
\[
E_l = \left\{ \sum_{V \in \mathcal{V}_l} \int_{T_1}^{T_2} (v^* J_{0,s}^{-1} V(X_s, \alpha_s))^2 ds \leq \varepsilon^m(l) \right\}.
\]

Clearly $\{v^* C_l v \leq \varepsilon\} \subseteq E_0$. Consider the decomposition
\[
E_0 \subseteq (E_0 \cap E_1^c) \cup (E_1 \cap E_2^c) \cup \cdots \cup (E_{l_0-1} \cap E_{l_0}^c) \cup F,
\]
where $F = E_0 \cap E_1 \cap \cdots \cap E_{l_0}$. Then for any unit vector $v$ we have
\[
\mathbb{P}\{|v^* C_l v \leq \varepsilon| N_t = k\} \leq \mathbb{P}(E_0| N_t = k) \\
\leq \mathbb{P}(F| N_t = k) + \sum_{l=0}^{l_0-1} \mathbb{P}(E_l \cap E_{l+1}^c| N_t = k).
\]

We are going to estimate each term in the above sum. This will be done in two steps.

**Step 1.** We can write
\[
\mathbb{P}(F| N_t = k) \leq \mathbb{P}(F \cap G| N_t = k) + \mathbb{P}(G^c| N_t = k),
\]
where $G := \{ \sup_{0 \leq s \leq T_2} \| J_{0,s} \| \leq \frac{1}{\varepsilon^2}, 0 < 2\beta < m(l_0) \}$. First we claim that when $\varepsilon$ is sufficiently small, the intersection $F \cap G \cap \{ N_t = k \}$ is empty. In fact, taking into account the estimate (15), on $N_t = k$, we have
\[
\sum_{l=0}^{l_0} \sum_{V \in \mathcal{V}_l} \int_{T_1}^{T_2} (v^* J_{0,s}^{-1} V(X_s, \alpha_s))^2 ds \\
= \sum_{l=0}^{l_0} \sum_{V \in \mathcal{V}_l} \int_{T_1}^{T_2} \left( \frac{v^* J_{0,s}^{-1} V(X_s, \alpha_s)}{|v^* J_{0,s}^{-1}|} \right)^2 |v^* J_{0,s}^{-1}|^2 ds \geq \frac{t \varepsilon^2 \beta}{k+1},
\]

because $|v^* J_{0,s}^{-1}| \geq \frac{1}{\| J_{0,s} \|} \geq \varepsilon^\beta$, and $T_2 - T_1 \geq \frac{t}{k+1}$. On the other hand, the left-hand side of (17) is bounded by $(l_0 + 1)\varepsilon^m(l_0)$ on the set $F$. Thus $F \cap G \cap \{ N_t = k \} = \emptyset$, provided $\varepsilon < \varepsilon_1$, where $\varepsilon_1 = \left[ \frac{t \varepsilon^2 \beta}{(k+1)(l_0 + 1)} \right].$

Now we consider the second term in (16). Using Chebyshev inequality we obtain
\[
\mathbb{P}\left( \sup_{T_1 \leq s \leq T_2} |J_{0,s}| \geq \varepsilon^{-\beta} \right| N_t = k \leq \varepsilon \mathbb{P} \left( \sup_{T_1 \leq s \leq T_2} |J_{0,s}|^p | N_t = k \right).
\]

Taking into account that the Poisson random measure $N$ is independent of the Brownian motion $W$, we can estimate the above conditional expectation using Burkholder-Davis-Gundy’s inequality as in Lemma 4.3, and we obtain the estimate
\[
\mathbb{P}\left( \sup_{T_1 \leq s \leq T_2} |J_{0,s}| \geq \varepsilon^{-\beta} \right| N_t = k \leq C \varepsilon^{-\beta};
\]

\[
(18)
\]
Step 2. We shall bound the remaining parts. For any \( l = 0, \ldots, l_0 - 1 \) we have

\[
\mathbb{P}(E_l \cap E_{l+1}^C | N_t = k) = \mathbb{P}\left\{ \sum_{V \in \Sigma_l} \int_{T_1}^{T_2} (v^* J_{0,s}^{-1} V(X_s, \alpha_s))^2 ds \leq \varepsilon^{m(l)}, \quad \sum_{V \in \Sigma_{l+1}} \int_{T_1}^{T_2} (v^* J_{0,s}^{-1} V(X_s, \alpha_s))^2 ds > \varepsilon^{m(l+1)} | N_t = k \right\}
\]

\[
\leq \sum_{V \in \Sigma_l} \mathbb{P}\left\{ \int_{T_1}^{T_2} (v^* J_{0,s}^{-1} V(X_s, \alpha_T)) ds \leq \varepsilon^{m(l)} \right\}
\]

\[
\leq \sum_{l=1}^{d} \int_{T_1}^{T_2} (v^* J_{0,s}^{-1} [\sigma_l, V](X_s, \alpha_T)) ds + \int_{T_1}^{T_2} (v^* J_{0,s}^{-1} ([\sigma_0, V]) ds > \varepsilon^{m(l+1)} \frac{m(l+1)}{n(l)} | N_t = k \},
\]

where \( n(l) \) denotes the cardinality of the set \( \Sigma_l' \). Consider the continuous semimartingale \( \{ v^* J_{0,t}^{-1} V(X_t, \alpha_T), T_1 \leq t < T_2 \} \). For any \( t \in [T_1, T_2] \) Itô’s formula yields

\[
v^* J_{0,t}^{-1} V(X_t, \alpha_T)
\]

\[
= v^* J_{0,t_1}^{-1} V(X_{T_1}, \alpha_T) + \int_{T_1}^{t} v^* J_{0,s}^{-1} [\sigma_l, V](X_s, \alpha_T) dW_s^l
\]

\[
+ \int_{T_1}^{t} v^* J_{0,s}^{-1} [\sigma_0, V] ds + \frac{1}{2} \sum_{l=1}^{d} [\sigma_l, [\sigma_l, V]](X_s, \alpha_T) ds.
\]

Notice that \( 8m(l+1) < m(l) \) and also notice condition (13) holds for any \( p \) and the fact that the Poisson random measure \( N \) is independent of \( W \). An application of Lemma 4.5 to the semimartingale \( Y_t = v^* J_{0,t}^{-1} V(X_t, \alpha_T) \) with time interval \( [T_1, T_2] \) which satisfy \( T_2 - T_1 \geq \frac{T}{k+1} \) on the set \( N_t = k \) yields

\[
\mathbb{P}(E_l \cap E_{l+1}^C | N_t = k) \leq \varepsilon^p
\]

(19)

for any \( p \geq 2 \), and for \( \varepsilon < \varepsilon_0 \), where \( \varepsilon_0 = \delta_0 \left( \frac{T}{k+1} \right)^{\gamma_0} \). The exponents \( \delta_0 \) and \( \gamma_0 \) only depend on \( p \) and \( T \). Therefore, from (18) and (19) we obtain

\[
\mathbb{P}\{ v^* C_t \leq \varepsilon | N_t = k \} \leq \varepsilon^p,
\]

for any \( p \geq 2 \), and for \( \varepsilon < \min(\varepsilon_0, \varepsilon_1) \). Then, following the steps of [5, Lemma 2.3.1], we can obtain that

\[
\mathbb{P}\left\{ \inf_{|v| = 1} v^* C_t \leq \varepsilon | N_t = k \right\} \leq \varepsilon^p,
\]
for all $0 < \varepsilon \leq C_1 (\frac{t}{t + 1})^{C_2}$ and for all $p \geq 2$, where $C_1$ and $C_2$ are positive constants depending on $p$, $T$ and $n$. Consequently,

$$E|\det(C_2)|^{-p} \leq E(\inf_{|v|=1} v^* C_t v)^{-np} \leq \sum_{k=0}^{\infty} P(N_t = k) E\left(\inf_{|v|=1} v^* C_t v\right)^{-np},$$

$$\leq \sum_{k=0}^{\infty} \lambda^k \frac{1}{k!} e^\lambda C_1 \left(\frac{t}{k+1}\right)^{C_2} + \frac{1}{C_1} \left(\frac{k+1}{t}\right)^{C_2} < \infty.$$ 

The proof is now complete. \quad \Box

5. **Bismut type formula.** In this section, we prove a version of Bismut type formula for SDEs with Markovian switching. As an application, this formula is used to obtain the strong Feller property for the transition semigroup of $(X_t, \alpha_t)$.

**Theorem 5.1.** Suppose the condition (UHC) holds. Then for any $f \in C_b^2(\mathbb{R}^n \times S)$, we have

$$\nabla P_t f(x, \alpha) = E\left[f(X_t, \alpha_t) \int_0^t \sigma(X_s, \alpha_s)^* J_{s,t}^{-1} J_{0,t} dW_s\right],$$

(20)

where $M_t = \int_0^t J_{s,t} \sigma(X_s, \alpha_s) \sigma(X_s, \alpha_s)^* J_{s,t}^{-1}ds$ and the stochastic integral is interpreted in the Skorohod sense, that is, \(\int_0^t \sigma(X_s, \alpha_s)^* J_{s,t}^{-1} J_{0,t} dW_s\) is the divergence of the process \{\(\sigma(X_s, \alpha_s)^* J_{s,t}^{-1} J_{0,t} f_{[0,t]}(s), s \geq 0\}\).

**Proof.** For any $\xi \in \mathbb{R}^n$ with $|\xi| = 1$, let $h^\xi = (DX_t)^* J_{t}^{-1} J_{0,t} \xi$. Then we get

$$\langle DX_t, h^\xi \rangle_H = \langle DX_t, (DX_t)^* J_{t}^{-1} J_{0,t} \xi \rangle_H = \langle DX_t, (DX_t)^* M_t^{-1} J_{0,t} \xi \rangle = J_{0,t} \xi.$$ 

We claim that $h^\xi \in L^{p}(\mathbb{H} \otimes V)$ for any $p \geq 2$. In fact, we have

$$D_t^s h^\xi = (D_t^s (DX_t)^* J_{t}^{-1} J_{0,t} \xi + (DX_t)^* J_{t}^{-1} J_{0,t} (D_t^s J_{t}^{-1}) \xi + (DX_t)^* J_{t}^{-1} J_{0,t} (D_t^s J_{t}^{-1}) \xi - (DX_t)^* J_{t}^{-1} [D_t^s (DX_t), (DX_t)^*]_H + (DX_t, D_t^s (DX_t)^*_H] M_t^{-1} J_{0,t} \xi.$$ 

By Lemma 4.3, Lemma 4.4 and Theorem 4.6 we obtain

$$E_1\|h^\xi\|^p_{H \otimes V} + E_1\|D_t^s h^\xi\|^p_{H \otimes H \otimes V} \leq E\|h^\xi\|^p_H + \sum_{i=1}^d E_1 \int_0^t \|D_t^s h^\xi\|^p_H ds < \infty.$$ 

Notice that $h^\xi = \sigma(X_s, \alpha_s)^* J_{s,t}^{-1} J_{0,t} \xi$. Then, the derivative of $P_t f(x, \alpha)$ can be computed as follows

$$\langle \nabla P_t f(x, \alpha), \xi \rangle = E\langle \nabla \xi [f(X_t, \alpha_t)] \rangle = E\langle \nabla f(X_t, \alpha_t) J_{0,t} \xi \rangle = E \left[\langle \nabla f(X_t, \alpha_t), h^\xi \rangle_H \right] = E_1 \left[\langle Df(X_t, \alpha_t), h^\xi \rangle_H \right] = \delta(h^\xi) |V| = E \left[ f(X_t, \alpha_t) \delta(h^\xi) \right] = E \left[ f(X_t, \alpha_t) \int_0^t \sigma(X_s, \alpha_s)^* J_{s,t}^{-1} J_{0,t} \xi dW_s \right].$$
where the second and forth equalities follow from the chain rule and the sixth equality follows from the integration by parts formula, where the stochastic integral is interpreted in the Skorohod sense.

As an application of the above Bismut type formula, we intend to prove the strong Feller property. That is, we claim that for any \( t > 0 \) and for any bounded Borel measurable function \( f \) on \( \mathbb{R}^n \times \mathcal{S} \), \( P_t f(x, \alpha) \) is bounded and continuous in \((x, \alpha)\). Since \( \mathcal{S} \) is a finite set, it is sufficient to prove that for any \( \alpha \in \mathcal{S} \), \( P_t f(x, \alpha) \) is bounded and continuous with respect to \( x \).

**Theorem 5.2.** Suppose that condition (UHC) holds. Then for any \( f \in \mathcal{B}_0(\mathbb{R}^n \times \mathcal{S}) \), \( t > 0, \alpha \in \mathcal{S}, x \in \mathbb{R}^n \), we have

\[
\lim_{y \to x} |P_t f(y, \alpha) - P_t f(x, \alpha)| = 0.
\]

**Proof.** Fix \( x \in \mathbb{R}^n, \alpha \in \mathcal{S}, t > 0 \). First, we consider the case where \( f \in C^2_b(\mathbb{R}^n \times \mathcal{S}) \).

Applying (20), we have

\[
|\nabla P_t f(x, \alpha)| = \sup_{|\xi| = 1} |\langle \nabla P_t f(x, \alpha), \xi \rangle| \leq \sup_{|\xi| = 1} \mathbb{E} \left| f(X_t, \alpha_t) \int_0^t h_s^\xi dW_s \right| \leq \|f\|_\infty \sup_{|\xi| = 1} \mathbb{E} \left( \int_0^t h_s^\xi dW_s \right)^2,
\]

where \( h_s^\xi = \sigma(X_s, \alpha_s)^* J_{s,t} M_t^{-1} J_{0,t} \xi \). As we have seen, since the process \( h_s^\xi \) is not adapted, the integral is Skorohod integral and it can be estimated as follows:

\[
\mathbb{E} \left( \int_0^t h_s^\xi dW_s \right)^2 \leq \left( \mathbb{E} \left( \int_0^t h_s^\xi dW_s \right)^2 \right)^{1/2} = \mathbb{E} \int_0^t \|h_s^\xi\|^2_{\mathbb{R}^{2d} \otimes V} ds + \mathbb{E} \int_0^t \|D_s h_s^\xi\|^2_{\mathbb{R}^{2d} \otimes \mathbb{R}^d} ds + \mathbb{E} \int_0^t \|D_s h_s^\xi\|^2_{\mathbb{R}^{2d} \otimes \mathbb{R}^d} dr ds
\]

\[
\leq \mathbb{E} \int_0^t \|h_s^\xi\|^2_{\mathbb{R}^{2d} \otimes V} ds + \mathbb{E} \int_0^t \|D_s h_s^\xi\|^2_{\mathbb{R}^{2d} \otimes \mathbb{R}^d} dr ds
\]

\[
= \mathbb{E} \|h_0^\xi\|^2_H + \sum_{l=1}^d \mathbb{E} \int_0^t \|D^l h_s^\xi\|^2_H ds.
\]

Thus, we have

\[
\|\nabla P_t f(x, \alpha)\| \leq C_x \|f\|_\infty,
\]

where the constant \( C_x \) depends on the initial condition \( x \). In fact, we also have that for any \( y \in B_r(x) = \{y \in \mathbb{R}^n : |y - x| \leq r\} \), the following inequality holds

\[
\sup_{y \in B_r(x)} |\nabla P_t f(y, \alpha)| \leq C_{x,r} \|f\|_\infty.
\]

This implies easily for any \(|y - x| \leq 1\),

\[
|P_t f(y, \alpha) - P_t f(x, \alpha)| \leq C_{x,1} \|f\|_\infty |y - x|.
\]

Hence, the theorem holds for any \( f \in \mathcal{B}_0(\mathbb{R}^n \times \mathcal{S}) \) by a standard argument. \( \square \)
REFERENCES

[1] G. Basak, A. Bisi and M. Ghosh, Stability of a Random Diffusion with Linear Drift, *J. Math. Anal. Appl.*, 202 (1996), 604–622.

[2] B. Forster, E. Lütkebohmert and J. Teichmann, Absolutely continuous laws of jump-diffusions in finite and infinite dimensions with applications to mathematical finance, *SIAM J. Math. Anal.*, 40 (2009), 2132–2153.

[3] P. Malliavin, *Stochastic Analysis*, Springer-Verlag, Berlin, 1997.

[4] X. Mao, Stability of stochastic differential equations with Markovian switching, *Stochastic Process. Appl.*, 79 (1999), 45–67.

[5] D. Nualart, *The Malliavin Calculus and Related Topics*, Springer, Berlin, 2006.

[6] G. Yin and C. Zhu, *Hybrid Switching Diffusions: Properties and Applications*, Springer, New York, 2010.

[7] C. Yuan and X. Mao, Asymptotic stability in distribution of stochastic differential equations with Markovian switching, *Stochastic Process. Appl.*, 103 (2003), 277–291.

Received December 2017; revised May 2018.

E-mail address: yaozhong@ualberta.ca
E-mail address: nualart@math.ku.edu
E-mail address: xbsun@jsnu.edu.cn
E-mail address: ycxie@jsnu.edu.cn