DENSITY FUNCTIONS OF DISTRIBUTION DEPENDENT SDES DRIVEN BY LÉVY NOISES

YULIN SONG
Department of Mathematics, Nanjing University, Nanjing, 210093, China
(Communicated by Feng-Yu Wang)

Abstract. By Malliavin calculus for Wiener-Poisson functionals and Lions derivative for probability measures, existence and smoothness of density functions for distribution dependent SDEs with Lévy noises are derived.

1. Introduction.

1.1. Background and notations. Distribution dependent stochastic differential equations (DDSDEs in short), which are also called McKean-Vlasov equations due to McKean’s pioneering work [16] in connection with a mathematical foundation of the Boltzmann equation and the investigation of the Vlasov equation as a limit equation for particle systems [4], have crucial applications in characterizing nonlinear Fokker-Planck equations [1, 2, 11] and intrinsic link to mean field games. For DDSDEs, there exist fruitful results concerning to the well-posedness of solutions, the regularity and ergodicity of associated transition probabilities, see [5, 9, 12, 13, 19, 20, 21, 25] and references therein.

The smoothness of the density of non-stationary (time-dependent) diffusions with or without jumps have been brought into sharp focus for many years, see [6, 10, 14, 24] and references within. On the other hand, the regularity of the transition probabilities of SDEs driven by Lévy noise has also received increasing attentions. For smooth densities of SDEs with degenerate jump noises, we refer to [18, 23, 26] and references therein. In [9], by using Malliavin calculus and Lions derivative for probability measures the authors proved the smoothness of density functions of DDSDEs forced by Brownian motions. However, to the best of our knowledges there is few work to study the regularity of transition probabilities of solutions to DDSDEs driven by Lévy jump processes. So the objective of this paper is to fill the gap in a special case.

Let \( \{W_t\}_{t \in [0,1]} \) be an \( \mathbb{R}^d \)-valued Brownian motion defined on \( (\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P) \), which is a complete probability space with filtration. Let \( \{L_t\}_{t \in [0,1]} \) be a Lévy jump process. Assume that \( W \) and \( L \) are independent. Denote by \( \mathcal{P}(\mathbb{R}^d) \) the collection of probability measures on \( \mathbb{R}^d \). Define

\[
\mathcal{P}_2 := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |y|^2 \mu(dy) < \infty \right\}.
\]

2020 Mathematics Subject Classification. Primary: 60H10; Secondary: 60H07.
Key words and phrases. Distribution dependent SDEs, Lions derivative, Lévy noise, Malliavin calculus, density functions.
This work is supported by NNSFC (No. 11971227, 11790272).
Then $\mathcal{P}_2$ is a Polish space equipped with the Wasserstein distance
\[ W_2(\mu_1, \mu_2) := \inf_{\pi \in \mathcal{C}(\mu_1, \mu_2)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dx, dy) \right)^{1/2}, \]
where $\mathcal{C}(\mu_1, \mu_2)$ is the set of all joint probability measures $\pi$ with marginal laws $\mu_1$ and $\mu_2$. The reference $\sigma-$algebra on $\mathcal{P}_2$ is the Borel $\sigma-$algebra generated by all open sets of $\mathcal{P}_2$.

For measurable maps
\[ b : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^d, \quad A : \mathcal{P}_2 \to \mathbb{R}^d \otimes \mathbb{R}^d, \quad B : \mathcal{P}_2 \to \mathbb{R}^d \otimes \mathbb{R}^d, \]
consider the following DDSDEs on $\mathbb{R}^d$:
\[ X^x_t = x + \int_0^t b(X^x_s, \mu_s)ds + \int_0^t A(\mu_s)dW_s + \int_0^t B(\mu_s)dL_s, \quad (1.1) \]
where $\mu_s$ denotes the distribution of $X^x_s$ under $\mathbb{P}$ and $x$ is an element of $\mathbb{R}^d$.

We will use the following notations frequently:

- Denote by $\mathcal{B}(\mathbb{R}^d)$ the $\sigma$-algebra generated by all open sets of $\mathbb{R}^d$ and by $\mathcal{S}^d$ the unit sphere of $\mathbb{R}^d$. Let $C_c^\infty(\mathbb{R}^d)$ be the class of all smooth functions defined on $\mathbb{R}^d$ with compact supports. For $f \in C_c^\infty(\mathbb{R}^d)$, define $\|f\|_\infty := \sup_{x \in \mathbb{R}^d} |f(x)|$.

- For a function $f$ defined on $\mathbb{R}^d$ or on a subset of $\mathbb{R}^d$, denote its $i$-th derivative by $\nabla^i f$. For a function $g$ defined on $\mathbb{R}^d \times \mathcal{P}_2$, let $\partial_{x} g$ be the $i$-th partial derivative with respect to the first variable.

- For a vector $\xi \in \mathbb{R}^d$ and a matrix $A$, denote their transposes by $\xi^*$ and $A^*$ respectively. The Hilbert-Schmidt norm of $A$ is denoted by $\|A\|_{\text{HS}}$, which is defined by $\|A\|_{\text{HS}} := \sqrt{\sum_{i,j} a_{ij}^2}$.

- The letter $C$ with or without indices will denote an unimportant constant, whose values may change from one appearance to another.

### 1.2. Assumptions and main results.

Let $\Gamma_0 := \{ z \in \mathbb{R}^d : 0 < |z| < 1 \}$. Throughout this work, the Lévy process $L_t$ is assumed to be square integrable and have the representation $L_t = \int_0^t z \tilde{N}(dz, ds)$, where $\tilde{N}(dz, ds)$ denotes the associated martingale measure of $L_t$. We also assume the Lévy measure of $L_t$, which is denoted by $\nu(dz)$, is absolutely continuous with respect to the Lebesgue measure $dz$ on $\mathbb{R}^d$ with a positive density function; that is, there is a function $\kappa : \Gamma_0 \to (0, +\infty)$ such that
\[ \nu(dz) = \kappa(z)dz. \]
Moreover, we assume the following regularity and order conditions ($\mathbb{H}_m^\alpha$):

- For some $c_1 > 0$ and $\alpha \in (0, 2)$,
\[ \lim_{\epsilon \to 0} \epsilon^{\alpha-2} \int_{|z| \leq \epsilon} |z|^2 \nu(dz) = c_1. \]

- For some $m \in \mathbb{N}$, $\kappa \in C^m(\Gamma_0; (0, \infty))$ and any $1 \leq j \leq m$, there exists $c_j > 0$ such that,
\[ |\nabla^j \log \kappa(z)| \leq c_j |z|^{-j}, \quad \forall z \in \Gamma_0. \]

For $\alpha \in (0, 2)$, it is clear that a truncated stable process with Lévy measure $\mathcal{C}|z|^{-d+\alpha} \mathbb{1}_{\Gamma_0}(z)dz$ meets ($\mathbb{H}_m^\alpha$).

We list the assumptions on the coefficients.
There is a constant $C > 0$ such that for all $x \in \mathbb{R}^d$, $\mu_1, \mu_2 \in \mathcal{P}_2$,
\[|b(x, \mu_1) - b(x, \mu_2)| + |A(\mu_1) - A(\mu_2)| + |B(\mu_1) - B(\mu_2)| \leq C \mathcal{W}_2(\mu_1, \mu_2).\]
For any fixed $x \in \mathbb{R}^d$, let $\delta_x$ be the Dirac measure at $x$ and \{\(X_t^x\)\}_{t \in [0, 1]} be the solution to Equ. (1.1). The first result of this paper is:

**Theorem 1.1.** Assume (H) and (H\(^*\)) hold. If
\[\text{Rank}[A(\delta_x), B(\delta_x)] = d, \] then for each $t \in (0, 1]$, $X_t^x$ has a density $\rho_t(x, y)$. Moreover,
\[\lim_{z \to x} \int_{\mathbb{R}^d} |\rho_t(z, y) - \rho_t(x, y)| dy = 0. \]

**Remark 1.** Compared with the distribution independent case (see [18, Theorem 1.1] and [23, Theorem 1.1]), the condition (1.2) does not involve the drift term $b$. That is due to the non-homogeneous essence of Equ.(1.1). For non-homogeneous SDEs driven by Brownian motions, when the coefficients are no more regular in the time component, the “restricted Hörmander’s hypothesis”, which is not related to drift term, is required to investigating smooth densities (see [6] for more details). The same is true in jump cases. An interesting example (see Example 3.1) is given to indicate although DDSDEs (non-homogeneous SDEs) driven by purely jump Lévy noises satisfy the conditions posed in [23, Theorem 1.1] (analogous Hörmander’s condition in homogeneous jump-diffusion cases), there are no density functions for the solutions.

The following result is concerning to the smoothness of density functions for Equ.(1.1). The notation $C_{b,Lip}^{m,m}$ appearing below will be defined in Definition 2.3.

**Theorem 1.2.** For some integer $m$, assume (H\(_m\)) \(b(\cdot, \cdot) \in C_{b,Lip}^{m,m}(\mathbb{R}^d \times \mathcal{P}_2; \mathbb{R}^d)\) and $A(\cdot), B(\cdot) \in C_{b,Lip}^{m}(\mathcal{P}_2; \mathbb{R}^d)$. If
\[\inf_{|u|=1, \mu \in \mathcal{P}_2} \left( |uA(\mu)|^2 + |uB(\mu)|^2 \right) =: c_2 > 0, \]
then for any $k, n \in \mathbb{N}$ with $1 \leq k + n \leq m - 1$ and $\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_n \in \{1, \ldots, d\}$, there are $\gamma_{k,n} > 0$ and $C(k, n) > 0$ such that for all $f \in C^\infty(\mathbb{R}^d)$ and $t \in (0, 1]$,
\[\sup_{x \in \mathbb{R}^d} \left| (\partial_{x_{\alpha_1}} \cdots \partial_{x_{\alpha_k}} f)(X_t^x) \right| \leq C(k, n) \|f\|_{C^\infty} t^{-\gamma_{k,n}}. \]
In particular, if $m = \infty$, then $X_t^x$ has a smooth density $\rho_t(x, y)$ such that
\[(x, y) \mapsto \rho_t(x, y) \in C^\infty(\mathbb{R}^d \times \mathbb{R}^d), \quad t > 0. \]

**Remark 2.** From the proof of this theorem (in Section 3) we will see if the regularity of $y \mapsto \rho_t(x, y)$ is only considered, Equ.(1.1) can be treated as a non-homogeneous SDE. However, when we study the smoothness of $x \mapsto \rho_t(x, y)$, there is an essential difference between DDSDEs and general time dependent SDEs, since the distribution term appearing in the coefficients of DDSDEs is related to the initial value $x$. 

\[\|\partial_x b\|_\infty := \sup_{(x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2} |\partial_x b(x, \mu)| < \infty, \quad \|\partial^2_x b\|_\infty := \sup_{(x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2} |\partial^2_x b(x, \mu)| < \infty.\]
The rest of this paper is organized as follows. In Section 2, some basic elements of Lions derivative and Malliavin calculus for Wiener-Poisson functionals are introduced. Meanwhile, a sufficient condition for the existence of a class of Wiener-Poisson functionals is obtained. In Section 3, the proofs of the main results are given.

2. Preliminaries. In this section, we introduce some basic elements of differentiability of functions on $\mathcal{P}_2$ and Malliavin calculus for Wiener-Poisson functionals.

2.1. Derivative in the Wasserstein space. Now we introduce the notion of differentiability of functions on $\mathcal{P}_2$ which was first introduced by Lions [15] and revised in the notes by Cardaliaguet [7].

Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P})}$ be a complete probability space. Denote by $L^2(\tilde{\Omega}; \mathbb{R}^d)$ the Hilbert space consisting of all square integrable random variables valued on $\mathbb{R}^d$, equipped with the inner product defined as

$$\langle \xi_1, \xi_2 \rangle_{L^2} := \tilde{E}(\xi_1 \cdot \xi_2), \ \forall \xi_1, \xi_2 \in L^2(\tilde{\Omega}; \mathbb{R}^d).$$

Assume $\tilde{\mathcal{F}}$ is rich enough so that for each $\mu \in \mathcal{P}_2$ there exists a random variable $\xi \in L^2(\tilde{\Omega}; \mathbb{R}^d)$ such that $\tilde{P}_\xi = \mu$, i.e. $\mu$ is the distribution of $\xi$ under $\tilde{\mathbb{P}}$.

Let $f : \mathcal{P}_2 \rightarrow \mathbb{R}$ be a function. Define its lifted function over $L^2(\tilde{\Omega}; \mathbb{R}^d)$,

$$\tilde{f}(\xi) := f(\tilde{P}_\xi), \ \forall \xi \in L^2(\tilde{\Omega}; \mathbb{R}^d).$$

**Definition 2.1.** A function $f : \mathcal{P}_2 \rightarrow \mathbb{R}$ is said to be differentiable at $\mu_0 \in \mathcal{P}_2$ if there is a random variable $\xi_0 \in L^2(\tilde{\Omega}; \mathbb{R}^d)$ with $\tilde{P}_{\xi_0} = \mu_0$ such that the lifted function $\tilde{f}$ is Fréchet differentiable at $\xi_0$.

If $f$ is differentiable at $\mu_0$, there exists a linear continuous mapping $D\tilde{f}(\xi_0) : L^2(\tilde{\Omega}; \mathbb{R}^d) \rightarrow \mathbb{R}$ such that

$$\tilde{f}(\xi_0 + \eta) - \tilde{f}(\xi_0) = D\tilde{f}(\xi_0)(\eta) + o(\eta), \ \eta \in L^2(\tilde{\Omega}; \mathbb{R}^d),$$

as $\|\eta\|_{L^2} \rightarrow 0$. By Riesz representation theorem, there is a $(\tilde{\mathbb{P}}$-a.s.) unique random variable $\zeta \in L^2(\tilde{\Omega}; \mathbb{R}^d)$ such that

$$D\tilde{f}(\xi_0)(\eta) = \langle \eta, \zeta \rangle_{L^2},$$

for all $\eta \in L^2(\tilde{\Omega}; \mathbb{R}^d)$. According to Theorem 6.2 and Theorem 6.5 in [7], there is a Borel function $h_0 : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $\zeta = h_0(\xi_0)$, $\tilde{\mathbb{P}}$-a.s. and the function $h_0$ only depends on the law $\mu_0$, not on $\xi_0$ itself. Taking into account the definition of $\tilde{f}$, this allows us to write for any $\xi \in L^2(\tilde{\Omega}; \mathbb{R}^d)$,

$$f(\tilde{P}_\xi) - f(\tilde{P}_{\xi_0}) = \tilde{E}[h_0(\tilde{P}_{\xi_0}) \cdot (\xi - \xi_0)] + o(\|\xi - \xi_0\|_{L^2}). \quad (2.1)$$

We call $\partial_\mu f(\mu_0) \cdot \cdot := h_0(\cdot)$ the derivative of $f$ at $\mu_0$. Note that $\partial_\mu f(\mu_0)$ is only $\mu_0$-a.e. uniquely determined and it allows us to express $D\tilde{f}(\xi_0)$ as a function of any random variable $\xi_0$ with distribution $\mu_0$, irrespective of where this random variable is defined. In particular, the differentiation formula (2.1) is somehow invariant by modification of the probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P})}$ and of the variables $\xi_0$ and $\xi$ used for the representation of $f$, in the sense that $D\tilde{f}(\xi_0)$ always reads as $\partial_\mu f(\mu_0)$, whatever the choice of $\xi_0$ is.

Since we will consider functions $f : \mathcal{P}_2 \rightarrow \mathbb{R}$ which are differentiable at all elements of $\mathcal{P}_2$, we suppose that $\tilde{f} : L^2(\tilde{\Omega}; \mathbb{R}) \rightarrow \mathbb{R}$ is Fréchet differentiable over the whole space $L^2(\tilde{\Omega}; \mathbb{R}^d)$. In this case, we have the derivative $\partial_\mu f(\tilde{P}_\xi)$ defined
such that $\partial_\mu f(\tilde{\mathbb{P}}(\xi))(y) - \partial_\mu f(\tilde{\mathbb{P}}(\xi))(y') \leq K|y - y'|$, $\forall y, y' \in \mathbb{R}^d$.

**Definition 2.2.** A function $f: \mathcal{P}_2 \to \mathbb{R}$ is said to be continuously differentiable with Lipschitz-continuous and bounded derivatives, if there exists for each $\xi \in L^2(\Omega; \mathbb{R}^d)$ a $\tilde{\mathbb{P}}_\xi$-modification of $\partial_\mu f(\tilde{\mathbb{P}}(\xi))(\cdot)$, also denoted by $\partial_\mu f(\tilde{\mathbb{P}}(\xi))(\cdot)$, such that $\partial_\mu f: \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz-continuous and bounded derivatives, if there exists for all $\xi \in L^2(\Omega; \mathbb{R}^d)$ a $\tilde{\mathbb{P}}_\xi$-version of $\partial_\mu f(\tilde{\mathbb{P}}(\xi)): \mathbb{R}^d \to \mathbb{R}^d$ such that:

(i) $|\partial_\mu f(\mu)(y)| \leq C$, for all $\mu \in \mathcal{P}_2$ and $y \in \mathbb{R}^d$;

(ii) $|\partial_\mu f(\mu(y) - \partial_\mu f(\mu')(y')| \leq C(\mathcal{W}_2(\mu, \mu') + |y - y'|)$, for all $\mu, \mu' \in \mathcal{P}_2$ and $y, y' \in \mathbb{R}^d$. In this case, the function $\partial_\mu f$ is considered as the derivative of $f$ and the collection of all such function is denoted by $C^1_{\mathcal{P}_2}(\mathbb{P}_2)$.

**Remark 3.** It is known that (cf. [5, Remark 2.1]) if $f$ belongs to $C^1_{\mathcal{P}_2}(\mathbb{P}_2)$, then the version of $\partial_\mu f(\tilde{\mathbb{P}}(\xi))(\cdot)$ indicated in Definition 2.2 is unique.

For $y_1 \in \mathbb{R}^d$, if the derivative of $\mu \mapsto \partial_\mu f_1(\mu)(y_1)$ exists for each $i \in \{1, \cdots , n\}$, then we denote the second order derivative $(\partial^2_\mu f_1(\mu)(y_1, \cdot), \cdots , \partial^2_\mu f_n(\mu)(y_1, \cdot))$ by $\partial^2_\mu f(\mu)(y_1, \cdot)$.

Due to Remark 3, if $f^{(2)}: \mathcal{P}_2 \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is bounded and Lipschitz continuous with respect to $(\mu, y_1, y_2)$, i.e.

$$|\partial^2_\mu f(\mu)(y_1, y_2)| \leq C,$$

$$|\partial^2_\mu(\mu)(y_1, y_2) - \partial^2_\mu(\mu')(y_1', y_2')| \leq C(\mathcal{W}_2(\mu, \mu') + |y_1 - y_1'| + |y_2 - y_2'|)$$

for each $\mu, \mu' \in \mathcal{P}_2$ and $y_1, y_1', y_2, y_2' \in \mathbb{R}^d$, then it is unique. Generally, by induction we can define the $k$-th order derivative $\partial_k f$ through $(k-1)$-th order derivative of $f$.

For $f \in \mathcal{P}_2 \to \mathbb{R}$, let $\partial_k^\beta f: \mathcal{P}_2 \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d \to \mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d$ be its $k$-th order derivative. For an arbitrary integer $m$ and any multi-index $\beta = (\beta_1, \cdots , \beta_k)$ with $|\beta| = \beta_1 + \cdots + \beta_k = m$,

$$\partial^\beta_{(y_1, \cdots , y_k)} \partial_k^\beta f := \partial^\beta_{y_1} \cdots \partial^\beta_{y_k} \partial_k f: \mathcal{P}_2 \times \mathbb{R}^d \times \cdots \times \mathbb{R}^d \to \mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d$$

denotes the $m$-th order partial derivatives of $\partial_k f$ with respect to $(y_1, \cdots , y_k)$.

For arbitrary integer $n$ and $f = (f_1, \cdots , f_n): \mathcal{P}_2 \to \mathbb{R}^n$, if all elements of $f$ have $k$-th derivatives with respect to $\mu$, then we define $\partial_k^\beta f := (\partial_k^\beta f_1, \cdots , \partial_k^\beta f_n)$.

**Definition 2.3.** (1) A function $U: \mathcal{P}_2 \to \mathbb{R}^n$ is said to be in $C^m_{\mathcal{P}_2}(\mathbb{P}_2; \mathbb{R}^n)$, if for each integer $j$ and multi-index $\beta = (\beta_1, \cdots , \beta_j)$ with $|\beta| + j \leq m$, the derivative $\partial^\beta_{(y_1, \cdots , y_j)} \partial_j^\beta U \exists$ and satisfying that each of these derivatives is bounded and Lipschitz continuous, i.e. there is a constant $C > 0$ such that

$$|\partial^\beta_{(y_1, \cdots , y_j)} \partial_j^\beta U(\mu)(y_1, \cdots , y_j)| \leq C,$$

and

$$|\partial^\beta_{(y_1, \cdots , y_j)} \partial_j^\beta U(\mu)(y_1, \cdots , y_j) - \partial^\beta_{(y_1, \cdots , y_j)} \partial_j^\beta U(\mu')(y_1', \cdots , y_j')| \leq C\left(\mathcal{W}_2(\mu, \mu') + \sum_{k=1}^j |y_k - y'_k|\right)$$
for all $\mu, \mu' \in \mathcal{P}_2$ and $y_1, \ldots, y_j, y'_1, \ldots, y'_j \in \mathbb{R}^d$.

(2) A function $V : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^n$ is said to be in $C_{b, \text{Lip}}^{m,m}(\mathbb{R}^d \times \mathcal{P}_2; \mathbb{R}^n)$, if for each integer $i, j$ and multi-index $\beta = (\beta_1, \ldots, \beta_j)$ with $i + |\beta| + j \leq m$, the derivative $\partial^\beta_i \partial^\beta_j V(x, \mu)(y_1, \ldots, y_j)$ exists and satisfying that each of these derivatives is bounded and Lipschitz continuous, i.e.

$$|\partial^\beta_i \partial^\beta_j V(x, \mu)(y_1, \ldots, y_j)| \leq C,$$

and

$$|\partial^\beta_i \partial^\beta_j (y_1, \ldots, y_j) \partial^\beta_i V(x, \mu)(y_1, \ldots, y_j) - \partial^\beta_i \partial^\beta_j (y_1, \ldots, y_j) \partial^\beta_i V(x', \mu')(y_1, \ldots, y'_j)|$$

$$\leq C \left( |x - x'| + W_2(\mu, \mu') + \sum_{k=1}^j |y_k - y'_k| \right)$$

for all $x, x' \in \mathbb{R}^d, \mu, \mu' \in \mathcal{P}_2$ and $y_1, \ldots, y_j, y'_1, \ldots, y'_j \in \mathbb{R}^d$.

The following result (cf. [5, Lemma 5.1]) says that $\partial_x$ and $\partial_\mu$ can change order if the mixed derivatives are bounded and Lipschitz continuous.

**Lemma 2.4.** Let $g : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^n$ satisfy the following properties:

1. both of $\partial_x g$ and $\partial_\mu g$ are differentiable with respect to $x$ and $\mu$,
2. all of the derivatives up to order 2 are bounded and Lipschitz continuous.

Then the functions $\partial_x \partial_\mu g$ and $\partial_\mu \partial_x g$ are identical.

**Remark 4.** For $V \in C_{b, \text{Lip}}^{m,m}(\mathbb{R}^d \times \mathcal{P}_2; \mathbb{R}^n)$, $\partial^\beta_x$ and $\partial^\beta_\mu$, as well as $\partial^\beta_x$ and $\partial^\beta_\mu$ can commute when they act on $V$. Particularly, for integers $j, k \geq 0$ and multi-index $\beta = (\beta_1, \ldots, \beta_j)$ with $|\beta| \leq j$ and $|\beta| + j + k \leq m$, $\partial^k_x \partial^\beta_\mu \partial^\beta_x V$ coincides with $\partial^\beta_\mu \partial_x \partial^\beta_x \partial^k V$.

### 2.2. Malliavin calculus

In this subsection, we recall some basic facts about Malliavin’s approach to Malliavin calculus for jump processes (cf. [3, 23] etc.).

Let $\Lambda \subset \mathbb{R}^d$ be an open set containing the origin. Let us define

$$\Lambda_0 := \Lambda \setminus \{0\}, \quad g(z) := 1 \vee \text{dis}(z, \Lambda_0)^{-1}, \quad \text{(2.2)}$$

where $\text{dis}(z, \Lambda_0^c)$ is the distance of $z$ to the complement of $\Lambda_0$. Let $\Omega$ be the canonical space of all points $\omega = (w, \mu)$, where

- $w : [0, 1] \to \mathbb{R}^d$ is a continuous function with $w(0) = 0$;
- $\mu$ is an integer-valued measure on $[0, 1] \times \Lambda_0$ with $\mu(A) < +\infty$ for any compact set $A \subset [0, 1] \times \Lambda_0$.

Define the canonical process on $\Omega$ as follows: for $\omega = (w, \mu)$,

$$W_t(\omega) := w(t), \quad N(\omega; dz, dt) := \mu(\omega; dz, dt) := \mu(dz, dt).$$

Let $(\mathcal{F}_t)_{t \in [0, 1]}$ be the smallest right-continuous filtration on $\Omega$ such that $W$ and $N$ are optional. In the following, we write $\mathcal{F} := \mathcal{F}_1$, and endow $(\Omega, \mathcal{F})$ with the unique probability measure $\mathbb{P}$ such that

- $W$ is a standard $d$-dimensional Brownian motion;
- $N$ is a Poisson random measure with intensity $\nu(dz)dt$, where $\nu(dz) = \kappa(z)dz$ with $\kappa \in C^1(\Lambda_0; (0, \infty)), \int_{\Lambda_0} (1 \wedge |z|^2) \kappa(z)dz < +\infty, \quad |\nabla \log \kappa(z)| \leq C g(z)$,

where $g(z)$ is defined by (2.2);
• $W$ and $N$ are independent.

In the following we write
\[ \mathcal{N}(dz, ds) := N(dz, ds) - \nu(ds). \]

Let $p \geq 1$, $i = 1, 2$ and $k$ be an integer. We introduce the following spaces for later use.

- $L^p(\Omega)$: The space of all $\mathcal{F}$-measurable random variables with finite norm:
  \[ \|F\|_p := \left[ \mathbb{E}|F|^p \right]^\frac{1}{p}. \]

- $L^p_{\xi}(\Omega)$: The space of all predictable processes $\xi : \Omega \times [0, 1] \times \Lambda_0 \to \mathbb{R}^k$ with finite norm:
  \[ \|\xi\|_{L^p_{\xi}} := \left[ \mathbb{E} \left( \int_0^1 |\xi(s, z)|^p \nu(ds) \right)^\frac{p}{2} \right]^\frac{1}{p} < \infty. \]

- $H_p$: The space of all measurable adapted processes $h : \Omega \times [0, 1] \to \mathbb{R}^d$ with finite norm:
  \[ \|h\|_{\mathcal{H}_p} := \left[ \mathbb{E} \left( \int_0^1 |h(s)|^2 ds \right)^\frac{1}{2} \right] < +\infty. \]

- $V_p$: The space of all predictable processes $v : \Omega \times [0, 1] \times \Lambda_0 \to \mathbb{R}^d$ with finite norm:
  \[ \|v\|_{V_p} := \|\nabla_z v\|_{L^p_{\xi}} + \|v\|_{L^p_{\xi}} < \infty, \]
  where $g(z)$ is defined by (2.2). Below we shall write
  \[ \mathcal{H}_{\infty-} := \cap_{p \geq 1} \mathcal{H}_p, \quad \mathcal{V}_{\infty-} := \cap_{p \geq 1} \mathcal{V}_p. \]

- $H_{\infty}$: The space of all bounded measurable adapted processes $h : \Omega \times [0, 1] \to \mathbb{R}^d$.

- $V_0$: The space of all predictable processes $v : \Omega \times [0, 1] \times \Lambda_0 \to \mathbb{R}^d$ with the following properties: (i) $v$ and $\nabla_z v$ are bounded; (ii) there exists a compact subset $U \subset \Lambda_0$ such that
  \[ v(t, z) = 0, \quad \forall z \notin U. \]

Let $m$ be an integer and $C^\infty_{\mathcal{P}}(\mathbb{R}^m)$ be the class of all smooth functions on $\mathbb{R}^m$ which together with all of the derivatives has at most polynomial growth. Let $F C^\infty_{\mathcal{P}}$ be the class of all Wiener-Poisson functionals on $\Omega$ with the following form:

\[ F = f(W(h_1), \ldots, W(h_{m_1}), N(g_1), \ldots, N(g_{m_2})), \]

where $f \in C^\infty_{\mathcal{P}}(\mathbb{R}^{m_1+m_2})$, $h_1, \ldots, h_{m_1} \in H_0$ and $g_1, \ldots, g_{m_2} \in V_0$ are non-random and real-valued, and
\[ W(h_i) := \int_0^1 \langle h_i(s), dW_s \rangle_{\mathbb{R}^d}, \quad N(g_j) := \int_0^1 \int_{\Lambda_0} g_j(s, z)N(dz, ds). \]

For any $p > 1$ and $\Theta = (h, v) \in H_p \times V_p$, let us define
\[ D_\Theta F := \sum_{i=1}^{m_1} (\partial_i f)(\cdot) \int_0^1 \langle h(s), h_i(s) \rangle_{\mathbb{R}^d} ds \]
\[ + \sum_{j=1}^{m_2} (\partial_{j+m_1} f)(\cdot) \int_0^1 \int_{\Lambda_0} \langle v(s, z), \nabla_z g_j(s, z) \rangle_{\mathbb{R}^d} N(dz, ds), \]
where “(·)” stands for \( W(h_1), \cdots, W(h_{m_1}), N(g_1), \cdots, N(g_{m_2}) \).

**Definition 2.5.** For \( p > 1 \) and \( \Theta = (h, \nu) \in \mathbb{H}_p \times \mathbb{V}_p \), we define the first order Sobolev space \( \mathbb{D}^{1,p}_\Theta \) being the completion of \( FC^\infty_p \) in \( L^p(\Omega) \) with respect to the norm:

\[
\|F\|_{\Theta,1,p} := \|F\|_{L^p} + \|D\Theta F\|_{L^p}.
\]

We have the following integration by parts formula (cf. [23, Theorem 2.9]).

**Theorem 2.6.** Given \( \Theta = (h, \nu) \in \mathbb{H}_\infty \times \mathbb{V}_\infty \) and \( p > 1 \), for any \( F \in \mathbb{D}^{1,p}_\Theta \), we have

\[
\mathbb{E} D\Theta F = \mathbb{E}(F\delta(\Theta)),
\]

where

\[
\delta(\Theta) := \int_0^1 \langle h(s), dW_s \rangle - \int_0^1 \int_{\Lambda_0} \frac{\operatorname{div}(\kappa \nu)(s,z)}{\kappa(z)} \tilde{N}(dz,ds),
\]

and \( \operatorname{div}(\kappa \nu) := \sum_{i=1}^d \partial_{x_i}(\kappa \nu) \) stands for the divergence.

The following result provides a sufficient condition for the existence of density functions for a class of Wiener-Poisson functionals.

**Theorem 2.7.** Let \( F = (F^1, \cdots, F^d) \) be a Wiener-Poisson functional. Assume there exist \( \Theta_1, \cdots, \Theta_d \in \mathbb{H}_\infty \times \mathbb{V}_\infty \) such that for each \( i, j, k \in \{1, \cdots, d\}, D\Theta_i F^i \) is in \( \mathbb{D}^{1,p}_\Theta \) for some \( p > 1 \). If matrix \( D\Theta F := (D\Theta_i F^i)_{1 \leq i,j \leq d} \) is invertible almost surely, then the distribution of \( F \) is absolutely continuous with respect to Lebesgue measure on \( \mathbb{R}^d \).

**Proof.** Let \( \text{GL}(d) \) be the set of all \( d \times d \)-matrix. Define

\[
\mathbb{M}_n := \{ M \in \text{GL}(d) : |M| \leq n, \det(M) \geq 1/n \}.
\]

Then \( \mathbb{M}_n \) is a compact subset of \( \text{GL}(d) \). Let \( \varphi_n \in C^\infty(\mathbb{R}^d) \) be a smooth function with

\[
\varphi_n|_{\mathbb{M}_n^c} = 1, \quad \varphi_n|_{\mathbb{M}_n^c,1} = 0, \quad 0 \leq \varphi_n \leq 1.
\]

For each \( n \in \mathbb{N} \), define a finite measure \( \mu_n \) by

\[
\mu_n(A) := \mathbb{E} \left( I_A(F) \varphi_n(D\Theta F) \right), \quad \forall A \in \mathcal{B}(\mathbb{R}^d).
\]

For each \( f \in C^\infty_c(\mathbb{R}^d) \), by the chain rule (c.f. [22, Lemma 2.4]), we have

\[
D\Theta f(D\Theta F) = \nabla f(F) D\Theta F.
\]

Hence,

\[
\nabla f(F) = D\Theta f(D\Theta F)^{-1}, \tag{2.3}
\]

where \( (D\Theta F)^{-1} \) stands for the inverse matrix of \( D\Theta F \). Thus, by Theorem 2.6 we have for \( i = 1, \cdots, d, \)

\[
\int_{\mathbb{R}^d} \partial_i f(y) \mu_n(dy) = \mathbb{E} (\partial_i f(F) \varphi_n(D\Theta F))
\]

\[
= \sum_{j=1}^d \mathbb{E} \left( D\Theta_j(f(F)) (D\Theta F)^{-1}_{ji} \varphi_n(D\Theta F) \right)
\]

\[
= \sum_{j=1}^d \mathbb{E} \left[ D\Theta_j \left( f(F) (D\Theta F)^{-1}_{ji} \varphi_n(D\Theta F) \right) \right].
\]
\[- \sum_{j=1}^{d} \mathbb{E} \left[ f(F) D_{\Theta_j} \left( (D_{\Theta} F)_{ji}^{-1} \varphi_n(D_{\Theta} F) \right) \right] \]
\[= \sum_{j=1}^{d} \mathbb{E} \left[ f(F) \left( (D_{\Theta} F)_{ji}^{-1} \varphi_n(D_{\Theta} F) \delta(\Theta_j) + \left( (D_{\Theta} F)^{-1} D_{\Theta_j}(D_{\Theta} F) (D_{\Theta} F)^{-1} \varphi_n(D_{\Theta} F) \right)_{ji} - (D_{\Theta} F)^{-1} \sum_{l,k=1}^{d} (\nabla \varphi_n(D_{\Theta} F))_{kl} D_{\Theta_j} D_{\Theta_l} F^k \right) \right], \quad (2.4)\]

where we have used the fact
\[D_{\Theta_j} (D_{\Theta} F)^{-1} = - (D_{\Theta} F)^{-1} D_{\Theta_j} (D_{\Theta} F) (D_{\Theta} F)^{-1}.\]

Then we have
\[\left| \int_{\mathbb{R}^d} \partial_i f(y) \mu_n(dy) \right| \leq C_n \|f\|_{\infty},\]

for some \(C_n\) depending on \(n\) and independent of \(f\). Due to Lemma 2.1.1 in [17], \(\mu_n\) is absolutely continuous with respect to the Lebesgue measure on \(\mathbb{R}^d\). That is, for any \(A \in \mathcal{B}(\mathbb{R}^d)\) with \(\int_{\mathbb{R}^d} I_A(y)dy = 0\),
\[\mathbb{E}(I_A(F)\varphi_n(D_{\Theta} F)) = 0.\]

By dominated convergence theorem and letting \(n \to \infty\), we obtain
\[\mathbb{E}(I_A(F)) = 0,\]

which means that the distribution of \(F\) is absolutely continuous with respect to the Lebesgue measure. \(\square\)

3. Proofs of main results.

**Lemma 3.1.** Assume (H) holds. Then for any \(p \geq 2\),
\[\mathbb{E} \left( \sup_{t \in [0,1]} |X^x_t|^p \right) \leq C_p (1 + |x|^p), \quad \text{and} \quad \lim_{t \to s} \mathcal{W}_2(\mu_s, \mu_s) = 0.\]

where \(C_p\) is a constant depending on \(p\).

**Proof.** For any \(p \geq 2\),
\[\mathbb{E} \left( \sup_{t \in [0,1]} |X^x_t|^p \right) \leq C_p |x|^p + C_p \int_0^1 \mathbb{E}|b(X^x_s, \mu_s)|^p ds + C_p \left( \int_0^1 \|A(\mu_s)\|_{\text{HS}}^p ds \right)^{\frac{p}{2}} \]
\[+ C_p \left( \int_0^1 \|B(\mu_s)\|_{\text{HS}}^2 ds \int_\Gamma |z|^2 dz \right)^{\frac{p}{2}} \]
\[\leq C_p (|x|^p + |b(0, \delta_0)|^p + \|A(\delta_0)\|_{\text{HS}}^p + \|B(\delta_0)\|_{\text{HS}}^p) \]
\[+ C_p \int_0^1 \mathbb{E}|b(X^x_s, \mu_s) - b(0, \delta_0)|^p ds \]
\[+ C_p \int_0^1 \|A(\mu_s) - A(\delta_0)\|_{\text{HS}}^p ds \]
\[+ C_p \int_0^1 \|B(\mu_s) - B(\delta_0)\|_{\text{HS}}^p ds \]
Proposition 1. For some integer \( m \geq 2 \), assume \( b(\cdot, \cdot) \in C^{m,m}_{b, Lip}(\mathbb{R}^d \times \mathcal{P}_2; \mathbb{R}^d) \) and \( A(\cdot), B(\cdot) \in C^{m}_{b, Lip}(\mathcal{P}_2; \mathbb{R}^d) \). Then \( X_t^x \) is \( m \)-times differentiable with respect to \( x \) and for any \( p \geq 2 \),

\[
\sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0,1]} |\nabla^k X_t^x|^p \right) < \infty, \quad k = 1, \ldots, m. \tag{3.2}
\]

Proof. For \( \varepsilon > 0 \), let

\[
X_t^{x+\varepsilon y} = x + \varepsilon y + \int_0^t b(X_s^{x+\varepsilon y}, \mu_s^x)ds + \int_0^t A(\mu_s^x)dW_s + \int_0^t B(\mu_s^x)dL_s, \quad t \in [0,1],
\]

where \( \mu_s^x \) stands for the distribution of \( X_s^{x+\varepsilon y} \). Then for any \( p \geq 2 \), by the Lipschitz continuity of \( b, A \) and \( B \) and Gronwall’s inequality it is easy to prove

\[
\mathbb{E} \left( \sup_{t \in [0,1]} |X_t^{x+\varepsilon y} - X_t^x|^p \right) \leq C_p |y|^p e^{p}. \tag{3.3}
\]

Denote by \( (\tilde{W}, \tilde{L}) \) a copy of \((W, L)\) on a probability space \((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\), and by \( \{\tilde{X}^x\} \) the solution to the SDE for \( X^x \), but driven by \( \tilde{W} \) and \( \tilde{L} \). Then \((\tilde{W}, \tilde{L}, \tilde{X}^x)\), defined over \((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\), is an independent copy of \((W, L, X^x)\). In what follows, \( \tilde{\mathbb{E}} \) stands for the expectation with respect to the probability \( \tilde{\mathbb{P}} \).

We first prove that for any \( y \in \mathbb{R}^d \) the directional derivative defined by

\[
\nabla_y X_t^x := \lim_{\varepsilon \to 0} \varepsilon^{-1} \left( X_t^{x+\varepsilon y} - X_t^x \right) \quad \text{in } L^2(\Omega)
\]

exists and satisfies

\[
\nabla_y X_t^x = y + \int_0^t \partial_y b(X_s^x, \mu_s) \nabla_y X_s^x ds + \int_0^t \tilde{\mathbb{E}} \left( \partial_y b(X_s^x, \mu_s)(\tilde{X}_s^x) \nabla_y \tilde{X}_s^x \right) ds
\]
By the similar argument as above, we have

\[ + \int_0^t \tilde{E} \left( \partial_\mu A(\mu_s)(\tilde{X}_s^x) \nabla_y \tilde{X}_s^x \right) dW_s + \int_0^t \tilde{E} \left( \partial_\mu B(\mu_s)(\tilde{X}_s^x) \nabla_y \tilde{X}_s^x \right) dL_s, \]

where \( \nabla_y \tilde{X}_s^x \) is the directional derivative of \( \tilde{X}_s^x \).

Observe that

\[ b(X_s^{x+ey}, \mu_s^*) - b(X_s^x, \mu_s) \]

\[ = \int_0^1 \partial_\lambda \left( b(X_s^x + \lambda(X_s^{x+ey} - X_s^x), \mu_s^*) \right) d\lambda + \int_0^1 \partial_\lambda \left( b(X_s^x, \mu_s^{\epsilon,\lambda}) \right) d\lambda \]

\[ = \alpha_s^* (X_s^{x+ey} - X_s^x) + \tilde{E} \left( \beta_s^* (\tilde{X}_s^{x+ey} - \tilde{X}_s^x) \right), \]

where \( \mu_s^{\epsilon,\lambda} \) stands for the distribution of \( X_s^x + \lambda(X_s^{x+ey} - X_s^x) \) and

\[ \alpha_s^* := \int_0^1 \partial_\mu b(X_s^x + \lambda(X_s^{x+ey} - X_s^x), \mu_s^*) d\lambda, \]

\[ \beta_s^* := \int_0^1 \partial_\mu b(X_s^x, \mu_s^{\epsilon,\lambda}))(\tilde{X}_s^x + \lambda(\tilde{X}_s^{x+ey} - \tilde{X}_s^x)) d\lambda. \]

Moreover, for any \( p \geq 2 \) by the Lipschitz continuity of \( \partial_\mu b \) and \( \partial_\mu b \) we have

\[ \mathbb{E} \left( \sup_{s \in [0,1]} |\alpha_s^* - \partial_\mu b(X_s^x, \mu_s)|^p \right) \]

\[ \leq \mathbb{E} \left( \sup_{s \in [0,1]} \int_0^1 |\partial_\mu b(X_s^x + \lambda(X_s^{x+ey} - X_s^x), \mu_s^*) - \partial_\mu b(X_s^x, \mu_s)|^p d\lambda \right) \]

\[ \leq C_p \mathbb{E} \left( \sup_{s \in [0,1]} |X_s^{x+ey} - X_s^x|^p + \sup_{s \in [0,1]} \mathbb{W}_2^p(\mu_s^*, \mu_s) \right) \]

\[ \leq C_p \mathbb{E} \left( \sup_{s \in [0,1]} |X_s^{x+ey} - X_s^x|^p \right) \leq C_p |y|^p e^p, \quad (3.4) \]

and

\[ \mathbb{E} \left( \sup_{s \in [0,1]} |\beta_s^* - \partial_\mu b(X_s^x, \mu_s^*)(\tilde{X}_s^x)|^p \right) \]

\[ \leq C_p \mathbb{E} \left( \sup_{s \in [0,1]} |\tilde{X}_s^{x+ey} - \tilde{X}_s^x|^p + \sup_{s \in [0,1]} \mathbb{W}_2^p(\mu_s^{\epsilon,\lambda}, \mu_s^*) \right) \]

\[ \leq C_p \mathbb{E} \left( \sup_{s \in [0,1]} |\tilde{X}_s^{x+ey} - \tilde{X}_s^x|^p \right) \leq C_p |y|^p e^p. \quad (3.5) \]

By the similar argument as above, we have

\[ A(\mu_s^*) - A(\mu_s) = \mathbb{E} \left( \gamma_s^* (\tilde{X}_s^{x+ey} - \tilde{X}_s^x) \right), \]

\[ B(\mu_s^*) - B(\mu_s) = \mathbb{E} \left( \vartheta_s^* (\tilde{X}_s^{x+ey} - \tilde{X}_s^x) \right), \]

with

\[ \gamma_s^* = \int_0^1 \partial_\mu A(\mu_s^{\epsilon,\lambda})(\tilde{X}_s^x + \lambda(\tilde{X}_s^{x+ey} - \tilde{X}_s^x)) d\lambda, \]

\[ \vartheta_s^* = \int_0^1 \partial_\mu B(\mu_s^{\epsilon,\lambda})(\tilde{X}_s^x + \lambda(\tilde{X}_s^{x+ey} - \tilde{X}_s^x)) d\lambda, \]
Consider the following equation:

\[
\Delta \left( \sup_{\alpha \geq 0} |\gamma^\alpha_0 - \partial_{\mu} A(\mu_0)(X_s^\alpha)|^p \right) + \int_0^t \mathbb{E} \left( \sup_{\alpha \geq 0} |\gamma^\alpha_s - \partial_{\mu} B(\mu_s)(X_s^\alpha)|^p \right) \leq C_p |y|^p \psi_p. \quad (3.6)
\]

By classical Picard’s iteration, it is route to prove that there is a unique solution and an independent copy \( Y \) of \( Y \), defined over \( (\Omega, \mathcal{F}, \mathbb{P}) \). Then

\[
X_t^{x+e_y} - X_t^{\alpha} - eY_t^{x}(y)
\]

\[
= \int_0^t \partial_{\alpha} b(X_s^\alpha, \mu_s)(X_s^{\alpha+e_y} - X_s^\alpha - eY_s^{x}(y)) ds
\]

\[
+ \int_0^t \mathbb{E} \left( \partial_{\alpha} b(X_s^\alpha, \mu_s)(X_s^{x+e_y} - X_s^x) \right) ds
\]

\[
+ \int_0^t \mathbb{E} \left( \partial_{\alpha} A(\mu_s)(X_s^\alpha)(X_s^{x+e_y} - X_s^x) \right) ds
\]

By Burkholder-Davis-Gundy inequality one can obtain

\[
\mathbb{E} \left( \sup_{t \in [0, 1]} |X_t^{x+e_y} - X_t^{\alpha} - eY_t^{x}(y)|^2 \right)
\]

\[
\leq C \int_0^1 \mathbb{E} \left( \sup_{s \leq t} |X_s^{x+e_y} - X_s^{\alpha} - eY_s^{x}(y)|^2 \right) dt + C \int_0^1 \mathbb{E} \left( \sup_{s \leq t} |X_s^{x+e_y} - X_s^{\alpha} - eY_s^{x}(y)|^2 \right) dt
\]

\[
+ C \int_0^1 \mathbb{E} \left( |X_s^{x+e_y} - X_s^{\alpha}|^2 |\alpha_s - \partial_{\alpha} b(X_s^\alpha, \mu_s)|^2 \right) ds
\]

\[
+ C \mathbb{E} \left( |X_s^{x+e_y} - X_s^{\alpha}|^2 |\beta_s - \partial_{\beta} b(X_s^\alpha, \mu_s)(X_s^\alpha)|^2 \right) ds
\]

\[
+ C \mathbb{E} \left( |X_s^{x+e_y} - X_s^{\alpha}|^2 |\gamma_s - \partial_{\gamma} A(\mu_s)(X_s^\alpha)|^2 \right) ds
\]

\[
+ C \mathbb{E} \left( |X_s^{x+e_y} - X_s^{\alpha}|^2 |\delta_s - \partial_{\delta} B(\mu_s)(X_s^\alpha)|^2 \right) ds
\]
Gronwall’s inequality, together with (3.3), (3.4), (3.5) and (3.6), yields

\[ \leq C \int_0^1 \mathbb{E} \left( \sup_{s \leq t} |X_s^{x+\epsilon y} - X_s^x - cY_s^x(y)|^2 \right) dt + C \mathbb{E} \left( \sup_{s \in [0,1]} |X_s^{x+\epsilon y} - X_s^x|^4 \right)^{\frac{1}{2}} \times \left\{ \mathbb{E} \left( \sup_{s \in [0,1]} |\alpha_s^x - \partial_x b(X_s^x, \mu_s)|^4 \right)^{\frac{1}{2}} + \mathbb{E} \left( \sup_{s \in [0,1]} |\beta_s^x + \partial_y b(X_s^x, \mu_s)(X_s^x)|^4 \right)^{\frac{1}{2}} + \mathbb{E} \left( \sup_{s \in [0,1]} |\gamma_s^x - \partial_\mu A(\mu_s)(\tilde{X}_s^x)|^4 \right)^{\frac{1}{2}} + \mathbb{E} \left( \sup_{s \in [0,1]} |\varphi_s^x - \partial_\mu B(\mu_s)(\tilde{X}_s^x)|^4 \right)^{\frac{1}{2}} \right\} . \]

Thus, Gronwall’s inequality, together with (3.3), (3.4), (3.5) and (3.6), yields

\[ \mathbb{E} \left( \sup_{t \in [0,1]} |X_t^{x+\epsilon y} - X_t^x - cY_t^x(y)|^2 \right) \leq C |\epsilon|^2. \]  

(3.7)

Note that \( Y_t^x(y) \) is linear function with respect to \( y \), which combining (3.7), can claim the Jacobian of \( X_t^x \) denoted by \( Y_t^x \) exists and satisfies

\[
\nabla X_t^x = I + \int_0^t \partial_x b(X_s^x, \mu_s) \nabla X_s^x ds + \int_0^t \tilde{\mathbb{E}} \left( \partial_\mu b(X_s^x, \mu_s)(\tilde{X}_s^x) \nabla X_s^x \right) ds + \sum_{i=1}^d \int_0^t \tilde{\mathbb{E}} \left( \partial_\mu A_i(\mu_s)(\tilde{X}_s^x) \nabla X_s^x \right) dW_s^i + \sum_{i=1}^d \int_0^t \tilde{\mathbb{E}} \left( \partial_\mu B_i(\mu_s)(\tilde{X}_s^x) \nabla X_s^x \right) dL_s^i. \]  

(3.8)

For any \( p \geq 2 \), by Burkholder-Davis-Gundy inequality and Jensen’s inequality we arrive at

\[ \mathbb{E} \left( \sup_{t \in [0,1]} |\nabla X_t^x|^p \right) \leq C_p + C_p \int_0^1 \mathbb{E} \left( \sup_{s \in [0,t]} |\nabla X_s^x|^p \right) dt. \]

Hence,

\[ \mathbb{E} \left( \sup_{t \in [0,1]} |\nabla X_t^x|^p \right) < \infty. \]  

(3.9)

That is, (3.2) holds for \( k = 1 \).

Now suppose that for some \( n \in \mathbb{N} \), (3.2) holds for all \( k \) less than \( n \). Repeating the above argument \( n \)-times again, we have

\[
\nabla^{n+1} X_t^x = \int_0^t \partial_x b(X_s^x, \mu_s) \nabla^{n+1} X_s^x ds + \int_0^t \tilde{\mathbb{E}} \left( \partial_\mu b(X_s^x, \mu_s)(\tilde{X}_s^x) \nabla^{n+1} \tilde{X}_s^x \right) ds + \sum_{i=1}^d \int_0^t \tilde{\mathbb{E}} \left( \partial_\mu A_i(\mu_s)(\tilde{X}_s^x) \nabla^{n+1} \tilde{X}_s^x \right) dW_s^i + \sum_{i=1}^d \int_0^t \tilde{\mathbb{E}} \left( \partial_\mu B_i(\mu_s)(\tilde{X}_s^x) \nabla^{n+1} \tilde{X}_s^x \right) dL_s^i + \int_0^t I_1(s) ds + \sum_{i=1}^d \int_0^t I_{1,2}(s) dW_s^i + \sum_{i=1}^d \int_0^t I_{1,3}(s) dL_s^i,
\]

where

\[ I_1(s) = \int_0^s \mathbb{E} \left( \sup_{r \in [0,t]} |\nabla^{n+1} X_r^x|^p \right) dr \]

and

\[ I_{1,2}(s) = \int_0^s \mathbb{E} \left( \sup_{r \in [0,t]} |\nabla^{n+1} X_r^x|^p \right) dr \]

for any \( p \geq 2 \).
where \( I_1(s), I_{1,2}(s) \) and \( I_{1,3}(s) \) are \( \mathbb{R}^{(k+2)d} \)-valued “polynomial” of \( \nabla^2 X_t^x, \ldots, \nabla^k X_t^x \) and the derivatives of \( \partial_b b \) and \( \partial_b b \), the derivatives of \( \partial_i A_i \) and the derivatives of \( \partial_i B_i \) respectively. Thus, by Hölder’s inequality and the assumption that (3.2) holds for \( k \leq n \), we immediately obtain

\[
\mathbb{E} \left( \sup_{t \in [0,1]} |\nabla^{n+1} X_t^x|^p \right) \leq C_p + C_p \int_0^1 \mathbb{E} \left( \sup_{s \in [0,t]} |\nabla^{k+1} X_s^x|^p \right) ds.
\]

Gronwall’s inequality yields

\[
\mathbb{E} \left( \sup_{t \in [0,1]} |\nabla^{n+1} X_t^x|^p \right) < \infty.
\]

By induction, (3.2) holds for all \( k \in \mathbb{N} \).

Let \( \{J_t(x)\}_{t \in [0,1]} \) be the solution to the following linear equation:

\[
J_t(x) = I + \int_0^t \partial_x b(X_s^x,\mu_s)J_s(x)ds, \quad \forall t \in [0,1],
\]

where \( I \) stands for the identity matrix. Let \( \{K_t(x)\}_{t \in [0,1]} \) solve

\[
K_t(x) = I - \int_0^t K_s(x)\partial_x b(X_s^x,\mu_s)ds, \quad \forall t \in [0,1].
\]

Then it is easy to verify that \( J_t(x)K_t(x) = I \) for each \( t \in [0,1] \) and

\[
\sup_{t \in [0,1], x \in \mathbb{R}^d} |J_t(x)| \leq e^{\|\partial_x b\|_\infty}, \quad \sup_{t \in [0,1], x \in \mathbb{R}^d} |K_t(x)| \leq e^{\|\partial_x b\|_\infty}.
\]

Let \( \zeta(z) \) be a nonnegative smooth function with

\[
\zeta(z) = |z|^3, \quad |z| \leq \frac{1}{4}, \quad \zeta(z) = 0, \quad |z| > \frac{1}{2}.
\]

For \( j = 1, \ldots, d \), take

\[
h_j(s) = A(\mu_s)^*(K_s^*(x))_{.,j}, \quad v_j(s,z) = B(\mu_s)^*(K_s^*(x))_{.,j}\zeta(z), \quad \Theta_j = (h_j, v_j),
\]

where \( (K_s^*)_{.,j} \) stands for the \( j \)-th column vector of \( K_s^* \). Then \( \Theta_j = (h_j, v_j) \in \mathbb{H}_{\infty-} \times \mathbb{V}_{\infty-} \) for \( 1 \leq j \leq d \). It follows from [21, Proposition 3.3] that for any \( p > 1 \) and \( i \in \{1, \ldots, d\} \), \( X_t^x \) is in \( \mathbb{D}^{1,p}_{\Theta_i} \) and \( D_{\Theta_i} X_t^x \) satisfies

\[
D_{\Theta_i} X_t^x = \int_0^t \partial_x b(X_s^x,\mu_s)D_{\Theta_i} X_s^x ds + \int_0^t A(\mu_s)h_i(s)ds
\]

\[
+ \int_0^t \int_{\Gamma_0} B(\mu_s)v_i(s,z)N(dz,ds), \quad \forall t \in [0,1].
\]

Moreover, by (3.10), (3.11), (3.15) and using Itô’s formula one has

\[
D_{\Theta_i} X_t^x = J_t(x) \left( \int_0^t K_s(x)A(\mu_s)h_i(s)ds + \int_0^t \int_{\Gamma_0} K_s(x)B(\mu_s)v_i(s,z)N(dz,ds) \right).
\]

Define

\[
\hat{\Theta} := (\Theta_1, \cdots, \Theta_d), \quad (D_{\Theta_i} X_t^x)_{ij} := D_{\Theta_j} X_t^{x,i},
\]

and

\[
\Sigma_t^x := \int_0^t K_s(x)A(\mu_s)A(\mu_s)^*K_s(x)^*ds.
\]
+ \int_0^t K_s(x)B(\mu_s)B(\mu_s)^*K_s(x)^*\zeta(z)N(dz,ds), \quad (3.18)

where $X^{x,i}_t$ denotes the $i$-th element of $X^x_t$. Then, by (3.16) we have

$$D_\Theta X^x_t = J_t(x)\Sigma^x_t. \quad (3.19)$$

**Lemma 3.2.** Assume $(\mathbb{H}^m)$, $b(\cdot, \cdot) \in C_{b,Lip}^{m,m}(\mathbb{R}^d \times \mathcal{P}_2; \mathbb{R}^d)$ and $A(\cdot), B(\cdot) \in C_{b,Lip}^m(\mathcal{P}_2; \mathbb{R}^d)$. For any $n, k \in \mathbb{N}$ with $k+n \leq m$, $j_1, \cdots, j_n \in \{1, \cdots, d\}$ and $p \geq 2$, we have

$$\sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0,1]} \left| D_{\Theta_{j_1}} \cdots D_{\Theta_{j_n}} X^x_t \right|^p \right) < \infty, \quad (3.20)$$

$$\sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0,1]} \left| D_{\Theta_{j_1}} \cdots D_{\Theta_{j_n}} \nabla^k X^x_t \right|^p \right) < \infty, \quad (3.21)$$

$$\sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0,1]} \left| D_{\Theta_{j_1}} \cdots D_{\Theta_{j_n}} J_t(x) \right|^p \right) < \infty, \quad (3.22)$$

$$\sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0,1]} \left| D_{\Theta_{j_1}} \cdots D_{\Theta_{j_n}} K_t(x) \right|^p \right) < \infty, \quad (3.23)$$

$$\sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0,1]} \left| D_{\Theta_{j_1}} \cdots D_{\Theta_{j_n}} \Sigma^x_t \right|^p \right) < \infty. \quad (3.24)$$

Furthermore, if $(\mathbb{H}^m)$ holds with the same $m$, then for any $n \leq m-1$, we also have

$$\sup_{x \in \mathbb{R}^d} \mathbb{E} \left( \sup_{t \in [0,1]} \left| D_{\Theta_{j_1}} \cdots D_{\Theta_{j_n}} \delta_i(\Theta_t) \right|^p \right) < \infty, \quad i = 1, \cdots, d, \quad (3.25)$$

where $\delta_i(\Theta_t)$ is defined as

$$\delta_i(\Theta_t) := \int_0^t \langle h_i(s), dW^z_s \rangle - \int_0^t \int_0^s \frac{\text{div}(\mathbf{v}_t)(s,z)}{\kappa(z)} \tilde{N}(dz,ds). \quad (3.26)$$

**Proof.** Due to Lemma 3.1, for any $i \in \{1, \cdots, d\}$ we have $\Theta_i = (h_i, \mathbf{v}_i) \in \mathbb{H}^{\infty - \cdot} \times \mathbb{V}^{\infty - \cdot}$.

First, we prove (3.20) and (3.23). It follows from [21, Proposition 3.3] that for any $p > 1$ and $i \in \{1, \cdots, d\}$, $X^x_t$ is in $D^{1,p}_{\Theta_i}$ and $D_{\Theta_i} X^x_t$ and

$$\mathbb{E} \left( \sup_{t \leq 1} |D_{\Theta_i} X^x_t|^p \right) \leq C_p \left( \|h_i\|_{D^{1,p}_{\Theta_i}}^p + \|v_i\|_{L^p_{\Theta_i}}^p \right). \quad (3.27)$$

By (3.11) and Picard’s iteration argument it is routine to prove $K_t(x) \in D^{1,p}_{\Theta_i}$ and

$$D_{\Theta_i} K_t(x) = -\int_0^t \frac{K_s(x)}{\kappa(s)} \left( D_{\Theta_i} X^x_s \partial^2 b(X^x_s, \mu_s) + D_{\Theta_i} K_s(x) \partial_s b(X^x_s, \mu_s) \right) ds.$$

It follows from (3.12) and the condition $b(\cdot, \cdot) \in C_{b,Lip}^{m,m}(\mathbb{R}^d \times \mathcal{P}_2; \mathbb{R}^d)$ that

$$\mathbb{E} \left( \sup_{t \in [0,1]} |D_{\Theta_i} K_t(x)|^p \right) \leq C_p \int_0^1 \mathbb{E} (|D_{\Theta_i} X^x_s|^p) \, ds + C_p \int_0^1 \mathbb{E} \left( \sup_{s \leq t} |D_{\Theta_i} K_s(x)|^p \right) \, dt.$$
\[ \leq C_p \mathbb{E} \left( \sup_{s \in [0, 1]} |D_{\Theta, t} X^x_s|^p \right) + C_p \int_0^1 \mathbb{E} \left( \sup_{s \leq t} |D_{\Theta, s} K_i(x)|^p \right) dt. \]

This, together with (3.27) and Gronwall’s inequality, yields

\[ \mathbb{E} \left( \sup_{t \in [0, 1]} |D_{\Theta, t} K_i(x)|^p \right) \leq C_p \left( \|h_i\|_{L^p}^p + \|\nu_i\|_{L^p}^p \right). \]

That is, (3.20) and (3.23) hold for \( n = 1 \). For \( n > 1 \), noting (3.14) and repeating the above procedure \( n - 1 \) times, one can obtain (3.20) and (3.23).

Next, we prove (3.21). For any \( i \in \{1, \ldots, d\} \) and \( p > 1 \), by (3.8) and Picard iteration argument, it is routine to prove \( \nabla X^x_t \in D^{1,p}_\Theta \) and

\[ D_{\Theta, t} \nabla X^x_t = \int_0^t \partial_x^2 b(X^x_s, \mu_s) D_{\Theta, s} X^x_s \nabla X^x_s ds + \int_0^t \partial_x b(X^x_s, \mu_s) D_{\Theta, s} X^x_s ds \]
\[ + \int_0^t \mathbb{E} \left( \partial_x \partial_s b(X^x_s, \mu_s)(X^x_s) \nabla X^x_s \right) D_{\Theta, s} X^x_s ds \]
\[ + \sum_{j=1}^d \int_0^t \mathbb{E} \left( \partial_x A_i(\mu_s)(X^x_s) \nabla X^x_s \right) h_{i,j}(s) ds \]
\[ + \sum_{j=1}^d \int_0^t \int_{\Gamma_0} \mathbb{E} \left( \partial_x B_i(\mu_s)(X^x_s) \nabla X^x_s \right) \nu_{i,j}(s, z) N(dz, ds), \]

where \( h_{i,j} \) and \( \nu_{i,j} \) stand for the \( j \)-th element of \( h_i \) and \( \nu_i \) respectively. It follows from Burkholder-Davis-Gundy inequality that

\[ \mathbb{E} \left( \sup_{t \in [0, 1]} |D_{\Theta, t} \nabla X^x_t|^{2p} \right) \]
\[ \leq C_p \int_0^1 \mathbb{E}(|D_{\Theta, s} X^x_s|^{2p}) ds + C_p \int_0^1 \mathbb{E}|D_{\Theta, s} \nabla X^x_s|^{2p} ds \]
\[ + C_p \int_0^1 \mathbb{E}|D_{\Theta, s} X^x_s|^{2p} \mathbb{E}|\nabla X^x_s|^{2p} ds + C_p \mathbb{E} \left( \int_0^1 |h_i(s)|^2 \mathbb{E}|\nabla X^x_s|^{2p} ds \right)^{\frac{p}{2}} \]
\[ + C_p \mathbb{E} \left( \int_0^1 \int_{\Gamma_0} \mathbb{E}|\nabla X^x_s|^{2p} |\nu_i(s, z)|^p (dz) ds \right)^p \]
\[ + C_p \mathbb{E} \int_0^1 \int_{\Gamma_0} \mathbb{E}|\nabla X^x_s|^{2p} |\nu_i(s, z)|^p (dz) ds \]
\[ \leq C_p \left[ \mathbb{E} \left( \sup_{t \in [0, 1]} |D_{\Theta, t} X^x_t|^{2p} \right) \right]^{\frac{1}{2}} \left[ \mathbb{E} \left( \sup_{s \in [0, 1]} |\nabla X^x_t|^{2p} \right) \right]^{\frac{p}{2}} \]
\[ + C_p \int_0^1 \mathbb{E} \left( \sup_{s \leq t} |D_{\Theta, s} \nabla X^x_s|^{p} \right) dt \]
\[ + C_p \mathbb{E} \left( \sup_{t \in [0, 1]} |\nabla X^x_t|^{p} \right) \left( \|h_i\|_{L^p}^p + \|\nu_i\|_{L^p}^p \right). \]
Grönwall’s inequality implies
\[
E \left( \sup_{t \in [0,1]} |D_{\Theta_i} \nabla X_i^x|^p \right) \leq C_p \left( \|h_i\|_{\mathbb{L}^p}^p + \|\nu_i\|_{\mathbb{L}^p}^p + \|\omega_i\|_{\mathbb{L}^p}^p \right) < \infty.
\]
Now we have proved (3.21) for \( n = 1 \) and \( k = 1 \). For higher order derivatives, the estimates can be derived by induction.

In addition, by (3.10) and Picard’s iteration argument it is easy to prove \( J_k(x) \in D_{\Theta_i}^{1,p} \) and
\[
D_{\Theta_i} J_i(x) = \int_0^t \left( \partial^2_x b(X_s^x, \mu_s) D_{\Theta_i} X_s^x, J_s(x) + \partial_x b(X_s^x, \mu_s) D_{\Theta_i} \right) J_s(x) \right) ds.
\]
Then we have
\[
E \left( \sup_{t \in [0,1]} |D_{\Theta_i} J_i(x)|^p \right) \leq C_p \int_0^1 E \left( |D_{\Theta_i} X_s^x|^p \right) ds + C_p \int_0^1 E \left( \sup_{s \leq t} |D_{\Theta_i} J_s(x)|^p \right) dt
\]
\[
\leq C_p E \left( \sup_{s \in [0,1]} |D_{\Theta_i} X_s^x|^p \right) + C_p \int_0^1 E \left( \sup_{s \leq t} |D_{\Theta_i} J_s(x)|^p \right) dt.
\]
This, together with (3.27) and Gronwall’s inequality, yields
\[
E \left( \sup_{t \in [0,1]} |D_{\Theta_i} J_i(x)|^p \right) \leq C_p (1 + |x|^p) \left( \|h_i\|_{\mathbb{L}^p}^p + \|\nu_i\|_{\mathbb{L}^p}^p + \|\omega_i\|_{\mathbb{L}^p}^p \right).
\]
That is, (3.22) holds for \( n = 1 \). For \( n > 1 \), repeating the above procedure \( n - 1 \) times, one can obtain (3.22). By the same argument, we can get (3.23) which combines with (3.13) and (3.18) can derive (3.24).

Finally, let us prove (3.25). By (3.14) and (3.26), we have
\[
D_{\Theta_i} \delta_i(\Theta_i) = \int_0^t \langle (A(\mu_s)^* D_{\Theta_i} K_s(x)^*), \epsilon \rangle dW_s + \int_0^t \langle (A(\mu_s)^* K_s(x)^*), \epsilon \rangle dW_s
\]
\[
+ \int_0^t \int_{\Gamma_0} \langle D_{\Theta_i} K_s(x)^* B^*(\mu_s)^* - \nabla \log \kappa(z) + \nabla \zeta(z), \tilde{N}(dz, ds)
\]
\[
+ \int_0^t \int_{\Gamma_0} \langle K_s^x(x)^* B^*(\mu_s)^* - \nabla \zeta(z) - \nabla \zeta(z)^* - \nabla \zeta(z) \rangle K_s^x(x) B^*(\mu_s)^* \zeta(z), N(ds, dz)
\]
By \((\mathbb{H}_m^n)\) and (3.13), one can obtain
\[
|\zeta(z) \nabla \log \kappa(z) + \nabla \zeta(z)| \leq C|z|, \quad |\nabla \zeta(z) \nabla \log \kappa(z) + \nabla \zeta(z) \zeta(z)| \leq C|z|^4.
\]
These, together with (3.23) and Burkholder-Davis-Gundy inequality, yield
\[
\sup_{x \in \mathbb{R}^d} E \left( \sup_{t \in [0,1]} |D_{\Theta_i} \delta_i(\Theta_i)|^p \right) \leq C_p + C_p \sup_{x \in \mathbb{R}^d} E \left( \sup_{s \in [0,1]} |D_{\Theta_i} K_s(x)|^p \right) < \infty.
\]
That is, we prove (3.25) for \( n = 1 \). Repeating the above argument \( n - 1 \) times, we can obtain (3.25).

Now we are ready to give the proof of Theorem 1.1.
Lemma 3.3. Under the conditions of Theorem 1.2, for any \( f \) and \( \zeta \), provided \( \zeta > 0 \), by (3.19) and the fact \( J_t(x) \) has inverse matrix \( K_t(x) \), we only need to prove that the matrix \( \Sigma_t^\gamma \) is invertible almost surely for \( t \in (0,1) \).

For any fixed \( \omega_0 \in \Omega_0 \), suppose there is a vector \( \xi \in \mathbb{S}^{d-1} \) such that \( \xi^* \Sigma_t^\gamma (\omega_0) \xi = 0 \). Then we have

\[
\int_0^t |\xi^* K_s(x, \omega_0) A(\mu_s)|^2 ds + \int_0^t \int_{\Gamma_s} |\xi^* K_s(x, \omega_0) B(\mu_s)|^2 \zeta(z) N(dz, ds)(\omega_0) = 0.
\]

Therefore,

\[
|\xi^* K_s(x, \omega_0) A(\mu_s)|^2 = |\xi^* K_s(x, \omega_0) B(\mu_s)|^2 = 0 \text{ holds for a.e. } s \in (0,t].
\]

As for (1.3), it can be proved by the same argument used in the proof of \[23, \text{Theorem 1.1}\], so we omit the proof here. The proof is complete. \( \square \)

The following two results are main ingredients of the proof of Theorem 1.2. The first one was introduced in \[23, \text{Lemma 5.3}\].

Lemma 3.3. Let \( g_t \) be a nonnegative, bounded and predictable processes. Under \((\mathbb{H}^p_t)\), there exist constants \( \lambda_0, c_0 \geq 1 \) depending on the bound of \( g_t \) such that for all \( \epsilon \in (0,\lambda_0^{-1}) \),

\[
\mathbb{P} \left( \int_0^t \int_{\Gamma_s} g_t \zeta(z) N(dz, ds) \leq \epsilon, \int_0^t g_t^2 ds > \epsilon^2 \right) \leq \exp \{1 - c_0 \epsilon^{- \lambda_0} \},
\]

where \( \zeta \) is defined by (3.13).

Lemma 3.4. Under the conditions of Theorem 1.2, for any \( p \geq 1 \), there exist constant \( C_p > 0 \) and \( \gamma(p) > 0 \) such that for all \( t \in (0,1) \),

\[
\sup_{x \in \mathbb{R}^d} \mathbb{E}(\det \Sigma_t^\gamma)^{-p} \leq C_p t^{-\gamma(p)}.
\]

Proof. For any \( u \in \mathbb{S}^d \), recalling (3.11) we have

\[
|u K_s(x)| \geq 1 - \int_0^t |u K_s(x) \partial_x b(X^\gamma_t, \mu_s)| ds \geq 1 - \|\partial_x b\|_{\infty} \epsilon \|u\|_{\infty} \geq \frac{1}{2},
\]

provided \( t \leq \min \{1, \frac{1}{2\|\partial_x b\|_{\infty} \epsilon} \} \). By Lemma 3.3 and (1.4), we obtain

\[
\mathbb{P} (u \Sigma_t^\gamma u^* \leq \epsilon) \leq \mathbb{P} \left( u \Sigma_t^\gamma u^* \leq \epsilon, \int_0^t |u K_s(x) B(\mu_s)|^2 ds > \epsilon^2 \right)
\]

with
that for all $t$

$$\begin{align*}
+\mathbb{P}\left(\int_0^t (|uK_s(x)A(\mu_s)|^2 + |uK_s(x)B(\mu_s)|^2) \, ds \leq 2\epsilon^\frac{2}{3}\right)
\leq\mathbb{P}\left(\int_0^t |uK_s(x)B(\mu_s)|^2 \zeta(z)N(dz, ds) \leq \epsilon, \int_0^t |uK_s(x)B(\mu_s)|^2 ds > \epsilon^\frac{2}{3}\right)
+\mathbb{P}\left(\int_0^t (|uK_s(x)A(\mu_s)|^2 + |uK_s(x)B(\mu_s)|^2) \, ds \leq 2\epsilon^\frac{2}{3}\right)
\leq \exp\{1 - c_0\epsilon^{-\frac{2}{3}}\} + \mathbb{P}\left(c_2\int_0^t |uK_s(x)|^2 ds \leq \epsilon^\frac{2}{3}\right)
\leq \exp\{1 - c_0\epsilon^{-\frac{2}{3}}\} + \mathbb{P}\left(c_2t^\frac{2}{2} \leq \epsilon^\frac{2}{3}\right) = \exp\{1 - c_0\epsilon^{-\frac{2}{3}}\},
\end{align*}$$

for all $\epsilon \in (0, (\frac{c_2}{2})^\frac{2}{3})$. Then the statement follows from a standard compactness argument (cf. [17, Lemma 2.3.1]).

Now we give the proof of Theorem 1.2.

Proof. For any $k, n \in \mathbb{N}$ and $f \in C_\infty^\infty(\mathbb{R}^d)$, we have

$$\begin{align*}
(\partial_{x_{a_1}} \cdots \partial_{x_{a_k}}) \mathbb{E} \left( (\partial_{x_{\beta_1}} \cdots \partial_{x_{\beta_n}}) f(X_t^x) \right)
= \sum_{j=1}^{k} \mathbb{E} \left( (\partial_{x_{a_1}} \cdots \partial_{x_{a_j}} \partial_{x_{\beta_1}} \cdots \partial_{x_{\beta_n}}) f(X_t^x) \right)
\times G_j(\partial_{x_{a_1}} X_t^x, \cdots, \partial_{x_{a_j}} X_t^x, \partial_{x_{\beta_1}} X_t^x, \cdots, \partial_{x_{\beta_n}} X_t^x, (\partial_{x_{a_1}} \cdots \partial_{x_{a_j}} X_t^x) ),
\end{align*}$$

where $G_j$ is a real polynomial function of partial derivatives of $X_t^x$ up to order $j$. Repeating $(\frac{2n+k+1}{2})$ times the argument as used in (2.4), and by Lemma 3.2, Lemma 3.4 and H"{o}lder’s inequality, there are $p_1, p_2 > 1$ and $C > 0$ independent of $x$ such that for all $t \in (0, 1)$,

$$\begin{align*}
\left| (\partial_{x_{a_1}} \cdots \partial_{x_{a_k}}) \mathbb{E} \left( (\partial_{x_{\beta_1}} \cdots \partial_{x_{\beta_n}}) f(X_t^x) \right) \right|
\leq C \|f\|_\infty (\mathbb{E} (\det \Sigma_t^x)^{-p_1})^\frac{1}{p_2} \leq C \|f\|_\infty t^{-\gamma(p_1)/p_2}.
\end{align*}$$

The proof is finished.

The following example shows that DDSDEs driven by purely jump Lévy noises meet the conditions posed in [23, Theorem 1.1], but there are no densities for the distributions of solutions.

Example 3.1. For $\alpha \in (0, 2)$, let $\{(L_t^1, L_t^2)\}_{t \in [0, 1]}$ be a truncated 2-dimensional $\alpha$-stable Lévy process with characteristic measure $\int_{|z| \leq \alpha} |z| \, dz$, where $C$ is a positive constant. Assume $g : \mathbb{R} \to \mathbb{R}$ is a bounded and Lipschitz continuous function with positive lower bound. For any fixed constants $x_1, x_2$, consider the following SDE:

$$\begin{align*}
X_1(t) = x_1 + t + \int_0^t g(\mathbb{E} X_1(s)) dL_s^1, \quad X_2(t) = x_2 + X_1(t) \mathbb{E} X_1(t). \quad (3.28)
\end{align*}$$

Note that

$$\begin{align*}
\mathbb{E} X_1(t) = x_1 + t, \quad t \in [0, 1].
\end{align*}$$
By Itô’s formula and the representation of $X_1(t)$, the equation (3.28) can be formulated as
\[
\begin{align*}
\begin{pmatrix}
X_1(t) \\
X_2(t)
\end{pmatrix} &= \begin{pmatrix}
x_1 \\
x_2
\end{pmatrix} + \int_0^t \begin{pmatrix}
1 \\
x_1 + s
\end{pmatrix} ds + \int_0^t \begin{pmatrix}
g(x_1 + s) \\
(x_1 + s)g(x_1 + s)
\end{pmatrix} d\begin{pmatrix}
L_1^1 \\
L_2^2
\end{pmatrix}.
\end{align*}
\]
Define functions $b : \mathbb{R}^2 \times [0, 1] \rightarrow \mathbb{R}^2$ and $B : [0, t] \rightarrow \mathbb{R}^2 \otimes \mathbb{R}^2$ as
\[
b_t(y_1, y_2) = \begin{pmatrix}
1 \\
y_1 + x_1 + t
\end{pmatrix}, \quad B_t = \begin{pmatrix}
g(x_1 + t) \\
(x_1 + t)g(x_1 + t)
\end{pmatrix}, \quad t \in [0, 1], y_1, y_2 \in \mathbb{R}.
\]
By direct calculus, we have
\[
b_t(y_1, y_2) \cdot \nabla B_t - \nabla b_t(y_1, y_2) B_t = -\begin{pmatrix}
0 \\
g(x_1 + t)
\end{pmatrix}.
\]
It is clear that
\[
\text{Rank}[B_t, b_t(y_1, y_2) \cdot \nabla B_t - \nabla b_t(y_1, y_2) B_t] = 2, \quad \forall (y_1, y_2) \in \mathbb{R}^2, \quad t \in [0, 1].
\]
Therefore, Equ.(3.28) satisfies the conditions required in [23, Theorem 1.1]. However, there is no density for the distribution of $(X_1(t), X_2(t))$, since the support of $(X_1(t), X_2(t))$ lies on the line $y = x_2 + (x_1 + t)x$, $x \in \mathbb{R}$.

Acknowledgments. The author is grateful to the editor’s and referees’ valuable suggestions and comments.

REFERENCES

[1] V. Barbu and M. Röckner, Probabilistic representation for solutions to non-linear Fokker-Planck equations, *SIAM J. Math. Anal.*, 50 (2018), 4246–4260.
[2] V. Barbu and M. Röckner, From non-linear Fokker-Planck equations to solutions of distribution dependent SDE, *Ann. Probab.*, 48 (2020), 1902–1920.
[3] J. M. Bismut, Calcul des variations stochastiques et processus de sauts, *Z. Wahrsch. Verw. Gebiete*, 63 (1983), 147–225.
[4] W. Brown and K. Hepp, The Vlasov dynamics and its fluctuation in the $1/N$ limit of interacting particles, *Commun. Math. Phys.*, 66 (1977), 101–113.
[5] R. Buckdahn, J. Li, S. Peng and C. Rainer, Mean-field stochastic differential equations and associated PDEs, *Ann. Probab.*, 45 (2017), 824–878.
[6] P. Cardaliaguet and L. Mosnager, Hypoelliptic non-homogeneous diffusions, *Probab. Theory Related Fields*, 123 (2002), 453–483.
[7] P. Cardaliaguet, *Notes on Mean Field Games*, from P. L. Lions’ lectures at Collège de France, 2013.
[8] R. Carmona and F. Delarue, Forward-backward stochastic differential equations and controlled McKean-Vlasov dynamics, *Ann. Probab.*, 43 (2015), 2647–2700.
[9] D. Crisan and E. McMurray, Smoothing properties of McKean-Vlasov SDEs, *Probab. Theory Related Fields*, 171 (2018), 97–148.
[10] R. Höpfner, E. Löcherbach and M. Thieullen, Strongly degenerate time inhomogeneous SDEs: Densities and support properties. Application to Hodgkin-Huxley type systems, *Bernoulli*, 23 (2017), 2587–2616.
[11] X. Huang, M. Röckner and F. Y. Wang, Nonlinear Fokker–Planck equations for probability measures on path space and path-distribution dependent SDEs, *Discrete Contin. Dyn. Syst.*, 39 (2019), 3017–3035.
[12] X. Huang and Y. Song, Well-posedness and regularity for distribution dependent SPDEs with singular drifts, *Nonlinear Anal.*, 203 (2021), 112167.
[13] X. Huang and F. Y. Wang, Distribution dependent SDEs with singular coefficients, *Stochastic Process Appl.*, 129 (2019), 4747–4770.
[14] H. Kunita, Nondegenerate SDE’s with jumps and their hypoelliptic properties, *J. Math. Soc. Japan*, 65 (2013), 993–1035.
[15] P. Lions, *Cours au Collège de France: Théorie des jeuxchamps moyens*, available from: http://www.college-de-france.fr/default/EN/all/equi1/der/audiovideo.jsp.
[16] H. P. McKean, Propagation of chaos for a class of nonlinear parabolic equations, *Lecture Series in Differential Equations*, 7 (1967), 41–57.

[17] D. Nualart, *The Malliavin Calculus and Related Topics*, 2nd edition, Springer-Verlag, New York, 2006.

[18] E. Priola and J. Zabczyk, Densities for Ornstein-Uhlenbeck processes with jumps, *Bull. Lond. Math. Soc.*, 41 (2009), 41–50.

[19] P. Ren and F. Y. Wang, Bismut formula for Lions derivative of distribution dependent SDEs and applications, *J. Differ. Equ.*, 267 (2019), 4745–4777.

[20] M. Röckner and X. Zhang, Well-posedness of distribution dependent SDEs with singular drifts, *Bernoulli*, 27 (2021), 1131–1158.

[21] Y. Song, Gradient estimates and exponential ergodicity for mean-field SDEs with jumps, *J. Theoret. Probab.*, 33 (2020), 201–238.

[22] Y. Song and Y. Xie, Existence of density functions for the running maximum of a Lévy-Itô diffusion, *Potential Anal.*, 48 (2018), 35–48.

[23] Y. Song and X. Zhang, Regularity of density for SDEs driven by degenerate Lévy noises, *Electron. J. Probab.*, 20 (2015), 1–27.

[24] S. Taniguchi, Applications of Malliavin’s calculus to time-dependent systems of heat equations, *Osaka J. Math.*, 22 (1985), 307–320.

[25] F. Y. Wang, Distribution dependent SDEs for Landau type equations, *Stochastic Process. Appl.*, 128 (2018), 595–621.

[26] X. Zhang, Densities for SDEs driven by degenerate α-stable processes, *Ann. Probab.*, 42 (2014), 1885–1910.

Received November 2020; revised April 2021.

E-mail address: songyl@amss.ac.cn