Journal of Theoretical Probability (2020) 33:201–238
https://doi.org/10.1007/s10959-018-0845-x

Gradient Estimates and Exponential Ergodicity for Mean-Field SDEs with Jumps

Yulin Song[1]

Received: 5 March 2018 / Revised: 8 June 2018 / Published online: 18 July 2018

Abstract
In this paper, we study mean-field stochastic differential equations with jumps. By Malliavin calculus for Wiener–Poisson functionals, sharp gradient estimates are derived. Based on the gradient estimates, exponential convergence to the unique invariant measure in total variation distance is also obtained under a dissipative condition.

Keywords Malliavin calculus · Gradient estimates · Exponential ergodicity · Density functions · McKean–Vlasov equations

Mathematics Subject Classification (2010) 60H07 · 60H10

1 Introduction

Let $\{W_t\}_{t\geqslant 0}$ be a d-dimensional Brownian motion and $\{L_t\}_{t\geqslant 0}$ be a d-dimensional Lévy process on a complete probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Denote by $N(\mathrm{d}z, \mathrm{d}t)$ the jump measure of L_t. According to Lévy–Khintchine formula, the Lévy process L_t has the representation:

$$L_t = \int_0^t \int_{[0<|z|<1]} z\widehat{N}(\mathrm{d}z, \mathrm{d}s) + \int_0^t \int_{[|z|\geqslant 1]} zN(\mathrm{d}z, \mathrm{d}s),$$

where $\widehat{N}(\mathrm{d}z, \mathrm{d}s) := N(\mathrm{d}z, \mathrm{d}s) - \nu(\mathrm{d}z)\mathrm{d}s$ is the martingale measure and ν is the characteristic measure of N under $\mathbb{P}$ satisfying $\int_{\mathbb{R}^d\setminus\{0\}}(|z|^2 \wedge 1)\nu(\mathrm{d}z) < \infty$. In this paper, we further assume $\int_{[|z|\geqslant 1]} |z|^p \nu(\mathrm{d}z) < \infty$ for all $p \geqslant 1$. Let $\mathcal{P}_2$ be the collection of all probability measures with finite second moments on $\mathbb{R}^d$. Define the Wasserstein distance

✉ Yulin Song
songyl@amss.ac.cn

1 Department of Mathematics, Nanjing University, Nanjing 210093, Jiangsu Province, People's Republic of China

$$\mathbb{W}_2(\mu_1, \mu_2) := \inf_{\pi \in \mathfrak{C}(\mu_1, \mu_2)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{2}},$$

where $\mathfrak{C}(\mu_1, \mu_2)$ is the class of all couplings of μ_1 and μ_2. Then $(\mathcal{P}_2, \mathbb{W}_2)$ is a Polish space.

For measurable maps

$$b : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^d, \quad \sigma : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^d \otimes \mathbb{R}^d,$$

and

$$f : \mathcal{P}_2 \to \mathbb{R}^d \otimes \mathbb{R}^d,$$

we consider the following mean-field stochastic differential equations (SDEs):

$$\mathrm{d}X_t = b(X_t, \mathbb{P}_{X_t})\mathrm{d}t + \sigma(X_t, \mathbb{P}_{X_t})\mathrm{d}W_t + f(\mathbb{P}_{X_t})\mathrm{d}L_t, \tag{1.1}$$

where $\mathbb{P}_{X_t}$ denotes the distribution of X_t under $\mathbb{P}$.

1.1 Background and Notations

Mean-field SDEs, also known as Mckean–Vlasov equations, were first studied by Kac [15] in the framework of his study of the Boltzman equation for the particle density in diluted monatomic gases, as well as in that of the stochastic toy model for the Vlasov kinetic equation of plasma. In [19], McKean studied the propagation of chaos in physical systems of N-interacting particles related to Boltzmann's model for the statistical mechanics of rarefied gases. The limit of N-particle systems with weak interaction, formed by N equations forced by independent Brownian motions, can be described as the solution of a nonlinear deterministic evolution equation known as the McKean–Vlasov equations. These processes are *nonlinear* Markov processes. Their transition functions may not only depend on the current state of the processes, but also on the current distribution. Henceforth, many people paid their attention to the study of the equations: [2,10–12,17] and references therein for the study of chaos propagation and the limit equations; [5,13] for the regularity of the value function and associated PDEs; [26] for gradient estimates and Harnack inequality in the diffusion case; [14] for kinds of continuity and Harnack inequality for functional type of distribution dependent SDEs. For more fruitful results, we refer to [16].

For distribution-independent SDEs driven by jump processes, gradient estimates were obtained in [22,25,27,28] and references therein. For Eq. (1.1), when $f \equiv 0$ and σ does not depend on distribution, Wang [26] investigated gradient estimates by using coupling method. It seems that there is no corresponding results for the general type of equations like (1.1). Hence, in this manuscript we shall use Malliavin calculus for Wiener–Poisson functionals as a technical tool to derive gradient estimates for Eq. (1.1). The two main procedures and novelties are: (1) adopt the notion of derivatives with respect to probability measures first introduced by Lions [18] to study the Jacobian which involves the distribution term; (2) for different non-degenerate conditions, construct corresponding modified Malliavin matrixes which have close relation to the integration by part formula and help us to establish derivative formulas, then to obtain sharp estimates.

The second objective of present paper is to study the exponential ergodicity of Eq. (1.1). In diffusion case (i.e., $f \equiv 0$), the existence of the invariant measure was derived by Benachour et al. [2], Veretennikov [29] and so on; and exponential convergence in the Wasserstein metric was established by Cattiaux et al. [7], Wang [26] and references therein. When $f \equiv 0$ and $\sigma(x,\mu) \equiv I$, the exponential convergence to an invariant measure in total variation distance was investigated by Butkovsky [6] under the Veretennikov–Khasminskii condition. Based on the derivative estimates obtained in Theorem 1.2 and under a dissipative condition, we will prove the exponential convergence to the unique invariant measure in *total variation metric* of Eq. (1.1).

We will use the following notations frequently:

- Denote by $\mathscr{B}(\mathbb{R}^d)$ the σ-algebra generated by all open sets of $\mathbb{R}^d$ and by $\mathscr{B}_b(\mathbb{R}^d)$ the class of all bounded and $\mathscr{B}(\mathbb{R}^d)$-measurable functions with the norm $\|f\|_\infty := \sup_{x\in\mathbb{R}^d} |f(x)|$. $C_b^1(\mathbb{R}^d)$ is the collection of all bounded and differentiable functions with bounded and continuous derivatives. $\mathbb{S}^d$ stands for the unit sphere of $\mathbb{R}^d$. $\mathbb{R}_0^d$ denotes $\mathbb{R}^d \setminus \{0\}$.
- The Hilbert–Schmidt norm of a matrix A is denoted by $\|A\|_{HS}$, which is defined by $\|A\|_{HS} := \sqrt{\sum_{i,j} a_{ij}^2}$.
- 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

Assume there is a sub-σ-field $\mathscr{F}_0$ satisfying: $\mathscr{F}_0$ is independent of $\{W_t\}_{t\geqslant 0}$ and $\{L_t\}_{t\geqslant 0}$, and $\mathscr{F}_0$ is "rich enough" such that $\mathcal{P}_2 = \{\mathbb{P}_\xi : \xi \in L^2(\Omega, \mathscr{F}_0, \mathbb{P})\}$. Let $\{\mathscr{F}_t\}_{t\geqslant 0}$ be the filtration generated by $\{W_t\}_{t\geqslant 0}$ and $\{L_t\}_{t\geqslant 0}$, completed and augmented by $\mathscr{F}_0$; that is,

$$\mathscr{F}_t := \cap_{r>t} \sigma\{W_s, L_s : s \leqslant r\} \vee \mathscr{F}_0 \vee \mathscr{N}, t \in [0,1], \tag{1.2}$$

where $\mathscr{N}$ is the collection of all $\mathbb{P}$-null sets.

Definition 1.1 1. For any $s \geqslant 0$, a càdlàg $\mathscr{F}_t$ -adapted process $\{X_t\}_{t\geqslant s}$ on $\mathbb{R}^d$ is called a strong solution of Eq. (1.1) from time s, if

$$\int_s^t \mathbb{E}\left(|b(X_r, \mathbb{P}_{X_r})|^2 + \|\sigma(X_r, \mathbb{P}_{X_r})\|_{HS}^2 + |f(\mathbb{P}_{X_r})|^2\right) \mathrm{d}r < \infty, \quad t \geqslant s,$$

and $\mathbb{P}$-a.s.,

$$X_t = X_s + \int_s^t b(X_r, \mathbb{P}_{X_r}) \mathrm{d}r + \int_s^t \sigma(X_r, \mathbb{P}_{X_r}) \mathrm{d}W_r + \int_s^t f(\mathbb{P}_{X_r}) \mathrm{d}L_s, \ t \geqslant s.$$

2. A triple $(\tilde{X}, \tilde{W}, \tilde{L})$ is called a weak solution to Eq. (1.1) from time s, if $\tilde{W}$ is a d-dimensional Brownian motion with respect to a complete filtrated probability space $(\tilde{\Omega}, \{\tilde{\mathscr{F}}_t\}_{t\geqslant 0}, \tilde{\mathbb{P}})$, and $\tilde{L}$ is a Lévy process with characteristic measure ν under $\tilde{\mathbb{P}}$, such that $\tilde{X}_t$ solves

$$\tilde{X}_t = \tilde{X}_s + \int_s^t b(\tilde{X}_r, \tilde{\mathbb{P}}_{\tilde{X}_r})\mathrm{d}r + \int_s^t \sigma(\tilde{X}_r, \tilde{\mathbb{P}}_{\tilde{X}_r})\mathrm{d}\tilde{W}_r, + \int_s^t f(\tilde{\mathbb{P}}_{\tilde{X}_r})\mathrm{d}\tilde{L}_r,\ t \geqslant s.$$

3. Equation (1.1) is said to have weak uniqueness in $\mathcal{P}_2$, if for any $s \geqslant 0$, any two weak solution from time s with common initial distribution in $\mathcal{P}_2$ are equal in law. To be precise, if $s \geqslant 0$ and $(\tilde{X}_{s,t}, \tilde{W}_t, \tilde{L}_t)_{t\geqslant s}$ with respect to $(\tilde{\Omega}, \{\tilde{\mathscr{F}}_t\}_{t\geqslant 0}, \tilde{\mathbb{P}})$ and $(X_{s,t}, W_t, L_t)_{t\geqslant s}$ with respect to $(\Omega, \{\mathscr{F}_t\}_{t\geqslant 0}, \mathbb{P})$ are weak solutions of (1.1), then $\mathbb{P}_{X_{s,s}} = \tilde{\mathbb{P}}_{\tilde{X}_{s,s}}$ yields $\mathbb{P}_{X_{s,\cdot}} = \tilde{\mathbb{P}}_{\tilde{X}_{s,\cdot}}$.

Let B_0 be the unit open ball without origin. For the Lévy measure $\nu(\mathrm{d}z)$, we have the following assumption:

(H_ν) $\nu|_{B_0}$ is absolutely continuous with respect to the Lebesgue measure $\mathrm{d}z$; that is, there is a function $\kappa : B_0 \to (0, +\infty)$ such that

$$\nu(\mathrm{d}z)|_{B_0} = \kappa(z)\mathrm{d}z. \tag{1.3}$$

Moreover, we assume the following regularity and order conditions:

- for some $c_0 > 0$,

$$\kappa \in C^1(B_0; (0, \infty)), \quad |\nabla \log \kappa(z)| \leqslant c_0|z|^{-1}, \quad \forall z \in B_0. \tag{1.4}$$

- for some $c_1 > 0$ and $\alpha \in (0, 2)$,

$$\lim_{\epsilon\downarrow 0} \epsilon^{\alpha-2} \int_{[|z|\leqslant\epsilon]} |z|^2 \nu(\mathrm{d}z) = c_1. \tag{1.5}$$

Let us list assumptions on the coefficients.

$(H1)$ b and σ are twice differentiable with respect to the first variable x and the partial derivatives are bounded. b, σ and f, as well as their partial derivatives with respect to x, are Lipschitz continuous with respect to μ; that is, there exists a constant $C > 0$ such that

$$|b(x, \mu_1) - b(x, \mu_2)| + \|\sigma(x, \mu_1) - \sigma(x, \mu_2)\|_{HS} + |f(\mu_1) - f(\mu_2)| \leqslant C\mathbb{W}_2(\mu_1, \mu_2),$$

and

$$|\partial_x b(x, \mu_1) - \partial_x b(x, \mu_2)| + |\partial_x \sigma(x, \mu_1) - \partial_x \sigma(x, \mu_2)| \leqslant C\mathbb{W}_2(\mu_1, \mu_2),$$

for all $x \in \mathbb{R}^d$ and $\mu_1, \mu_2 \in \mathcal{P}_2$.

$(H2)$ For each $x \in \mathbb{R}^d$, each of the components of $b(x,\cdot)$, $\sigma(x,\cdot)$ and $f(\cdot)$ is in $C_b^{1,1}(\mathcal{P}_2)$ (see Definition 2.2 below) with $\sup_{x\in\mathbb{R}^d,\mu\in\mathcal{P}_2,y\in\mathbb{R}^d} |\partial_\mu b(x,\mu)(y)| < +\infty$ and $\sup_{x\in\mathbb{R}^d,\mu\in\mathcal{P}_2,y\in\mathbb{R}^d} |\partial_\mu \sigma(x,\mu)(y)| < +\infty$. Moreover, $\partial_\mu b(\cdot,\mu)(y)$ and $\partial_\mu \sigma(\cdot,\mu)(y)$ are differentiable with bounded derivatives; that is,

$$\|\partial_x\partial_\mu b\|_\infty := \sup_{x\in\mathbb{R}^d,\mu\in\mathcal{P}_2,y\in\mathbb{R}^d} |\partial_x\partial_\mu b(x,\mu)(y)| < +\infty,$$

and

$$\|\partial_x\partial_\mu \sigma\|_\infty := \sup_{x\in\mathbb{R}^d,\mu\in\mathcal{P}_2,y\in\mathbb{R}^d} |\partial_x\partial_\mu \sigma(x,\mu)(y)| < +\infty.$$

For $x \in \mathbb{R}^d$, let $\{X_t^x\}_{t\geqslant 0}$ be the solution to Eq. (1.1) with initial value x.

Theorem 1.2 *Assume (H_ν), $(H1)$ and $(H2)$. The following statements hold:*

1. *If $\|\sigma^{-1}\|_\infty := \sup_{y\in\mathbb{R}^d,\mu\in\mathcal{P}_2} |\sigma^{-1}(y,\mu)| < \infty$, then there exists $C > 0$ such that for each $t \in (0,1]$, $x, y \in \mathbb{R}^d$ and $g \in C_b^1(\mathbb{R}^d)$,*

$$|\mathbb{E}\nabla g(X_t^x)| \leqslant C\|g\|_\infty(1+|x|)t^{-\frac{1}{2}}, \tag{1.6}$$

and

$$|\nabla_y \mathbb{E} g(X_t^x)| \leqslant C\|g\|_\infty(1+|x|)|y|t^{-\frac{1}{2}}. \tag{1.7}$$

2. *If $\|f^{-1}\|_\infty := \sup_{\mu\in\mathcal{P}_2} |f^{-1}(\mu)| < \infty$, then there exists $C > 0$ such that for each $t \in (0,1]$, $x, y \in \mathbb{R}^d$ and $g \in C_b^1(\mathbb{R}^d)$*

$$|\mathbb{E}\nabla g(X_t^x)| \leqslant C\|g\|_\infty(1+|x|)t^{-\frac{1}{\alpha}}, \tag{1.8}$$

and

$$|\nabla_y \mathbb{E} g(X_t^x)| \leqslant C\|g\|_\infty(1+|x|)|y|t^{-\frac{1}{\alpha}}. \tag{1.9}$$

Remark 1.3 1. If $\|\sigma^{-1}\|_\infty < \infty$ holds, the process L can be any Lévy process which is independent of W and has finite p-th moment for all $p \geqslant 2$. In this case, the condition (H_ν) can be removed.
2. If $\|f^{-1}\|_\infty < \infty$ holds, for any $\alpha \in (0,2)$ the order $\frac{1}{\alpha}$ in the gradient estimates is sharp in short-time when L is a truncated α-stable process with characteristic measure $\frac{C_\alpha}{|z|^{d+\alpha}} I_{[0<|z|<1]}\mathrm{d}z$ for some $C_\alpha > 0$.

As an immediate result of Theorem 1.2, we have

Corollary 1.4 *Assume (H_ν), $(H1)$ and $(H2)$.*

1. *If* $\|\sigma^{-1}\|_\infty < \infty$ *holds, then there exists* $C > 0$ *such that for each* $t \in (0, 1]$ *and* $x_1, x_2 \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} |p_t(x_1, y) - p_t(x_2, y)|\mathrm{d}y \leqslant C(1 + |x_1| + |x_2|)|x_1 - x_2|t^{-\frac{1}{2}},$$

where $p_t(x_1, y)$ *and* $p_t(x_2, y)$ *denote the density functions of* $X_t^{x_1}$ *and* $X_t^{x_2}$, *respectively.*

2. *If* $\|f^{-1}\|_\infty < \infty$ *holds, then there exists* $C > 0$ *such that for each* $t \in (0, 1]$ *and* $x_1, x_2 \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} |p_t(x_1, y) - p_t(x_2, y)|\mathrm{d}y \leqslant C(1 + |x_1| + |x_2|)|x_1 - x_2|t^{-\frac{1}{\alpha}},$$

where $p_t(x_1, y)$ *and* $p_t(x_2, y)$ *denote the density functions of* $X_t^{x_1}$ *and* $X_t^{x_2}$ *respectively.*

The proofs of Theorem 1.2 and the Corollary will be shown in Sect. 3.

It is well known that under the Lipschitz condition, Eq. (1.1) has a unique strong solution (see Theorem 3.1 below). Hence, the solution is a Markov process. Precisely speaking, letting $\{X_{s,t}^\xi\}_{t\geqslant s}$ denote the solution of Eq. (1.1) from time s with $\mathscr{F}_s$ -measurable and square-integrable initial value $X_{s,s}^\xi = \xi$, the existence and uniqueness imply

$$X_{s,t}^\xi = X_{r,t}^{X_{s,r}^\xi}, \quad t \geqslant r \geqslant s \geqslant 0.$$

Due to this property, we may define a *nonlinear* semigroup $\{P_{s,t}^*\}_{t\geqslant s}$ on $\mathcal{P}_2$ by letting $P_{s,t}^*\mu = \mathbb{P}_{X_{s,t}^\xi}$ for $\mathbb{P}_\xi = \mu \in \mathcal{P}_2$. For simplicity, we will use P_t^* to denote $P_{0,t}^*$. For more detailed discussion about this kind of nonlinear semigroup, we refer to [26, p. 598].

A probability measure $\hat{\mu}$ is said to be an invariant measure of P_t^* if $P_t^*\hat{\mu} = \hat{\mu}$ for all $t \geqslant 0$. The solution is called to be exponentially ergodic if for any $\mu \in \mathcal{P}_2$, $P_t^*\mu$ converges to $\hat{\mu}$ exponentially in the sense of total variation distance. In order to investigate the exponential ergodicity of P_t^*, we give the following dissipative condition.

($H3$) There exist constants C_1 and C_2 with $C_2 > C_1 \geqslant 0$ such that for each $x_1, x_2 \in \mathbb{R}^d$ and $\mu_1, \mu_2 \in \mathcal{P}_2$,

$$2\langle b(x_1, \mu_1) - b(x_2, \mu_2), x_1 - x_2\rangle + \|\sigma(x_1, \mu_1) - \sigma(x_2, \mu_2)\|_{HS}^2 + \int_{\mathbb{R}_0^d} |z|^2\nu(\mathrm{d}z)|f(\mu_1) - f(\mu_2)|^2 \leqslant C_1\mathbb{W}_2(\mu_1, \mu_2)^2 - C_2|x_1 - x_2|^2.$$

Define the total variation distance on $\mathcal{P}_2$ as

$$\|\mu_1 - \mu_2\|_{TV} := \sup_{A\in\mathscr{B}(\mathbb{R}^d)} |\mu_1(A) - \mu_2(A)|, \quad \mu_1, \mu_2 \in \mathcal{P}_2.$$

We have the following exponentially ergodic property.

Theorem 1.5 *Let (H_ν) and $(H1)$–$(H3)$ hold. Assume $\|\sigma^{-1}\|_\infty < \infty$ or $\|f^{-1}\|_\infty < \infty$. Then there is a unique invariant measure $\hat{\mu}$ for P_t^* such that for any $\mu \in \mathcal{P}_2$,*

$$\|P_t^*\mu - \hat{\mu}(\cdot)\|_{TV} \leqslant C\left[1 + \left(\int_{\mathbb{R}^d} |x|^2 \mu(\mathrm{d}x)\right)^{\frac{1}{2}}\right] e^{-\frac{1}{2}(C_2 - C_1)t},$$

where C is a constant independent of μ and t.

The rest of this manuscript is organized as follows. In Sect. 2, we introduce some preliminaries : Lions' definition of the derivative of functions defined on $\mathcal{P}_2$ and Malliavin calculus for Wiener–Poisson functionals. In Sect. 3, we give the proofs of the main results. An example is shown in Sect. 4.

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 [18] and revised in the notes by Cardaliaguet [8].

Let $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ 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

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

Assume $\tilde{\mathscr{F}}$ 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 $\tilde{\mathbb{P}}_\xi = \mu$, i.e., μ is the distribution of ξ under $\tilde{\mathbb{P}}$.

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

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

Definition 2.1 A function $f : \mathcal{P}_2 \to \mathbb{R}$ 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 $\tilde{\mathbb{P}}_{\xi_0} = \mu_0$ such that the lifted function $\tilde{f}$ is Fréchet differentiable at ξ_0

If f is differentiable at μ_0, there exists a linear continuous mapping $D\tilde{f}(\xi_0) : L^2(\tilde{\Omega}; \mathbb{R}^d) \to \mathbb{R}$ such that

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

as $|\eta|_{L^2} \to 0$. By Riesz' representation theorem, there is a ($\mathbb{P}$-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 [8], there is a Borel function $h_0 : \mathbb{R}^d \to \mathbb{R}^d$ such that $\zeta = h_0(\xi_0)$, $\mathbb{P}-$a.s. and the function h_0 only depends on the law μ_0, not on ξ_0 itself. Taking into account the definition of $\tilde{f}$, this allows to write for any $\xi \in L^2(\tilde{\Omega}; \mathbb{R}^d)$,

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

We call $\partial_\mu f(\mu_0)(\cdot) := h_0(\cdot)$ the derivative of f at μ_0. Note that $\partial_\mu f(\mu_0)$ is only μ_0-a.e. uniquely determined, and it allows us to express $D\tilde{f}(\xi_0)$ as a function of any random variable ξ_0 with distribution μ_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 $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ and of the variables ξ_0 and ξ 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 ξ_0 is.

Since we will consider functions $f : \mathcal{P}_2 \to \mathbb{R}$ which are differentiable at all elements of $\mathcal{P}_2$, we suppose that $\tilde{f} : L^2(\tilde{\Omega}; \mathbb{R}) \to \mathbb{R}$ is Fréchet differentiable over the whole space $L^2(\tilde{\Omega}; \mathbb{R}^d)$. In this case, we have the derivative $\partial_\mu f(\tilde{\mathbb{P}}_\xi)$ defined $\tilde{\mathbb{P}}_\xi$-a.e. for all $\xi \in L^2(\tilde{\Omega}; \mathbb{R}^d)$. Due to Lemma 3.2 in [9], if the Fréchet derivative $D\tilde{f} : L^2(\tilde{\Omega}; \mathbb{R}^d) \to L(L^2(\tilde{\Omega}; \mathbb{R}^d))$ is Lipschitz continuous with a Lipschitz constant $K \in (0, +\infty)$, then there is for all $\xi \in L^2(\tilde{\Omega}; \mathbb{R}^d)$ a $\tilde{\mathbb{P}}_\xi$-version of $\partial_\mu f(\tilde{\mathbb{P}}_\xi) : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$|\partial_\mu f(\tilde{\mathbb{P}}_\xi)(y_1) - \partial_\mu f(\tilde{\mathbb{P}}_\xi)(y_2)| \leqslant K|y_1 - y_2|, \quad \forall y_1, y_2 \in \mathbb{R}^d.$$

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 all $\xi \in L^2(\tilde{\Omega}; \mathbb{R}^d)$ 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 bounded and Lipschitz continuous, that is, there is some constant $C > 0$ such that:

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

(ii) $|\partial_\mu f(\mu_1)(y_1) - \partial_\mu f(\mu_2)(y_2)| \leqslant C\,(W_2(\mu_1, \mu_2) + |y_1 - y_2|)$, for all $\mu_1, \mu_2 \in \mathcal{P}_2$ and $y_1, y_2 \in \mathbb{R}^d$. In this case, the function $\partial_\mu f$ is considered as the derivative of f and the collection of all such function f is denoted by $C_b^{1,1}(\mathcal{P}_2)$.

Remark 2.3 It is known that (cf. [5, Remark 2.1]) if f belongs to $C_b^{1,1}(\mathcal{P}_2)$, then the version of $\partial_\mu f(\mathbb{P}_\xi)(\cdot)$ indicated in Definition 2.2 is unique.

Example 2.4 Given two twice continuously differentiable functions $h : \mathbb{R}^d \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ with bounded derivatives, we consider

$$f(\tilde{\mathbb{P}}_\xi) := g(\tilde{\mathbb{E}}h(\xi)), \quad \xi \in L^2(\tilde{\Omega}, \mathbb{R}^d).$$

Then, given any $\xi_0 \in L^2(\tilde{\Omega}, \mathbb{R}^d)$,

$$\tilde{f}(\xi) := f(\tilde{\mathbb{P}}_\xi) = g(\tilde{\mathbb{E}}h(\xi))$$

is Fréchet differentiable in ξ_0, and

$$\begin{aligned}\tilde{f}(\xi_0 + \eta) - \tilde{f}(\xi_0) &= \int_0^1 g'(\tilde{\mathbb{E}}h(\xi_0 + s\eta))\tilde{\mathbb{E}}\left(h'(\xi_0 + s\eta)\eta\right) \mathrm{d}s \\ &= g'(\tilde{\mathbb{E}}h(\xi_0))\tilde{\mathbb{E}}\left(h'(\xi_0)\eta\right) + o\left(\|\eta\|_{L^2}\right).\end{aligned}$$

Thus,

$$D\tilde{f}(\xi_0)(\eta) = \tilde{\mathbb{E}}\left(g'(\tilde{\mathbb{E}}h(\xi_0))\nabla h(\xi_0)\eta\right), \quad \eta \in L^2(\tilde{\Omega}; \mathbb{R}^d);$$

that is,

$$\partial_\mu f(\tilde{\mathbb{P}}_{\xi_0})(y) = g'(\tilde{\mathbb{E}}h(\xi_0))\nabla h(y), \quad \forall y \in \mathbb{R}^d.$$

Moreover, we see that

$$\partial_y \partial_\mu f(\tilde{\mathbb{P}}_{\xi_0})(y) = g'(\tilde{\mathbb{E}}h(\xi_0))\nabla^2 h(y), \quad \forall y \in \mathbb{R}^d.$$

2.2 Malliavin Calculus

In this section, we recall some basic facts about Bismut's approach to Malliavin calculus for jump processes (cf. [3,4,24] etc.).

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

$$\Gamma_0 := \Gamma \setminus \{0\}, \quad \varrho(z) := 1 \vee \mathbf{d}(z, \Gamma_0^c)^{-1}, \tag{2.2}$$

where $\mathbf{d}(z, \Gamma_0^c)$ is the distance of z to the complement of Γ_0. Let Ω be the canonical space of all points $\omega = (w, \mu)$, where

- $w : [0, 1] \to \mathbb{R}^d$ is a continuous function with $w(0) = 0$;
- μ is an integer-valued measure on $[0, 1] \times \Gamma_0$ with $\mu(A) < +\infty$ for any compact set $A \subset [0, 1] \times \Gamma_0$.

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

$$W_t(\omega) := w(t), \quad N(\omega; \mathrm{d}t, \mathrm{d}z) := \mu(\omega; \mathrm{d}t, \mathrm{d}z) := \mu(\mathrm{d}t, \mathrm{d}z).$$

Let $(\mathscr{F}_t)_{t\in[0,1]}$ be the smallest right-continuous filtration on Ω such that W and N are optional. In the following, we write $\mathscr{F} := \mathscr{F}_1$, and endow $(\Omega, \mathscr{F})$ 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 $\nu(\mathrm{d}z)\mathrm{d}t$, where $\nu(\mathrm{d}z) = \kappa(z)\mathrm{d}z$ with

$$\kappa \in C^1(\Gamma_0; (0,\infty)), \ \int_{\Gamma_0} (1 \wedge |z|^2)\kappa(z)\mathrm{d}z < +\infty, \ \ |\nabla \log \kappa(z)| \leqslant C\varrho(z), \tag{2.3}$$

  where $\varrho(z)$ is defined by (2.2);
- W and N are independent.

In the following, we write

$$\widehat{N}(\mathrm{d}z, \mathrm{d}s) := N(\mathrm{d}z, \mathrm{d}s) - \nu(\mathrm{d}z)\mathrm{d}s.$$

2.3 Function Spaces

Let $p \geqslant 1$ and k be an integer. We introduce the following spaces for later use.

- $L^p(\Omega)$: The space of all $\mathscr{F}$-measurable random variables with finite norm:

$$\|F\|_p := \left[\mathbb{E}|F|^p\right]^{\frac{1}{p}}.$$

- $\mathbb{L}^1_p$: The space of all predictable processes: $\xi : \Omega \times [0,1] \times \Gamma_0 \to \mathbb{R}^k$ with finite norm:

$$\|\xi\|_{\mathbb{L}^1_p} := \left[\mathbb{E}\left(\int_0^1\int_{\Gamma_0} |\xi(s,z)|\nu(\mathrm{d}z)\mathrm{d}s\right)^p\right]^{\frac{1}{p}} + \left[\mathbb{E}\int_0^1\int_{\Gamma_0} |\xi(s,z)|^p\nu(\mathrm{d}z)\mathrm{d}s\right]^{\frac{1}{p}} < \infty. \tag{2.4}$$

- $\mathbb{L}^2_p$: The space of all predictable processes: $\xi : \Omega \times [0,1] \times \Gamma_0 \to \mathbb{R}^k$ with finite norm:

$$\|\xi\|_{\mathbb{L}^2_p} := \left[\mathbb{E}\left(\int_0^1\int_{\Gamma_0} |\xi(s,z)|^2\nu(\mathrm{d}z)\mathrm{d}s\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} + \left[\mathbb{E}\int_0^1\int_{\Gamma_0} |\xi(s,z)|^p\nu(\mathrm{d}z)\mathrm{d}s\right]^{\frac{1}{p}} < \infty.$$

- $\mathbb{H}_p$: The space of all measurable adapted processes $h : \Omega \times [0,1] \to \mathbb{R}^d$ with finite norm:

$$\|h\|_{\mathbb{H}_p} := \left[\mathbb{E}\left(\int_0^1 |h(s)|^2\mathrm{d}s\right)^{\frac{p}{2}}\right]^{\frac{1}{p}} < +\infty. \tag{2.5}$$

- $\mathbb{V}_p$: The space of all predictable processes $\mathbf{v} : \Omega \times [0, 1] \times \Gamma_0 \to \mathbb{R}^d$ with finite norm:

$$\|\mathbf{v}\|_{\mathbb{V}_p} := \|\nabla_z \mathbf{v}\|_{\mathbb{L}^1_p} + \|\mathbf{v}\varrho\|_{\mathbb{L}^1_p} < \infty, \tag{2.6}$$

where $\varrho(z)$ is defined by (2.2). Below we shall write

$$\mathbb{H}_{\infty-} := \cap_{p\geqslant 1}\mathbb{H}_p, \quad \mathbb{V}_{\infty-} := \cap_{p\geqslant 1}\mathbb{V}_p.$$

- $\mathbb{H}_0$: The space of all bounded measurable adapted processes $h : \Omega \times [0, 1] \to \mathbb{R}^d$.
- $\mathbb{V}_0$: The space of all predictable processes $\mathbf{v} : \Omega \times [0, 1] \times \Gamma_0 \to \mathbb{R}^d$ with the following properties: (i) $\mathbf{v}$ and $\nabla_z \mathbf{v}$ are bounded; (ii) there exists a compact subset $U \subset \Gamma_0$ such that

$$\mathbf{v}(t, z) = 0, \quad \forall z \notin U.$$

- For any $t \in (0, 1]$, $\mathbb{L}^1_p(t)$, $\mathbb{H}_p(t)$ and $\mathbb{V}_p(t)$ are the corresponding spaces as defined in (2.4), (2.5) and (2.6) when the integral interval $[0, 1]$ is changed into $[0, t]$.

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

$$F = f(W(h_1), \dots, W(h_{m_1}), N(g_1), \dots, N(g_{m_2})),$$

where $f \in C^\infty_{\mathrm{p}}(\mathbb{R}^{m_1+m_2})$, $h_1, \dots, h_{m_1} \in \mathbb{H}_0$ and $g_1, \dots, g_{m_2} \in \mathbb{V}_0$ are non-random and real-valued, and

$$W(h_i) := \int_0^1 \langle h_i(s), \mathrm{d}W_s\rangle_{\mathbb{R}^d}, \quad N(g_j) := \int_0^1\int_{\Gamma_0} g_j(s, z)N(\mathrm{d}s, \mathrm{d}z).$$

For any $p > 1$ and $\Theta = (h, \mathbf{v}) \in \mathbb{H}_p \times \mathbb{V}_p$, let us define

$$\begin{aligned} D_\Theta F := &\sum_{i=1}^{m_1} (\partial_i f)(\cdot) \int_0^1 \langle h(s), h_i(s)\rangle_{\mathbb{R}^d} \mathrm{d}s \\ &+ \sum_{j=1}^{m_2} (\partial_{j+m_1} f)(\cdot) \int_0^1\int_{\Gamma_0} \langle \mathbf{v}(s, z), \nabla_z g_j(s, z)\rangle_{\mathbb{R}^d} N(\mathrm{d}s, \mathrm{d}z), \end{aligned} \tag{2.7}$$

where "$(\cdot)$" stands for $W(h_1), \dots, W(h_{m_1}), N(g_1), \dots, N(g_{m_2})$.

Definition 2.5 For $p > 1$ and $\Theta = (h, \mathbf{v}) \in \mathbb{H}_p \times \mathbb{V}_p$, we define the first order Sobolev space $\mathbb{D}^{1,p}_\Theta$ being the completion of $\mathcal{F}C^\infty_{\mathrm{p}}$ 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. [24, Theorem 2.9]).

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

$$\mathbb{E}D_{\Theta}F = \mathbb{E}(F\delta(\Theta)), \tag{2.8}$$

where

$$\delta(\Theta) := \int_0^1 \langle h(s), \mathrm{d}W_s\rangle - \int_0^1\int_{\Gamma_0} \frac{\operatorname{div}(\kappa\mathbf{v})(s,z)}{\kappa(z)}\widehat{N}(\mathrm{d}z, \mathrm{d}s),$$

and $\operatorname{div}(\kappa\mathbf{v}) := \sum_{i=1}^d \partial_{z_i}(\kappa\mathbf{v}_i)$ *stands for the divergence.*

The following Burkholder–Davis–Gundy inequality (c.f. [20, Theorem 48] and [24, Lemma 2.3]) will be used frequently.

Lemma 2.7 1. *For any* $p \geqslant 1$*, there is a constant* $C_p > 0$ *such that for any càdlàg martingale* M_t*,*

$$\mathbb{E}\left(\sup_{s\leqslant t}|M_s|^p\right) \leqslant C_p\mathbb{E}[M, M]_t^{\frac{p}{2}}. \tag{2.9}$$

2. *For any* $p \geqslant 1$*, there is a constant* $C_p > 0$ *such that for any* $\zeta \in \mathbb{L}^1_p$*,*

$$\mathbb{E}\left(\sup_{t\in[0,1]}\left|\int_0^t\int_{B_0}\zeta(s,z)N(\mathrm{d}z,\mathrm{d}s)\right|^p\right) \leqslant C_p\|\zeta\|^p_{\mathbb{L}^1_p}. \tag{2.10}$$

3 Proofs of Main Results

Let's first show the existence and uniqueness of the solution to Eq. (1.1).

Theorem 3.1 *Assume that there is a constant* $C > 0$ *such that*

$$\begin{aligned}&|b(x_1,\mu_1) - b(x_2,\mu_2)|^2 + \|\sigma(x_1,\mu_1) - \sigma(x_2,\mu_2)\|^2_{HS} + |f(\mu_1) - f(\mu_2)|^2\\ &\quad\leqslant C\left(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1,\mu_2)^2\right), \quad x_1, x_2 \in \mathbb{R}^d,\ \mu_1,\mu_2 \in \mathcal{P}_2.\end{aligned} \tag{3.1}$$

Then Eq. (1.1) *admits a unique strong/weak solution. Moreover, for any* $s \geqslant 0$*,* $T \geqslant s$ *and* $p \geqslant 2$*,* $\mathbb{E}|X_{s,s}|^p < \infty$ *implies*

$$\mathbb{E}\left(\sup_{t\in[s,T]}|X_{s,t}|^p\right) \leqslant C_{s,T}\left(1 + \mathbb{E}|X_{s,s}|^p\right), \tag{3.2}$$

where $C_{s,T}$ *is a constant depending on* s *and* T*.*

Proof For the existence and uniqueness of the strong solution, we refer to [13, Theorem 3.1]. And for (3.2), it can be easily derived by Lemma 2.7 and Gronwall's inequality, so we omit the proof. We only prove the uniqueness of the weak solution. Let (X_t, W_t, L_t) and $(\tilde{X}_t, \tilde{W}_t, \tilde{L}_t)$ with respect to $(\Omega, \mathscr{F}_t, \mathbb{P})$ and $(\tilde{\Omega}, \tilde{\mathscr{F}}_t, \tilde{\mathbb{P}})$ respectively be two weak solutions with $\mathbb{P}_{X_0} = \tilde{\mathbb{P}}_{\tilde{X}_0}$. Then X_t solves Eq. (1.1) while $\tilde{X}_t$ solves

$$\mathrm{d}\tilde{X}_t = b(\tilde{X}_t, \tilde{\mathbb{P}}_{\tilde{X}_t})\mathrm{d}t + \sigma(\tilde{X}_t, \tilde{\mathbb{P}}_{\tilde{X}_t})\mathrm{d}\tilde{W}_t + f(\tilde{\mathbb{P}}_{\tilde{X}_t})\mathrm{d}\tilde{L}_t. \tag{3.3}$$

To prove $\mathbb{P}_{X_\cdot} = \tilde{\mathbb{P}}_{\tilde{X}_\cdot}$, let

$$\overline{b}_t(x) = b(x, \mathbb{P}_{X_t}), \ \ \overline{\sigma}_t(x) = \sigma(x, \mathbb{P}_{X_t}), \ \ \overline{f}_t = f(\mathbb{P}_{X_t}).$$

Due to (3.1) and (3.2), it is easy to verify that $\overline{b}$ and $\overline{\sigma}$ are Lipschitz continuous and $\overline{f}$ is bounded on [0, 1]. Therefore, the SDE

$$\mathrm{d}\overline{X}_t = \overline{b}_t(\overline{X}_t)\mathrm{d}t + \overline{\sigma}(\overline{X}_t)\mathrm{d}\tilde{W}_t + \overline{f}_t\mathrm{d}\tilde{L}_t, \ \ \overline{X}_0 = \tilde{X}_0 \tag{3.4}$$

has a unique strong solution. Due to Yamada–Watanabe theorem for nonhomogeneous SDEs with jumps (see [1]), it also has the uniqueness of the weak solution. Noting that

$$\mathrm{d}X_t = \overline{b}_t(X_t)\mathrm{d}t + \overline{\sigma}(X_t)\mathrm{d}W_t + \overline{f}_t\mathrm{d}L_t, \ \ \mathbb{P}_{X_0} = \tilde{\mathbb{P}}_{\tilde{X}_0},$$

we have $\tilde{\mathbb{P}}_{\overline{X}_\cdot} = \mathbb{P}_{X_\cdot}$. Therefore, (3.4) can be written as

$$\mathrm{d}\overline{X}_t = b\left(\overline{X}_t, \tilde{\mathbb{P}}_{\overline{X}_t}\right)\mathrm{d}t + \sigma\left(\overline{X}_t, \tilde{\mathbb{P}}_{\overline{X}_t}\right)\mathrm{d}\tilde{W}_t + f\left(\tilde{\mathbb{P}}_{\overline{X}_t}\right)\mathrm{d}\tilde{L}_t.$$

By the uniqueness of (3.3), we obtain $\overline{X} = \tilde{X}$. Therefore, $\tilde{\mathbb{P}}_{\tilde{X}_\cdot} = \mathbb{P}_{X_\cdot}$. □

For any $x \in \mathbb{R}^d$, denote by X_t^x the solution to Eq. (1.1) with initial value x. Assume $(H1)$ holds. Let $\{J_t\}_{t\in[0,1]}$ satisfy the following linear matrix-valued equation:

$$J_t = I + \int_0^t \partial_x b\big(X_s^x, \mathbb{P}_{X_s^x}\big) J_s \mathrm{d}s + \sum_{k=1}^d \int_0^t \partial_x \sigma_k \left(X_s^x, \mathbb{P}_{X_s^x}\right) J_s \mathrm{d}W_s^k, \ \ t \in [0, 1], \tag{3.5}$$

where σ_k denotes the k-th column of σ and W^k is the k-th element of W. Then by Itô's formula, we can easily obtain that the inverse matrix of J_t denoted by K_t satisfies:

$$K_t = I - \int_0^t K_s \partial_x b\big(X_s^x, \mathbb{P}_{X_s^x}\big)\mathrm{d}s + \sum_{k=1}^d \int_0^t K_s \left(\partial_x \sigma_k\big(X_s^x, \mathbb{P}_{X_s^x}\big)\right)^2 \mathrm{d}s$$
$$- \sum_{k=1}^d \int_0^t K_s \partial_x \sigma_k\big(X_s^x, \mathbb{P}_{X_s^x}\big)\mathrm{d}W_s^k \tag{3.6}$$

for all $t \in [0, 1]$.

Lemma 3.2 *Assume $(H1)$ holds. For any $p \geqslant 2$, we have*

$$\mathbb{E}\left(\sup_{t\in[0,1]} |J_t|^p\right) < \infty, \qquad \mathbb{E}\left(\sup_{t\in[0,1]} |K_t|^p\right) < \infty. \tag{3.7}$$

These can be easily derived by (2.9) and Gronwall's inequality, so we omit the proof.

3.1 Malliavin Derivatives and Their Estimates

Proposition 3.3 *Assume (H_ν) and $(H1)$. For any $p \geqslant 2$, $\Theta := (h, \mathbf{v}) \in \mathbb{H}_{\infty-} \times \mathbb{V}_{\infty-}$ and $t \in [0, 1]$, X_t^x is in $\mathbb{D}_\Theta^{1,p}$ and*

$$D_\Theta X_t^x = \int_0^t \partial_x b\big(X_s^x, \mathbb{P}_{X_s^x}\big) D_\Theta X_s^x \mathrm{d}s + \int_0^t \partial_x \sigma\big(X_s^x, \mathbb{P}_{X_s^x}\big) D_\Theta X_s^x \mathrm{d}W_s$$
$$+ \int_0^t \sigma\big(X_s^x, \mathbb{P}_{X_s^x}\big) h(s)\mathrm{d}s + \int_0^t \int_{B_0} f\big(\mathbb{P}_{X_s^x}\big)\mathbf{v}(s, z) N(\mathrm{d}z, \mathrm{d}s). \tag{3.8}$$

Moreover, there exists $C_p > 0$ such that

$$\mathbb{E}\left(\sup_{s\leqslant t} |D_\Theta X_s^x|^p\right) \leqslant C_p(1 + |x|^p)\left(\|h\|^p_{\mathbb{H}_{2p}(t)} + \|\mathbf{v}\|^p_{\mathbb{L}^1_p(t)}\right), \quad \forall t \in [0, 1]. \tag{3.9}$$

Proof Define the following Picard iteration: for all $t \in [0, 1]$, $X_t^{x,0} \equiv x$ and

$$X_t^{x,n+1} = x + \int_0^t b\big(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\big)\mathrm{d}s + \int_0^t \sigma\big(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\big)\mathrm{d}W_s$$
$$+ \int_0^t f\big(\mathbb{P}_{X_s^{x,n}}\big)\mathrm{d}L_s, \quad n \geqslant 0.$$

Then from the proof of Theorem 3.1 in [13] we have for any $p \geqslant 2$,

$$\lim_{n\to\infty} \mathbb{E}\left(\sup_{t\in[0,1]} |X_t^{x,n} - X_t^x|^p\right) = 0. \tag{3.10}$$

Now let's prove the following statement: for any $n \geqslant 1$,

$$X_t^{x,n} \in \mathbb{D}_\Theta^{1,p}, \quad \forall t \in [0,1] \text{ and } \mathbb{E}\left(\sup_{t\in[0,1]} |D_\Theta X_t^{x,n}|^p\right) < \infty, \quad \forall p \geqslant 2. \tag{3.11}$$

Due to (2.7), (2.9) and (2.10), it is clear that (3.11) holds for $n = 1$. Suppose that (3.11) holds for some n. Then, by($H1$) and the chain rule [23, Lemma 2.4], we have $\sigma\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) \in \mathbb{D}_\Theta^{1,p}$ and

$$D_\Theta \sigma\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) = \partial_x \sigma\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) D_\Theta X_s^{x,n}. \tag{3.12}$$

Also by (2.7), we have

$$D_\Theta \int_0^t f\left(\mathbb{P}_{X_s^{x,n}}\right) \mathrm{d}L_s = \int_0^t \int_{B_0} f\left(\mathbb{P}_{X_s^{x,n}}\right) \mathbf{v}(s,z) N(\mathrm{d}z, \mathrm{d}s).$$

Using the chain rule and Lemma 2.3 in [23], one can show that $\int_0^t b(X_s^n, \mathbb{P}_{X_s^{x,n}})\mathrm{d}s \in \mathbb{D}_\Theta^{1,p}$ and

$$D_\Theta \int_0^t b\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) \mathrm{d}s = \int_0^t \partial_x b\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) D_\Theta X_s^{x,n} \mathrm{d}s.$$

Therefore, $X_t^{x,n+1} \in \mathbb{D}_\Theta^{1,p}$ and

$$\begin{aligned} D_\Theta X_t^{x,n+1} &= \int_0^t \partial_x b\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) D_\Theta X_s^{x,n} \mathrm{d}s + \int_0^t \partial_x \sigma\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) D_\Theta X_s^{x,n} \mathrm{d}W_s \\ &\quad + \int_0^t \sigma\left(X_s^{x,n}, \mathbb{P}_{X_s^{x,n}}\right) h(s) \mathrm{d}s + \int_0^t \int_{\Gamma_0} f\left(\mathbb{P}_{X_s^{x,n}}\right) \mathbf{v}(s,z) N(\mathrm{d}z, \mathrm{d}s). \end{aligned} \tag{3.13}$$

By (2.9) and ($B1$), we can easily have $\mathbb{E}\left(\sup_{t\in[0,1]} |D_\Theta X_t^{x,n+1}|^p\right) < \infty$. So we have proved (3.11).

Due to ($H1$), (3.2) and the condition $(h, \mathbf{v}) \in \mathbb{H}_{\infty-} \times \mathbb{V}_{\infty-}$, the linear Eq. (3.8) has a unique solution denoted by $\{Y_t\}_{t\in[0,1]}$. For any $p \geqslant 2$, by (2.9) and (2.10) one can arrive at

$$\begin{aligned} \mathbb{E}\left(\sup_{s\leqslant t} |Y_s|^p\right) &\leqslant C_p \int_0^t \mathbb{E}|\partial_x b\left(X_s^x, \mathbb{P}_{X_s^x}\right) Y_s|^p \mathrm{d}s \\ &\quad + C_p \mathbb{E}\left(\int_0^t |\partial_x \sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right) Y_s|^2 \mathrm{d}s\right)^{\frac{p}{2}} \\ &\quad + C_p \mathbb{E}\left(\int_0^t |\sigma(X_s^x, \mathbb{P}_{X_s^x}) h(s)| \mathrm{d}s\right)^p \end{aligned}$$

$$
\begin{aligned}
&+ C_p\mathbb{E}\left(\int_0^t\int_{B_0}|f(\mathbb{P}_{X_s^x})\mathbf{v}(s,z)|N(\mathrm{d}z,\mathrm{d}s)\right)^p\\
&\leqslant C_p\int_0^t \mathbb{E}\left(\sup_{s\leqslant r}|Y_s|^p\right)\mathrm{d}r + C_p\left[\mathbb{E}\left(\sup_{t\in[0,1]}|X_t^x|^{2p}+1\right)\right]^{\frac{1}{2}}\\
&\left(\|h\|^p_{\mathbb{H}_{2p}(t)}+\|\mathbf{v}\|^p_{\mathbb{L}^1_p(t)}\right).
\end{aligned}
$$

Gronwall's inequality, together with (3.2), implies

$$
\mathbb{E}\left(\sup_{s\leqslant t}|Y_s|^p\right)\leqslant C_p\left(1+|x|^p\right)\left(\|h\|^p_{\mathbb{H}_{2p}(t)}+\|\mathbf{v}\|^p_{\mathbb{L}^1_p(t)}\right).
$$

It follows from (3.8) and (3.13) that

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{t\in[0,1]}|D_\Theta X_s^{x,n+1}-Y_t|^p\right)\\
&\quad\leqslant C_p\int_0^1\mathbb{E}|\partial_x b\big(X_s^{x,n},\mathbb{P}_{X_s^{x,n}}\big)D_\Theta X_s^{x,n}-\partial_x b\big(X_s^x,\mathbb{P}_{X_s^x}\big)Y_s|^p\mathrm{d}s\\
&\qquad+C_p\int_0^1\mathbb{E}|\partial_x\sigma\big(X_s^{x,n},\mathbb{P}_{X_s^{x,n}}\big)D_\Theta X_s^{x,n}-\partial_x\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)Y_s|^p\mathrm{d}s\\
&\qquad+C_p\mathbb{E}\left(\int_0^1|\sigma\big(X_s^{x,n},\mathbb{P}_{X_s^{x,n}}\big)-\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)||h(s)|\mathrm{d}s\right)^p\\
&\qquad+C_p\mathbb{E}\left(\int_0^1\int_{\Gamma_0}|f\big(\mathbb{P}_{X_s^{x,n}}\big)-f(\mathbb{P}_{X_s^x})||\mathbf{v}(s,z)|N(\mathrm{d}z,\mathrm{d}s)\right)^p\\
&\quad\leqslant C_p\int_0^1\mathbb{E}|\partial_x b\big(X_s^{x,n},\mathbb{P}_{X_s^{x,n}}\big)-\partial_x b\big(X_s^x,\mathbb{P}_{X_s^x}\big)|^p|Y_s|^p\mathrm{d}s\\
&\qquad+C_p\int_0^1\mathbb{E}|\partial_x\sigma\big(X_s^{x,n},\mathbb{P}_{X_s^{x,n}}\big)-\partial_x\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)|^p|Y_s|^p\mathrm{d}s\\
&\qquad+C_p\int_0^1\mathbb{E}|D_\Theta X_s^{x,n}-Y_s|^p\mathrm{d}s\\
&\qquad+C_p\mathbb{E}\left[\int_0^1|X_s^{x,n}-X_s^x|^2+\mathbb{W}_2(\mathbb{P}_{X_s^{x,n}},\mathbb{P}_{X_s^x})^2\mathrm{d}s\int_0^1|h(s)|^2\mathrm{d}s\right]^{\frac{p}{2}}\\
&\qquad+C_p\sup_{s\in[0,1]}\mathbb{W}_2(\mathbb{P}_{X_s^{x,n}},\mathbb{P}_{X_s^x})^p\|\mathbf{v}\|^p_{\mathbb{L}^1_p}\\
&\quad\leqslant C_p\int_0^1\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta X_s^{x,n}-Y_s|^p\right)\mathrm{d}t\\
&\qquad+C_p\left(\mathbb{E}\sup_{t\in[0,1]}|X_s^{x,n}-X_s^x|^{2p}\right)^{\frac{1}{2}}\left(\left(\mathbb{E}\sup_{t\in[0,1]}|Y_t|^{2p}\right)^{\frac{1}{2}}+\|h\|^p_{\mathbb{H}_{2p}}+\|\mathbf{v}\|^p_{\mathbb{L}^1_p}\right).
\end{aligned}
$$

Gronwall's inequality implies

$$\limsup_{n\to\infty}\mathbb{E}\left(\sup_{t\in[0,1]}|D_\Theta X_t^{x,n+1}-Y_t|^p\right)\leqslant C_p\limsup_{n\to\infty}\left(\mathbb{E}\sup_{t\in[0,1]}|X_t^{x,n}-X_t^x|^{2p}\right)^{\frac{1}{2}}=0.$$

Combining this with (3.10) and the fact $\mathbb{W}_2(\mathbb{P}_{X_s^{x,n}},\mathbb{P}_{X_s^x})^p\leqslant\mathbb{E}|X_s^{x,n}-X_s^x|^p$, and letting $n\to\infty$ in (3.13), we obtain $X_t^x\in\mathbb{D}_\Theta^{1,p}$ and $D_\Theta X_t^x$ satisfies Eq. (3.8). □

Lemma 3.4 *Assume (H_v) and $(H1)$. For any $\Theta:(h,\mathbf{v})\in\mathbb{H}_{\infty-}\times\mathbb{V}_{\infty-}$, we have $K_t\in\mathbb{D}_\Theta^{1,2}$. Moreover, there exists a constant $C>0$ such that*

$$\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta K_s|^2\right)\leqslant C\left(\|h\|^2_{\mathbb{H}_4(t)}+\|h\|^2_{\mathbb{H}_8(t)}+\|\mathbf{v}\|^2_{\mathbb{L}^1_4(t)}\right),\quad\forall t\in[0,1].\tag{3.14}$$

Proof Define the following Picard's iteration: for each $t\in[0,1]$, $K_t^{(0)}=I$ and for $n\geqslant 0$,

$$\begin{aligned}K_t^{(n+1)}=I&-\int_0^t K_s^{(n)}\partial_x b\big(X_s^x,\mathbb{P}_{X_s^x}\big)\mathrm{d}s+\sum_{k=1}^d\int_0^t K_s^{(n)}\left(\partial_x\sigma_k(X_s^x,\mathbb{P}_{X_s^x})\right)^2\mathrm{d}s\\&-\sum_{k=1}^d\int_0^t K_s^{(n)}\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)\mathrm{d}W_s^k,\quad t\in[0,1].\end{aligned}$$

Then for any $p\geqslant 2$, it is routine to prove that

$$\lim_{n\to\infty}\mathbb{E}\left(\sup_{t\in[0,1]}|K_t^{(n)}-K_t|^p\right)=0.\tag{3.15}$$

By induction, Proposition 1.3.2 and 1.2.4 in [21], and Proposition 3.3 we have $K_t^{(n+1)}$ is Malliavin differentiable along Θ. Moreover,

$$\begin{aligned}D_\Theta K_t^{(n+1)}=&-\int_0^t D_\Theta K_s^{(n)}\partial_x b(X_s^x,\mathbb{P}_{X_s^x})\mathrm{d}s-\int_0^t K_s^{(n)}\partial_x^2 b\big(X_s^x,\mathbb{P}_{X_s^x}\big)D_\Theta X_s^x\mathrm{d}s\\&+\sum_{k=1}^d\int_0^t D_\Theta K_s^{(n)}\left(\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)\right)^2\mathrm{d}s\\&+2\sum_{k=1}^d\int_0^t K_s^{(n)}\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)\partial_x^2\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)D_\Theta X_s^x\mathrm{d}s\end{aligned}$$

$$-\sum_{k=1}^{d}\int_0^t D_\Theta K_s^{(n)}\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)\mathrm{d}W_s^k-\sum_{k=1}^{d}\int_0^t K_s^{(n)}\partial_x^2\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)D_\Theta X_s^x\mathrm{d}W_s^k$$
$$-\sum_{k=1}^{d}\int_0^t K_s^{(n)}\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)h_k(s)\mathrm{d}s,$$

where h_k denotes the k-th component of h. Let $\{Y_t\}_{t\in[0,1]}$ solve the following linear equation:

$$\begin{aligned}Y_t=&-\int_0^t Y_s\partial_x b\big(X_s^x,\mathbb{P}_{X_s^x}\big)\mathrm{d}s-\int_0^t K_s\partial_x^2 b\big(X_s^x,\mathbb{P}_{X_s^x}\big)D_\Theta X_s^x\mathrm{d}s\\&+\sum_{k=1}^{d}\int_0^t Y_s\left(\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)\right)^2\mathrm{d}s+2\sum_{k=1}^{d}\int_0^t K_s\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)\partial_x^2\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)D_\Theta X_s^x\mathrm{d}s\\&-\sum_{k=1}^{d}\int_0^t Y_s\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)\mathrm{d}W_s^k-\sum_{k=1}^{d}\int_0^t K_s\partial_x^2\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)D_\Theta X_s^x\mathrm{d}W_s^k\\&-\sum_{k=1}^{d}\int_0^t K_s\partial_x\sigma_k\big(X_s^x,\mathbb{P}_{X_s^x}\big)h_k(s)\mathrm{d}s.\end{aligned}\tag{3.16}$$

Then by Hölder's inequality and (2.9), we can arrive at

$$\begin{aligned}&\mathbb{E}\left(\sup_{t\in[0,1]}|D_\Theta K_s^{(n+1)}-Y_s|^2\right)\\&\leqslant C\int_0^1\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta K_s^{(n)}-Y_s|^2\right)\mathrm{d}s+C\int_0^1\mathbb{E}|K_s^{(n)}-K_s|^2|D_\Theta X_s^x|^2\mathrm{d}s+C\mathbb{E}\left(\int_0^1|K_s^{(n)}-K_s||h(s)|\mathrm{d}s\right)^2\\&\leqslant C\int_0^1\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta K_s^{(n)}-Y_s|^2\right)\mathrm{d}s+C\left[\mathbb{E}\left(\sup_{t\in[0,1]}|K_t^{(n)}-K_t|^4\right)\right]^{\frac12}\left[\left(\mathbb{E}\sup_{t\in[0,1]}|D_\Theta X_t^x|^4\right)^{\frac12}+\|h\|_{\mathbb{H}_4}^2\right].\end{aligned}$$

Gronwall's inequality, together with (3.9), yields

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}\left(\sup_{t\in[0,1]}|D_\Theta K_s^{(n+1)}-Y_s|^2\right)\leqslant\lim_{n\to\infty}C\left[\mathbb{E}\left(\sup_{t\in[0,1]}|K_t^{(n)}-K_t|^4\right)\right]^{\frac12}\\\left[\left(\mathbb{E}\sup_{t\in[0,1]}|D_\Theta X_t^x|^4\right)^{\frac12}+\|h\|_{\mathbb{H}_4}^2\right]=0.\end{aligned}$$

Combining this with (3.15), we obtain $K_t\in\mathbb{D}_\Theta^{1,2}$ and $D_\Theta K_t=Y_t$ for all $t\in[0,1]$ a.s.. Moreover, by (3.9), (3.7) and (3.16) we have

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta K_s|^2\right)\\
&\quad\leqslant C\int_0^t \mathbb{E}\left(\sup_{s\leqslant r}|D_\Theta K_s|^2\right)\mathrm{d}r + C\int_0^t \mathbb{E}|K_s|^2|D_\Theta X_s^x|^2\mathrm{d}s + C\mathbb{E}\left(\int_0^t |K_s||h(s)|\mathrm{d}s\right)^2\\
&\quad\leqslant C\int_0^t \mathbb{E}\left(\sup_{s\leqslant r}|D_\Theta K_s|^2\right)\mathrm{d}r + C\left[\mathbb{E}\left(\sup_{s\in[0,1]}|K_s|^4\right)\right]^{\frac{1}{2}}\left\{\left[\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta X_s^x|^4\right)\right]^{\frac{1}{2}}\right.\\
&\qquad\left.+\|h\|^2_{\mathbb{H}_4(t)}\right\}\\
&\quad\leqslant C\int_0^t \mathbb{E}\left(\sup_{s\leqslant r}|D_\Theta K_s|^2\right)\mathrm{d}r + C(1+|x|^2)\left(\|h\|^2_{\mathbb{H}_8(t)}+\|h\|^2_{\mathbb{H}_4(t)}+\|\mathbf{v}\|^2_{\mathbb{L}^1_4(t)}\right).
\end{aligned}
$$

Hence,

$$
\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta K_s|^2\right)\leqslant C(1+|x|^2)\left(\|h\|^2_{\mathbb{H}_8(t)}+\|h\|^2_{\mathbb{H}_4(t)}+\|\mathbf{v}\|^2_{\mathbb{L}^1_4(t)}\right),
$$

where C is a constant independent of t. □

3.2 Directional Derivative with Respect to Initial Value

Recall that for given $x, y \in \mathbb{R}^d$, the directional derivative of X_t^x along the direction y is defined as

$$
\nabla_y X_t^x := L^2 - \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(X_t^{x+\epsilon y}-X_t^x\right),\quad \forall t\in[0,1].
$$

Denote by $(\tilde{W},\tilde{L})$ a copy of (W, L) on some complete probability space $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{\mathbb{P}})$, and by $\{\tilde{X}_t^x\}_{t\geqslant 0}$ the copy of the solution to the SDE (1.1), but driven by Brownian motion $\tilde{W}$ and Lévy process $\tilde{L}$. Obviously, $(\tilde{W},\tilde{L},\tilde{X}^x)$ is an independent copy of (W, L, X^x), defined over $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{\mathbb{P}})$. And for all $t\in[0,1]$, $\nabla_y\tilde{X}_t^x$ is the directional derivative of $\tilde{X}_t^x$ along the direction y.

Proposition 3.5 *Assume (H1) and (H2). Then for any $t\in[0,1]$ and $x, y\in\mathbb{R}^d$, we have*

$$
\begin{aligned}
\nabla_y X_t^x &= y + \int_0^t \partial_x b\big(X_s^x,\mathbb{P}_{X_s^x}\big)\nabla_y X_s^x\mathrm{d}s + \int_0^t \tilde{\mathbb{E}}\left(\partial_\mu b\big(X_s^x,\mathbb{P}_{X_s^x}\big)(\tilde{X}_s^x)\nabla_y\tilde{X}_s^x\right)\mathrm{d}s\\
&\quad+\int_0^t \partial_x\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)\nabla_y X_s^x\mathrm{d}W_s + \int_0^t \tilde{\mathbb{E}}\left(\partial_\mu\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)(\tilde{X}_s^x)\nabla_y\tilde{X}_s^x\right)\mathrm{d}W_s\\
&\quad+\int_0^t \tilde{\mathbb{E}}\left(\partial_\mu f\big(\mathbb{P}_{X_s^x}\big)(\tilde{X}_s^x)\nabla_y\tilde{X}_s^x\right)\mathrm{d}L_s. \qquad (3.17)
\end{aligned}
$$

Moreover, for any $p \geqslant 2$ there exists $C_p > 0$ such that

$$\mathbb{E}\left(\sup_{t\in[0,1]} |\nabla_y X_t^x|^p\right) \leqslant C_p|y|^p, \tag{3.18}$$

where C_p is a constant independent of x, y and t.

Proof For the sake of convenience, we assume $b \equiv 0$. For $\epsilon > 0$, let

$$X_t^{x+\epsilon y} = x + \epsilon y + \int_0^t \sigma\left(X_s^{x+\epsilon x}, \mathbb{P}_{X_s^{x+\epsilon y}}\right) \mathrm{d}W_s + \int_0^t f\left(\mathbb{P}_{X_s^{x+\epsilon y}}\right) \mathrm{d}L_s, \quad t \in [0,1].$$

Then for any $p \geqslant 2$, by the Lipschitz continuity of σ and f and Gronwall's inequality it is easy to prove

$$\mathbb{E}\left(\sup_{t\in[0,1]} |X_t^{x+\epsilon y} - X_t^x|^p\right) \leqslant C_p|y|^p\epsilon^p. \tag{3.19}$$

Observe that

$$\begin{aligned}
&\sigma\left(X_s^{x+\epsilon y}, \mathbb{P}_{X_s^{x+\epsilon y}}\right) - \sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right)\\
&\quad = \int_0^1 \partial_\lambda \left(\sigma\left(X_s^x + \lambda(X_s^{x+\epsilon y} - X_s^x), \mathbb{P}_{X_s^{x+\epsilon y}}\right)\right) \mathrm{d}\lambda\\
&\qquad + \int_0^1 \partial_\lambda \left(\sigma\left(X_s^x, \mathbb{P}_{X_s^x + \lambda(X_s^{x+\epsilon y} - X_s^x)}\right)\right) \mathrm{d}\lambda\\
&\quad = \alpha_s^\epsilon\left(X_s^{x+\epsilon y} - X_s^x\right) + \tilde{\mathbb{E}}\left(\beta_s^\epsilon\left(\tilde{X}_s^{x+\epsilon y} - \tilde{X}_s^x\right)\right),
\end{aligned}$$

where

$$\alpha_s^\epsilon := \int_0^1 \partial_x\sigma\left(X_s^x + \lambda\left(X_s^{x+\epsilon y} - X_s^x\right), \mathbb{P}_{X_s^{x+\epsilon y}}\right)\mathrm{d}\lambda$$

and

$$\beta_s^\epsilon := \int_0^1 \partial_\mu\sigma\left(X_s^x, \mathbb{P}_{X_s^x + \lambda\left(X_s^{x+\epsilon y} - X_s^x\right)}\right)\left(\tilde{X}_s^x + \lambda\left(\tilde{X}_s^{x+\epsilon y} - \tilde{X}_s^x\right)\right)\mathrm{d}\lambda.$$

Moreover, for any $p \geqslant 2$ by the Lipschitz continuity of $\partial_x\sigma$ and $\partial_\mu\sigma$ we have

$$\begin{aligned}
&\mathbb{E}\left(\sup_{s\in[0,1]} |\alpha_s^\epsilon - \partial_x\sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right)|^p\right)\\
&\quad \leqslant \mathbb{E}\left(\sup_{s\in[0,1]} \int_0^1 |\partial_x\sigma\left(X_s^x + \lambda\left(X_s^{x+\epsilon y} - X_s^x\right), \mathbb{P}_{X_s^{x+\epsilon y}}\right) - \partial_x\sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right)|^p \mathrm{d}\lambda\right)
\end{aligned}$$

$$\leqslant C_p\mathbb{E}\left(\sup_{s\in[0,1]}|X_s^{x+\epsilon y}-X_s^x|^p+\sup_{s\in[0,1]}\mathbb{W}_2^p\big(\mathbb{P}_{X_s^{x+\epsilon y}},\mathbb{P}_{X_s^x}\big)\right)$$
$$\leqslant C_p\mathbb{E}\left(\sup_{s\in[0,1]}|X_s^{x+\epsilon y}-X_s^x|^p\right)\leqslant C_p|y|^p\epsilon^p, \tag{3.20}$$

and

$$\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\beta_s^\epsilon-\partial_\mu\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)\big(\tilde{X}_s^x\big)|^p\right)$$
$$\leqslant C_p\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x|^p+\sup_{s\in[0,1]}\sup_{\lambda\in[0,1]}\mathbb{W}_2\big(\mathbb{P}_{X_s^x+\lambda\left(X_s^{x+\epsilon y}-X_s^x\right)},\mathbb{P}_{X_s^x}\big)^p\right)$$
$$\leqslant C_p\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x|^p\right)\leqslant C_p|y|^p\epsilon^p. \tag{3.21}$$

By the similar argument as above, we have

$$f\big(\mathbb{P}_{X_s^{x+\epsilon y}}\big)-f\big(\mathbb{P}_{X_s^x}\big)=\tilde{\mathbb{E}}\left(\gamma_s^\epsilon\big(\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x\big)\right)$$

for some process γ^ϵ with

$$\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\gamma_s^\epsilon-\partial_\mu f\big(\mathbb{P}_{X_s^x}\big)(\tilde{X}_s^x)|^p\right)\leqslant C_p|y|^p\epsilon^p. \tag{3.22}$$

Consider the following equation:

$$Y_t^x(y)=y+\int_0^t\partial_x\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)Y_s^x(y)\mathrm{d}W_s+\int_0^t\tilde{\mathbb{E}}\left[\partial_\mu\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)\big(\tilde{X}_s^x\big)\tilde{Y}_s^x(y)\right]\mathrm{d}W_s$$
$$+\int_0^t\tilde{\mathbb{E}}\left[\partial_\mu\sigma\big(\mathbb{P}_{X_s^x}\big)\big(\tilde{X}_s^x\big)\tilde{Y}_s^x(y)\right]\mathrm{d}L_s. \tag{3.23}$$

By classical Picard's iteration, it is not difficult to prove that there is a unique solution and an independent copy $\tilde{Y}$ of Y, defined on $(\tilde{\Omega},\tilde{\mathscr{F}},\tilde{\mathbb{P}})$. Then

$$X_t^{x+\epsilon y}-X_t^x-\epsilon Y_t^x(y)=\int_0^t\partial_x\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)\big(X_s^{x+\epsilon y}-X_s^x-\epsilon Y_s^x(y)\big)\mathrm{d}W_s$$
$$+\int_0^t\tilde{\mathbb{E}}\left(\partial_\mu\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)\big(\tilde{X}_s^x\big)\big(\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x-\epsilon\tilde{Y}_s^x(y)\big)\right)\mathrm{d}W_s$$
$$+\int_0^t\big(\alpha_s^\epsilon-\partial_x\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)\big)\big(X_s^{x+\epsilon y}-X_s^x\big)\mathrm{d}W_s$$
$$+\int_0^t\tilde{\mathbb{E}}\left[\big(\beta_s^\epsilon-\partial_\mu\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)\big(\tilde{X}_s^x\big)\big)\big(\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x\big)\right]\mathrm{d}W_s$$

$$
\begin{aligned}
&+\int_0^t \tilde{\mathbb{E}}\left(\partial_\mu f\left(\mathbb{P}_{X_s^x}\right)\left(\tilde{X}_s^x\right)\left(\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x-\epsilon \tilde{Y}_s^x(y)\right)\right) \mathrm{d} L_s \\
&+\int_0^t \tilde{\mathbb{E}}\left[\left(\gamma_s^\epsilon-\partial_\mu f\left(\mathbb{P}_{X_s^x}\right)\left(\tilde{X}_s^x\right)\right)\left(\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x\right)\right] \mathrm{d} L_s .
\end{aligned}
$$

Hence, it follows from Lemma 2.7 that

$$
\begin{aligned}
&\mathbb{E}\left(\sup_{t\in[0,1]} |X_t^{x+\epsilon y}-X_t^x-\epsilon Y_t^x(y)|^2\right) \\
&\quad\leqslant C\int_0^1 \mathbb{E}\left(\sup_{s\leqslant t}|X_s^{x+\epsilon y}-X_s^x-\epsilon Y_s^x(y)|^2\right)\mathrm{d}t \\
&\qquad + C\int_0^1 \tilde{\mathbb{E}}\left(\sup_{s\leqslant t}|\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x-\epsilon \tilde{Y}_s^x(y)|^2\right)\mathrm{d}t \\
&\qquad + C\int_0^1 \mathbb{E}\left(|X_s^{x+\epsilon y}-X_s^x|^2|\alpha_s^\epsilon-\partial_x\sigma\left(X_s^x,\mathbb{P}_{X_s^x}\right)|^2\right)\mathrm{d}s \\
&\qquad + C\mathbb{E}\int_0^1 \tilde{\mathbb{E}}\left(|\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x|^2|\beta_s^\epsilon-\partial_\mu\sigma\left(X_s^x,\mathbb{P}_{X_s^x}\right)\left(\tilde{X}_s^x\right)|^2\right)\mathrm{d}s \\
&\qquad + C\mathbb{E}\int_0^1\int_{\mathbb{R}_0^d} \tilde{\mathbb{E}}\left(|\tilde{X}_s^{x+\epsilon y}-\tilde{X}_s^x|^2|\gamma_s^\epsilon-\partial_\mu f\left(\mathbb{P}_{X_s^x}\right)\left(\tilde{X}_s^x\right)|^2\right)|z|^2\nu(\mathrm{d}z)\mathrm{d}s \\
&\quad\leqslant C\int_0^1 \mathbb{E}\left(\sup_{s\leqslant t}|X_s^{x+\epsilon y}-X_s^x-\epsilon Y_s^x(y)|^2\right)\mathrm{d}t \\
&\qquad + C\left[\mathbb{E}\left(\sup_{s\in[0,1]}|X_s^{x+\epsilon y}-X_s^x|^4\right)\right]^{\frac12}\left[\mathbb{E}\left(\sup_{s\in[0,1]}|\alpha_s^\epsilon-\partial_x\sigma\left(X_s^x,\mathbb{P}_{X_s^x}\right)|^4\right)\right]^{\frac12} \\
&\qquad + C\left[\mathbb{E}\left(\sup_{s\in[0,1]}|X_s^{x+\epsilon y}-X_s^x|^4\right)\right]^{\frac12}\left[\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\beta_s^\epsilon-\partial_\mu\sigma\left(X_s^x,\mathbb{P}_{X_s^x}\right)\left(\tilde{X}_s^x\right)|^4\right)\right]^{\frac12} \\
&\qquad + C\left[\mathbb{E}\left(\sup_{s\in[0,1]}|X_s^{x+\epsilon y}-X_s^x|^4\right)\right]^{\frac12}\left[\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\gamma_s^\epsilon-\partial_\mu f\left(\mathbb{P}_{X_s^x}\right)\left(\tilde{X}_s^x\right)|^4\right)\right]^{\frac12}.
\end{aligned}
$$

Gronwall's inequality, together with (3.19), (3.20), (3.21) and (3.22), yields

$$
\mathbb{E}\left(\sup_{t\in[0,1]}|X_t^{x+\epsilon y}-X_t^x-\epsilon Y_t^x(y)|^2\right)\leqslant C|y|^4\epsilon^4.
$$

Thus,

$$
\lim_{\epsilon\to 0}\mathbb{E}\left(\sup_{t\in[0,1]}\left|\epsilon^{-1}\left(X_t^{x+\epsilon y}-X_t^x\right)-Y_t^x(y)\right|^2\right)=0.
$$

For (3.18), it is due to Lemma 2.7 and Grondwall's inequality. □

Lemma 3.6 *Assume (H_ν), $(H1)$ and $(H2)$. For any $x, y \in \mathbb{R}^d$ and $\Theta = (h, \mathbf{v}) \in \mathbb{H}_{\infty-} \times \mathbb{V}_{\infty-}$, we have $\nabla_y X_t^x \in \mathbb{D}_\Theta^{1,2}$ and*

$$\mathbb{E}\left(\sup_{s\leqslant t} |D_\Theta \nabla_y X_s^x|^2\right) \leqslant C(1+|x|^2)|y|^2 \Big(\|h\|^2_{\mathbb{H}_4(t)} + \|h\|^2_{\mathbb{H}_8(t)} + \|\mathbf{v}\|^2_{\mathbb{L}_4(t)} + \|\mathbf{v}\|^2_{\mathbb{V}_2(t)} + \|\mathbf{v}\|^2_{\mathbb{V}_4(t)}\Big), \tag{3.24}$$

where C is a constant independent of x, y and t.

Proof By the similar argument as discussed in the proof of Lemma 3.4, using Picard's iteration we can prove that $\nabla_y X_t^x$ is Malliavin differentiable. And by (3.17), we have

$$\begin{aligned} D_\Theta \nabla_y X_t^x &= \int_0^t \partial_x^2 b\big(X_s^x, \mathbb{P}_{X_s^x}\big) D_\Theta X_s^x \nabla_y X_s^x \mathrm{d}s + \int_0^t \partial_x b\big(X_s^x, \mathbb{P}_{X_s^x}\big) D_\Theta \nabla_y X_s^x \mathrm{d}s \\ &+ \int_0^t \tilde{\mathbb{E}}\left(\partial_x \partial_\mu b\big(X_s^x, \mathbb{P}_{X_s^x}\big)\big(\tilde{X}_s^x\big) D_\Theta X_s^x \nabla_y \tilde{X}_s^x\right) \mathrm{d}s \\ &+ \int_0^t \partial_x^2 \sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right) D_\Theta X_s^x \nabla_y X_s^x \mathrm{d}W_s \\ &+ \int_0^t \partial_x \sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right) D_\Theta \nabla_y X_s^x \mathrm{d}W_s \\ &+ \int_0^t \partial_x \sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right) \nabla_y X_s^x h(s) \mathrm{d}s \\ &+ \int_0^t \tilde{\mathbb{E}}\left(\partial_\mu \sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right)\big(\tilde{X}_s^x\big) \nabla_y \tilde{X}_s^x\right) h(s) \mathrm{d}s \\ &+ \int_0^t \tilde{\mathbb{E}}\left(\partial_x \partial_\mu \sigma\left(X_s^x, \mathbb{P}_{X_s^x}\right)\big(\tilde{X}_s^x\big) D_\Theta X_s^x \nabla_y \tilde{X}_s^x\right) \mathrm{d}W_s \\ &+ \int_0^t \int_{B_0} \tilde{\mathbb{E}}\left(\partial_\mu f\left(\mathbb{P}_{X_s^x}\right)\big(\tilde{X}_s^x\big) \nabla_y \tilde{X}_s^x\right) \mathbf{v}(s,z) N(\mathrm{d}z, \mathrm{d}s). \end{aligned}$$

Then by Lemma 2.7 and Hölder's inequality one can obtain

$$\begin{aligned} \mathbb{E}\left(\sup_{s\leqslant t} |D_\Theta \nabla_y X_s^x|^2\right) &\leqslant C \int_0^t \mathbb{E}|D_\Theta X_s^x|^2 |\nabla_y X_s^x|^2 \mathrm{d}s + C \int_0^t \mathbb{E}|D_\Theta \nabla_y X_s^x|^2 \mathrm{d}s \\ &+ C \int_0^t \mathbb{E}|D_\Theta X_s^x|^2 \tilde{\mathbb{E}}|\nabla_y \tilde{X}_s^x|^2 \mathrm{d}s + C\mathbb{E}\left(\int_0^t |\nabla_y X_s^x||h(s)|\mathrm{d}s\right)^2 \\ &+ C\mathbb{E}\left(\int_0^t \tilde{\mathbb{E}}|\nabla_y \tilde{X}_s^x||h(s)|\mathrm{d}s\right)^2 + C\mathbb{E}\left(\int_0^t\int_{B_0} \tilde{\mathbb{E}}|\nabla_y \tilde{X}_s^x||\mathbf{v}(s,z)|\nu(\mathrm{d}z)\mathrm{d}s\right)^2 \\ &+ C\mathbb{E}\left(\int_0^t\int_{B_0} \tilde{\mathbb{E}}|\nabla_y \tilde{X}_s^x|^2 |\mathbf{v}(s,z)|^2 \nu(\mathrm{d}z)\mathrm{d}s\right) \end{aligned}$$

$$\leqslant C\int_0^t \mathbb{E}\left(\sup_{s\leqslant r}|D_\Theta\nabla_y X_s^x|^2\right)\mathrm{d}r + C\left[\mathbb{E}\left(\sup_{t\in[0,1]}|\nabla_y X_t^x|^4\right)\right]^{\frac{1}{2}}\|h\|^2_{\mathbb{H}_4(t)}$$
$$+ C\left[\mathbb{E}\left(\sup_{t\in[0,1]}|\nabla_y X_t^x|^4\right)\right]^{\frac{1}{2}}\left[\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta X_s^x|^4\right)\right]^{\frac{1}{2}}$$
$$+ C\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\nabla_y\tilde{X}_s^x|^2\right)\|\mathbf{v}\|^2_{\mathbb{V}_2(t)} + C\left[\tilde{\mathbb{E}}\left(\sup_{s\in[0,1]}|\nabla_y\tilde{X}_s^x|^4\right)\right]^{\frac{1}{2}}\|\mathbf{v}\|^2_{\mathbb{V}_4(t)}.$$

Gronwall's inequality, together with (3.9) and (3.18), gives

$$\mathbb{E}\left(\sup_{s\leqslant t}|D_\Theta\nabla_y X_s^x|^2\right)\leqslant C(1+|x|^2)|y|^2\left(\|h\|^2_{\mathbb{H}_4(t)}+\|h\|^2_{\mathbb{H}_8(t)}+\|\mathbf{v}\|^2_{\mathbb{L}^1_4(t)}\right.$$
$$\left.+\|\mathbf{v}\|^2_{\mathbb{V}_2(t)}+\|\mathbf{v}\|^2_{\mathbb{V}_4(t)}\right),$$

where C is a constant independent of x, y and t. □

3.3 Prooof of Theorem 1.2

The following lemma, which was introduced in [24, Lemma 5.2] and [27, Lemma 2.5,2.6], is very useful to derive the gradient estimates.

Lemma 3.7 *Under (1.5), we have the following statements:*

1. *for any $p\geqslant 2$, there exist constants $\epsilon_0, C_0, C_1>0$ such that for all $\epsilon\in(0,\epsilon_0)$,*

$$C_0\epsilon^{p-\alpha}\leqslant\int_{[|z|\leqslant\epsilon]}|z|^p\nu(\mathrm{d}z)\leqslant C_1\epsilon^{p-\alpha}. \tag{3.25}$$

2. *for any $p\geqslant 2$, there exists a constant $C_p>0$ such that for each $t,\epsilon\in(0,1)$,*

$$\mathbb{E}\left(\int_0^t\int_{[0<|z|\leqslant\epsilon]}|z|^3N(\mathrm{d}z,\mathrm{d}s)\right)^{-p}\leqslant C_p\left((t\epsilon^{3-\alpha})^{-p}+t^{-\frac{3p}{\alpha}}\right). \tag{3.26}$$

For any $\Theta=(h,\mathbf{v})\in\mathbb{H}_{\infty-}\times\mathbb{V}_{\infty-}$, by (3.5), (3.6), (3.8) and applying Itô's formula to $K_tD_\Theta X_t^x$, one can easily have

$$D_\Theta X_t^x = J_t\left(\int_0^t K_s\sigma\big(X_s^x,\mathbb{P}_{X_s^x}\big)h(s)\mathrm{d}s+\int_0^t\int_{\Gamma_0}K_sf\big(\mathbb{P}_{X_s^x}\big)\mathbf{v}(s,z)N(\mathrm{d}z,\mathrm{d}s)\right),$$
$$\forall t\in[0,1]. \tag{3.27}$$

For each fixed $t \in (0, 1)$, let $\zeta_t(z)$ be a smooth, nonnegative and real-valued function such that

$$\zeta_t(z) = |z|^3, \text{ if } |z| \leqslant \frac{1}{4} t^{\frac{1}{\alpha}} \quad \text{and} \quad \zeta_t(z) = 0, \text{ if } |z| \geqslant \frac{1}{2} t^{\frac{1}{\alpha}}$$

with $|\nabla_z \zeta_t(z)| \leqslant C|z|^2$ and $|\zeta_t(z)| \leqslant C|z|^3$, where C is a constant independent of t.

In what follows, we choose some specific $(h, \mathbf{v}) \in \mathbb{H}_{\infty-} \times \mathbb{V}_{\infty-}$ in the following two cases.

1. If $\|\sigma^{-1}\|_\infty := \sup_{x \in \mathbb{R}^d, \mu \in \mathcal{P}_2} |\sigma^{-1}(x, \mu)| < \infty$, for any fixed $t \in (0, 1]$ and $1 \leqslant j \leqslant d$, we set

$$h_{t,j}(s) = \frac{1}{t} \sigma^{-1}\big(X_s^x, \mathbb{P}_{X_s^x}\big)(J_s)_{\cdot j}, \quad \forall s \in [0, t] \quad \text{and} \quad \mathbf{v} \equiv 0, \tag{3.28}$$

where $(J_s)_{\cdot j}$ stands for the j-th column of J_s.

2. If $\|f^{-1}\|_\infty := \sup_{\mu \in \mathcal{P}_2} |f^{-1}(\mu)| < \infty$, for any fixed $t \in (0, 1]$ and $1 \leqslant j \leqslant d$, set

$$h \equiv 0 \quad \text{and} \quad \mathbf{v}_{t,j}(s, z) = f^{-1}\big(\mathbb{P}_{X_s^x}\big)(J_s)_{\cdot j} \zeta(z), \quad \forall s \in [0, t], z \in \Gamma_0. \tag{3.29}$$

Define

$$\delta_t(\mathbf{v}_{t,j}) := \int_0^t \int_{B_0} \frac{\operatorname{div}(\kappa(z) \mathbf{v}_{t,j}(s, z))}{\kappa(z)} \widehat{N}(\mathrm{d}z, \mathrm{d}s),$$

and

$$G_{t,j} := \int_0^t \int_{B_0} \langle \nabla_z \zeta_t(z), \mathbf{v}_{t,j}(s, z) \rangle N(\mathrm{d}z, \mathrm{d}s).$$

We have the following estimates.

Lemma 3.8 1. *Assume $\|\sigma^{-1}\|_\infty < \infty$. For any $p \geqslant 2$, we have*

$$\|h_{t,j}\|_{\mathbb{H}_p(t)} \leqslant C_p t^{-\frac{1}{2}}, \quad 1 \leqslant j \leqslant d, \tag{3.30}$$

where C_P is a constant independent of t.

2. *Assume $\|f^{-1}\|_\infty < \infty$. For any $p \geqslant 2$, we have*

$$\|\mathbf{v}_{t,j}\|_{\mathbb{L}^1_p(t)} \leqslant C_p t^{\frac{3}{\alpha}}, \quad \|\mathbf{v}_{t,j}\|_{\mathbb{V}_p(t)} \leqslant C_p t^{\frac{2}{\alpha}}, \quad 1 \leqslant j \leqslant d, \tag{3.31}$$

and

$$\mathbb{E}|\delta_t(\mathbf{v}_{t,j})|^p \leqslant C_p t^{\frac{2p}{\alpha}}, \quad \mathbb{E}|G_{t,j}|^p \leqslant C_p t^{\frac{5p}{\alpha}}, \quad 1 \leqslant j \leqslant d. \tag{3.32}$$

where C_p is a constant independent of t.

Proof 1. By (3.7), we have

$$\|h_{t,j}\|_{\mathbb{H}_p(t)} = \left[\mathbb{E}\left(\int_0^t |\frac{1}{t}\sigma^{-1}\big(X_s^x, \mathbb{P}_{X_s^x}\big)(J_s)_j|^2 \mathrm{d}s\right)^{\frac{p}{2}}\right]^{\frac{1}{p}}$$
$$\leqslant \frac{1}{t}\|\sigma^{-1}\|_\infty \left[\mathbb{E}\left(\sup_{t\in[0,1]} |J_s|^p\right)\right]^{\frac{1}{p}} t^{\frac{1}{2}} \leqslant C_p t^{-\frac{1}{2}},$$

where C_p is a constant independent of t.

2. For any $p \geqslant 2$ and $j = 1, \ldots, d$, by (2.10) and (3.25) we can obtain

$$\|\mathbf{v}_{t,j}\|^p_{\mathbb{L}^1_p(t)} \leqslant C_p \mathbb{E}\int_0^t\int_{B_0} |\mathbf{v}_{t,j}(s,z)|^p \nu(\mathrm{d}z)\mathrm{d}s + C_p\mathbb{E}\left(\int_0^t\int_{B_0} |\mathbf{v}_{t,j}(s,z)|\nu(\mathrm{d}z)\mathrm{d}s\right)^p$$
$$\leqslant C_p \mathbb{E}\int_0^t\int_{B_0} |J_s|^p|\zeta_t(z)|^p \nu(\mathrm{d}z)\mathrm{d}s + C_p\mathbb{E}\left(\int_0^t\int_{B_0} |J_s||\zeta_t(z)|\nu(\mathrm{d}z)\mathrm{d}s\right)^p$$
$$\leqslant C_p \mathbb{E}\int_0^t\int_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]} |J_s|^p|z|^{3p} \nu(\mathrm{d}z)\mathrm{d}s$$
$$+ C_p\mathbb{E}\left(\int_0^t\int_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]} |J_s||z|^3\nu(\mathrm{d}z)\mathrm{d}s\right)^p$$
$$\leqslant C_p t^{\frac{3p-\alpha}{\alpha}} t + C_p t^p \left(t^{\frac{3-\alpha}{\alpha}}\right)^p \leqslant C_p t^{\frac{3p}{\alpha}}.$$

Observe that

$$|\nabla_z \mathbf{v}_{t,j}(s,z)| = |f^{-1}(\mu_s)(J_{s-})_{\cdot j}\nabla_z\zeta_t(z)| \leqslant C|J_s||z|^2 I_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]}.$$

Then we have

$$\|\mathbf{v}_{t,j}\|^p_{\mathbb{V}_p(t)} \leqslant C_p\|\nabla_z\mathbf{v}_{t,j}\|^p_{\mathbb{L}^1_p(t)} + C_p\|\varrho\mathbf{v}_{t,j}\|^p_{\mathbb{L}^1_p(t)}$$
$$\leqslant C_p \mathbb{E}\int_0^t\int_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]} |J_s|^p|z|^{2p} \nu(\mathrm{d}z)\mathrm{d}s$$
$$+ C_p\mathbb{E}\left(\int_0^t\int_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]} |J_s||z|^2\nu(\mathrm{d}z)\mathrm{d}s\right)^p$$
$$\leqslant C_p t^{\frac{2p-\alpha}{\alpha}} t + C_p (t^{\frac{2-\alpha}{\alpha}} t)^p \leqslant C_p t^{\frac{2p}{\alpha}}.$$

Since

$$\left|\frac{\mathrm{div}(\kappa(z)\mathbf{v}_{t,j}(s,z))}{\kappa(z)}\right| = \Big|\langle \nabla\log\kappa(z), f^{-1}\big(\mathbb{P}_{X_s^x}\big)(J_s)_{\cdot j}\zeta_t(z)\rangle$$
$$+\langle f^{-1}\big(\mathbb{P}_{X_s^x}\big)(J_s)_{\cdot j}, \nabla_z\zeta_t(z)\rangle\Big|$$
$$\leqslant \frac{C}{|z|}|J_s||z|^3 I_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]} + C|J_s||z|^2 I_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]} \leqslant C|J_s||z|^2 I_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]},$$

then by (2.9) and (3.25) we have for any $p \geqslant 2$,

$$
\begin{aligned}
\mathbb{E}|\delta_t(\mathbf{v}_{t,j})|^p &= \mathbb{E}\left|\int_0^t\int_{B_0} \frac{\mathrm{div}(\kappa(z)\mathbf{v}_{t,j}(s,z))}{\kappa(z)}\widehat{N}(\mathrm{d}z,\mathrm{d}s)\right|^p \\
&\leqslant C_p\mathbb{E}\left(\int_0^t\int_{B_0}\left|\frac{\mathrm{div}(\kappa(z)\mathbf{v}_{t,j}(s,z))}{\kappa(z)}\right|^2\nu(\mathrm{d}z)\mathrm{d}s\right)^{\frac{p}{2}} \\
&\leqslant C_p\left(\int_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]}|z|^4\nu(\mathrm{d}z)\right)^{\frac{p}{2}}\mathbb{E}\left(\sup_{s\in[0,1]}|J_s|^p\right)t^{\frac{p}{2}} \\
&\leqslant C_p t^{\frac{p(4-\alpha)}{2\alpha}}t^{\frac{p}{2}} = C_p t^{\frac{2p}{\alpha}}.
\end{aligned}
$$

It follows from (2.10), (3.7) and (3.25) that

$$
\begin{aligned}
\mathbb{E}|G_{t,j}|^p &= \mathbb{E}\left|\int_0^t\int_{B_0}\langle\nabla_z\zeta_t(z),\mathbf{v}_{t,j}(s,z)\rangle N(\mathrm{d}z,\mathrm{d}s)\right|^p \\
&\leqslant C_p\mathbb{E}\int_0^t\int_{B_0}\left|\langle\nabla_z\zeta_t(z),\mathbf{v}_{t,j}(s,z)\rangle\right|^p\nu(\mathrm{d}z)\mathrm{d}s \\
&\quad + C_p\mathbb{E}\left(\int_0^t\int_{B_0}\left|\langle\nabla_z\zeta_t(z),\mathbf{v}_{t,j}(s,z)\rangle\right|\nu(\mathrm{d}z)\mathrm{d}s\right)^p \\
&\leqslant C_p\int_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]}|z|^{5p}\nu(\mathrm{d}z)\mathbb{E}\left(\sup_{s\in[0,1]}|J_s|^p\right) \\
&\quad + C_p\left(\int_{[0<|z|\leqslant t^{\frac{1}{\alpha}}]}|z|^5\nu(\mathrm{d}z)\right)^p\mathbb{E}\left(\sup_{s\in[0,1]}|J_s|^p\right) \\
&\leqslant C_p t^{\frac{5p-\alpha}{\alpha}}t + C_p t^{\frac{(5-\alpha)p}{\alpha}}t^p \leqslant C_p t^{\frac{5p}{\alpha}}.
\end{aligned}
$$

□

Now we are ready to give the proof of Theorem 1.2.

Proof For any $\Theta = (h,\mathbf{v}) \in \mathbb{H}_{\infty-}\times\mathbb{V}_{\infty-}$, by (3.5), (3.6), (3.8) and applying Itô's formula to $K_t D_\Theta X_t^x$, one can easily have

$$
D_\Theta X_t^x = J_t\left(\int_0^t K_s\sigma\left(X_s^x,\mathbb{P}_{X_s^x}\right)h(s)\mathrm{d}s + \int_0^t\int_{B_0}K_s f\left(\mathbb{P}_{X_s^x}\right)\mathbf{v}(s,z)N(\mathrm{d}z,\mathrm{d}s)\right),\quad \forall t\in[0,1]. \tag{3.33}
$$

1. Assume $\|\sigma^{-1}\|_\infty < \infty$. For any fixed $t \in (0,1]$ and $1 \leqslant j \leqslant d$, set $h_{t,j}$ and $\mathbf{v}$ as in (3.28). Define a matrix M_t by

$$
(M_t)_{ij} := D_{(h_{t,j},0)}X_t^{x,i},\quad 1\leqslant i,j\leqslant d,
$$

where $X_t^{x,i}$ stands for the i-th element of X_t^x. Then by (3.33) we obtain

$$M_t = \left(D_{(h_{t,1},0)}X_t^x, \dots, D_{(h_{t,d},0)}X_t^x\right) = J_t.$$

For any $g \in C_b^1(\mathbb{R}^d)$, by Theorem 2.6 we have

$$\begin{aligned}
\mathbb{E}\nabla g(X_t^x) &= \mathbb{E}\nabla g(X_t^x)M_tK_t = \sum_{i=1}^d \mathbb{E}\left(D_{(h_{t,i},0)}g(X_t^x)(K_t)_{i\cdot}\right)\\
&= \sum_{i=1}^d \mathbb{E}D_{(h_{t,i},0)}\left(g(X_t^x)(K_t)_{i\cdot}\right) - \sum_{i=1}^d \mathbb{E}\left(g(X_t^x)D_{(h_{t,i},0)}(K_t)_{i\cdot}\right)\\
&= \mathbb{E}\left[g(X_t^x)\sum_{i=1}^d\left((K_t)_{i\cdot}\int_0^t \langle h_{t,i}(s), \mathrm{d}W_s\rangle - D_{(h_{t,i},0)}(K_t)_{i\cdot}\right)\right],
\end{aligned}$$

where $(K_t)_{i\cdot}$ stands for the i-th row of K_t. Moreover, it follows from Hölder's inequality, (3.7), (3.14) and (3.30) that

$$\begin{aligned}
|\mathbb{E}\nabla g(X_t^x)| &\leqslant C\|g\|_\infty(1+|x|)\sum_{i=1}^d\left(\|h_{t,i}\|_{\mathbb{H}_2(t)} + \|h_{t,i}\|_{\mathbb{H}_4(t)} + \|h_{t,i}\|_{\mathbb{H}_8(t)}\right)\\
&\leqslant C\|g\|_\infty(1+|x|)t^{-\frac{1}{2}},
\end{aligned}$$

where C is a constant independent of x and t.
For any $y \in \mathbb{R}^d$ and $g \in C_b^1(\mathbb{R}^d)$, also by Theorem 2.6 one has

$$\begin{aligned}
\nabla_y\mathbb{E}g(X_t^x) &= \mathbb{E}\left(\nabla g(X_t^x)\nabla_y X_t^x\right)\\
&= \mathbb{E}\left(\nabla g(X_t^x)M_tK_t\nabla_y X_t^x\right) = \sum_{j=1}^d \mathbb{E}\left(D_{(h_{t,j},0)}g(X_t^x)(K_t\nabla_y X_t^x)_{j\cdot}\right)\\
&= \sum_{j=1}^d \mathbb{E}D_{(h_{t,j},0)}\left(g(X_t^x)(K_t\nabla_y X_t^x)_{j\cdot}\right)\\
&\quad - \sum_{j=1}^d \mathbb{E}\left(g(X_t^x)D_{(h_{t,j},0)}(K_t\nabla_y X_t^x)_{j\cdot}\right)\\
&= \mathbb{E}\Bigg[g(X_t^x)\sum_{j=1}^d\Bigg((K_t)_{j\cdot}\nabla_y X_t^x\int_0^t\langle h_{t,j}(s), \mathrm{d}W_s\rangle\\
&\qquad - D_{(h_{t,j},0)}(K_t)_{j\cdot}\nabla_y X_t^x - (K_t)_{j\cdot}(D_{(h_{t,j},0)}\nabla_y X_t^x)\Bigg)\Bigg].
\end{aligned}$$

Hence, by Hölder's inequality, (3.7), (3.9), (3.14), (3.18), (3.24) and (3.30) one can arrive at

$$
\begin{aligned}
|\nabla_y \mathbb{E} g(X_t^x)| &\leqslant C\|g\|_\infty \sum_{j=1}^d \Bigg[\Big(\mathbb{E}\sup_{t\in[0,1]}|K_t|^4\Big)^{\frac14}\Big(\mathbb{E}\sup_{t\in[0,1]}|\nabla_y X_t^x|^4\Big)^{\frac14}\|h_{t,j}\|_{\mathbb{H}_2(t)} \\
&\quad + \Big(\mathbb{E}\sup_{s\leqslant t}|D_{(h_{t,j},0)}K_s|^2\Big)^{\frac12}\Big(\mathbb{E}\sup_{t\in[0,1]}|\nabla_y X_t^x|^2\Big)^{\frac12} \\
&\quad + \Big(\mathbb{E}\sup_{t\in[0,1]}|K_t|^2\Big)^{\frac12}\Big(\mathbb{E}\sup_{t\in[0,1]}|D_{(h_{t,j},0)}\nabla_y X_t^x|^2\Big)^{\frac12}\Bigg] \\
&\leqslant C\|g\|_\infty(1+|x|)|y|\sum_{j=1}^d\big(\|h_{t,j}\|_{\mathbb{H}_2(t)}+\|h_{t,j}\|_{\mathbb{H}_4(t)}+\|h_{t,j}\|_{\mathbb{H}_8(t)}\big) \\
&\leqslant C\|g\|_\infty(1+|x|)|y|t^{-\frac12}.
\end{aligned}
$$

2. Assume $\|f^{-1}\|_\infty < \infty$. For any fixed $t\in(0,1]$ and $1\leqslant j\leqslant d$, set

$$h\equiv 0 \quad\text{and}\quad \mathbf{v}_{t,j}(s,z) = f^{-1}\big(\mathbb{P}_{X_s^x}\big)(J_s)_{\cdot j}\zeta_t(z), \quad \forall s\in[0,t],\ z\in B_0.$$

Define a matrix $\hat{M}_t$ by

$$(\hat{M}_t)_{ij} := D_{(0,\mathbf{v}_{t,j})}X_t^{x,i}, \quad 1\leqslant i,j\leqslant d.$$

Then by (3.33) we obtain

$$\hat{M}_t = \big(D_{(0,\mathbf{v}_{t,j})}X_t^x,\ldots,D_{(0,\mathbf{v}_{t,j})}X_t^x\big) = J_t\int_0^t\int_{B_0}\zeta_t(z)N(\mathrm{d}z,\mathrm{d}s) =: J_tH_t.$$

For any $g\in C_b^1(\mathbb{R}^d)$, due to Theorem 2.6 we have

$$
\begin{aligned}
\mathbb{E}\nabla g(X_t^x) &= \mathbb{E}\Big(\nabla g(X_t^x)\hat{M}_tK_tH_t^{-1}\Big) = \sum_{j=1}^d \mathbb{E}\big(D_{(0,\mathbf{v}_{t,j})}g(X_t^x)(K_t)_{j\cdot}H_t^{-1}\big) \\
&= \sum_{j=1}^d \mathbb{E}D_{(0,\mathbf{v}_{t,j})}\big(g(X_t^x)(K_t)_{j\cdot}H_t^{-1}\big) - \sum_{j=1}^d\mathbb{E}\big(g(X_t^x)D_{(0,\mathbf{v}_{t,j})}\big((K_t)_{j\cdot}H_t^{-1}\big)\big) \\
&= \mathbb{E}\Bigg[g(X_t^x)\sum_{j=1}^d\big((K_t)_{j\cdot}H_t^{-1}\delta_t(\mathbf{v}_{t,j}) - D_{(0,\mathbf{v}_{t,j})}(K_t)_{j\cdot}H_t^{-1} + (K_t)_{j\cdot}H_t^{-2}G_{t,j}\big)\Bigg].
\end{aligned}
$$

Moreover, it follows from (3.7), (3.14), (3.26), (3.31) and (3.32) that

$$
\begin{aligned}
|\mathbb{E}\nabla g(X_t^x)| &\leqslant C\|g\|_\infty\sum_{j=1}^d\Big(\|K_t\|_{L^4}\|H_t^{-1}\|_{L^4}\|\delta_t(\mathbf{v}_{t,j})\|_{L^2} + \|D_{(0,\mathbf{v}_{t,j})}K_t\|_{L^2}\|H_t^{-1}\|_{L^2} \\
&\quad + \|K_t\|_{L^4}\|H_t^{-2}\|_{L^4}\|G_{t,j}\|_{L^2}\Big)
\end{aligned}
$$

$$\leqslant C\|g\|_\infty(1+|x|)\left(t^{-\frac{3}{\alpha}}t^{\frac{2}{\alpha}}+t^{-\frac{3}{\alpha}}t^{\frac{3}{\alpha}}+t^{-\frac{6}{\alpha}}t^{\frac{5}{\alpha}}\right)$$
$$\leqslant C\|g\|_\infty(1+|x|)t^{-\frac{1}{\alpha}}.$$

where C is a constant independent of x and t.
For any $y \in \mathbb{R}^d$ and $g \in C_b^1(\mathbb{R}^d)$,

$$\begin{aligned}
\nabla_y \mathbb{E}g(X_t^x) &= \mathbb{E}\left(\nabla g(X_t^x)\nabla_y X_t^x\right) = \mathbb{E}\left(\nabla g(X_t^x)\hat{M}_t K_t H_t^{-1}\nabla_y X_t^x\right)\\
&= \sum_{j=1}^d \mathbb{E}\left(D_{(0,\mathbf{v}_{t,j})} g(X_t^x)(K_t H_t^{-1}\nabla_y X_t^x)_{j\cdot}\right)\\
&= \sum_{j=1}^d \mathbb{E}D_{(0,\mathbf{v}_{t,j})}\left(g(X_t^x)(K_t H_t^{-1}\nabla_y X_t^x)_{j\cdot}\right)\\
&\quad - \sum_{j=1}^d \mathbb{E}\left(g(X_t^x)D_{(0,\mathbf{v}_{t,j})}(K_t H_t^{-1}\nabla_y X_t^x)_{j\cdot}\right)\\
&= \mathbb{E}\Bigg[g(X_t^x)\sum_{j=1}^d \Big((K_t)_{j\cdot}H_t^{-1}\nabla_y X_t^x \delta_t(\mathbf{v}_{t,j})\\
&\quad - D_{(0,\mathbf{v}_{t,j})}(K_t)_{j\cdot}H_t^{-1}\nabla_y X_t^x + (K_t)_{j\cdot}H_t^{-2}G_{t,j}\nabla_y X_t^x\\
&\quad - (K_t)_{j\cdot}H_t^{-1}(D_{(0,\mathbf{v}_{t,j})}\nabla_y X_t^x)\Big)\Bigg].
\end{aligned}$$

Hence, by Hölder's inequality, (3.7), (3.14), (3.24), (3.26), (3.31) and (3.32) one can arrive at

$$\begin{aligned}
|\nabla_y \mathbb{E}g(X_t^x)| &\leqslant C\|g\|_\infty \sum_{j=1}^d \Big[\|K_t\|_{L^8}\|H_t^{-1}\|_{L^4}\|\nabla_y X_t^x\|_{L^8}\|\delta_t(\mathbf{v}_{t,j})\|_{L^2}\\
&\quad + \|D_{(0,\mathbf{v}_{t,j})}K_t\|_{L^2}\|H_t^{-1}\|_{L^4}\|\nabla_y X_t^x\|_{L^4}\\
&\quad + \|K_t\|_{L^4}\|H_t^{-2}\|_{L^2}\|G_{t,j}\|_{L^8}\|\nabla_y X_t^x\|_{L^8}\\
&\quad + \|K_t\|_{L^4}\|H_t^{-1}\|_{L^4}\|D_{(0,\mathbf{v}_{t,j})}\nabla_y X_t^x\|_{L^2}\Big]\\
&\leqslant C\|g\|_\infty(1+|x|)|y|\left(t^{-\frac{3}{\alpha}}t^{\frac{2}{\alpha}}+t^{-\frac{3}{\alpha}}t^{\frac{3}{\alpha}}+t^{-\frac{6}{\alpha}}t^{\frac{5}{\alpha}}+t^{-\frac{3}{\alpha}}t^{\frac{2}{\alpha}}\right)\\
&\leqslant C\|g\|_\infty(1+|x|)|y|t^{-\frac{1}{\alpha}}.
\end{aligned}$$

□

Let's give the proof of Corollary 1.4.

Proof We only prove the first statement, since the second one can be obtained by the same argument. For any $x_1, x_2 \in \mathbb{R}^d$, according to Lemma 2.1.1 in [21] and (1.6), both of $X_t^{x_1}$ and $X_t^{x_2}$ have density functions denoted by $p_t(x_1, y)$ and $p_t(x_2, y)$ respectively. By (1.7), we have

$$\begin{aligned}\left|\mathbb{E}g(X_t^{x_1}) - \mathbb{E}g(X_t^{x_2})\right| &= \left|\int_0^1 \frac{\mathrm{d}}{\mathrm{d}r}\mathbb{E}g(X_t^{x_2+r(x_1-x_2)})\mathrm{d}r\right| \\ &\leqslant \int_0^1 \left|\nabla_{x_1-x_2}\mathbb{E}f(X_t^{x_2+r(x_1-x_2)})\right| \mathrm{d}r \\ &\leqslant C\|g\|_\infty(1+|x_1|+|x_2|)|x_1-x_2|t^{-\frac{1}{2}}.\end{aligned}$$

Hence,

$$\begin{aligned}\int_{\mathbb{R}^d}|p_t(x_1,y)-p_t(x_2,y)|\mathrm{d}y &= \sup_{\|g\|_\infty\leqslant 1, g\in\mathscr{B}_b(\mathbb{R}^d)}\left|\mathbb{E}g(X_t^{x_1}) - \mathbb{E}g(X_t^{x_2})\right| \\ &= \sup_{\|g\|_\infty\leqslant 1, g\in C_b^1(\mathbb{R}^d)}\left|\mathbb{E}g(X_t^{x_1}) - \mathbb{E}g(X_t^{x_2})\right| \\ &\leqslant C(1+|x_1|+|x_2|)|x_1-x_2|t^{-\frac{1}{2}}.\end{aligned}$$

□

3.4 Proof of Theorem 1.5

Proof The proof is divided into three steps.

Step 1 We first prove that for any $\mu_1, \mu_2 \in \mathcal{P}_2$,

$$\mathbb{W}_2(P_{s,t}^*\mu_1, P_{s,t}^*\mu_2)^2 \leqslant \mathbb{W}_2(\mu_1,\mu_2)^2\mathrm{e}^{-(C_2-C_1)(t-s)}. \tag{3.34}$$

Without loss of generality, we only prove the case for $s = 0$. Let ξ_1 and ξ_2 be two square-integrable and $\mathscr{F}_0$-measurable random variables such that

$$\mathbb{W}_2(\mu_1,\mu_2)^2 = \mathbb{E}|\xi_1-\xi_2|^2.$$

Denote by $X_t^{\xi_1}$ and $X_t^{\xi_2}$ the solutions to (1.1) with initial value ξ_1 and ξ_2, respectively. By $(H3)$ and Itô's formula, we have

$$\begin{aligned}&\mathbb{E}\left(|X_t^{\xi_1}-X_t^{\xi_2}|^2\mathrm{e}^{(C_2-C_1)t}\right) \\ &= \mathbb{W}_2(\mu_1,\mu_2)^2 + 2\int_0^t \mathbb{E}\Big(\Big\langle X_s^{\xi_1}-X_s^{\xi_2}, b\Big(X_s^{\xi_1},\mathbb{P}_{X_s^{\xi_1}}\Big) - b\Big(X_s^{\xi_2},\mathbb{P}_{X_s^{\xi_2}}\Big)\Big\rangle \\ &\quad + \|\sigma(X_s^{\xi_1},\mathbb{P}_{X_s^{\xi_1}}) - \sigma\Big(X_s^{\xi_2},\mathbb{P}_{X_s^{\xi_2}}\Big)\|_{HS}^2 \\ &\quad + \int_{\mathbb{R}_0^d}|z|^2\nu(\mathrm{d}s)|f(\mathbb{P}_{X_s^{\xi_1}}) - f(\mathbb{P}_{X_s^{\xi_2}})|^2\Big)\mathrm{e}^{(C_2-C_1)s}\mathrm{d}s\end{aligned}$$

$$
\begin{aligned}
&+ (C_2 - C_1)\int_0^t \mathbb{E}|X_s^{\xi_1} - X_s^{\xi_2}|^2 \mathrm{e}^{(C_2-C_1)s}\mathrm{d}s\\
&\leqslant \mathbb{W}_2(\mu_1,\mu_2)^2 + \int_0^t \mathbb{E}\Big(C_1\mathbb{W}_2(\mathbb{P}_{X_s^{\xi_1}},\mathbb{P}_{X_s^{\xi_1}})^2 - C_2|X_s^{\xi_1} - X_s^{\xi_2}|^2\Big)\mathrm{e}^{(C_2-C_1)s}\mathrm{d}s\\
&+ (C_2 - C_1)\int_0^t \mathbb{E}|X_s^{\xi_1} - X_s^{\xi_2}|^2 \mathrm{e}^{(C_2-C_1)s}\mathrm{d}s\\
&\leqslant \mathbb{W}_2(\mu_1,\mu_2)^2. \qquad (3.35)
\end{aligned}
$$

Hence,

$$
\mathbb{W}_2(P_t^*\mu_1, P_t^*\mu_2)^2 \leqslant \mathbb{E}\left(|X_t^{\xi_1} - X_t^{\xi_2}|^2\right) \leqslant \mathbb{W}_2(\mu_1,\mu_2)^2\mathrm{e}^{-(C_2-C_1)t}.
$$

Step 2 We prove the existence and uniqueness of the invariant measure. Let X_t^0 denote the solution with initial value 0 and $\epsilon_0 := \frac{C_2-C_1}{4}$. By Itô's formula, $(H3)$, (3.2) and Young's inequality, we have

$$
\begin{aligned}
&\mathbb{E}\left(|X_t^0|^2\mathrm{e}^{(C_2-C_1-2\epsilon_0)t}\right)\\
&= \mathbb{E}\int_0^t \Bigg(2\langle b(X_s^0,\mathbb{P}_{X_s^0}), X_s^0\rangle + \|\sigma(X_s^0,\mathbb{P}_{X_s^0})\|_{HS}^2\\
&\quad + \int_{\mathbb{R}_0^d}|z|^2\nu(\mathrm{d}z)|f(\mathbb{P}_{X_s^0})|^2\Bigg)\mathrm{e}^{(C_2-C_1-2\epsilon_0)s}\mathrm{d}s\\
&\quad + (C_2 - C_1 - 2\epsilon_0)\int_0^t \mathbb{E}|X_s^0|^2\mathrm{e}^{(C_2-C_1-2\epsilon_0)s}\mathrm{d}s\\
&\leqslant C_0 + \int_0^t (C_1+\epsilon_0)\mathbb{W}_2(\mathbb{P}_{X_s^0},\delta_0)^2 - (C_2-\epsilon_0)\mathbb{E}|X_s^0|^2\mathrm{d}s\\
&\quad + (C_2 - C_1 - 2\epsilon_0)\int_0^t \mathbb{E}|X_s^0|^2\mathrm{e}^{(C_2-C_1-2\epsilon_0)s}\mathrm{d}s\\
&\leqslant C_0,
\end{aligned}
$$

where C_0 is a constant depending on ϵ_0 and the values of b,σ at the point $(0,\delta_0)$ and f at δ_0. Then we have

$$
\sup_{t\geqslant 0}\mathbb{E}|X_t^0|^2 \leqslant C_0\sup_{t\geqslant 0}\mathrm{e}^{-(C_2-C_1-2\epsilon_0)t} \leqslant C_0. \qquad (3.36)
$$

Recalling the weak uniqueness of the solution, we have

$$
P_t^*(P_s^*\delta_0) = P_{t+s}^*\delta_0, \quad s,t\geqslant 0.
$$

This, together with (3.34) and (3.36), yields

$$
\mathbb{W}_2(P_{t+s}^*\delta_0, P_t^*\delta_0)^2 \leqslant \mathrm{e}^{-(C_2-C_1)t}\mathbb{E}|X_s^0|^2 \leqslant C_0\mathrm{e}^{-(C_2-C_1)t}.
$$

Then,

$$\lim_{t\to\infty}\sup_{s\geqslant 0}\mathbb{W}_2(P_{t+s}^*\delta_0, P_t^*\delta_0)=0, \tag{3.37}$$

which means that $\{P_t^*\delta_0\}_{t\geqslant 0}$ is a $\mathbb{W}_2$-Cauchy family when $t\to\infty$. Then, there is a unique probability measure $\hat{\mu}\in\mathcal{P}_2$ such that

$$\lim_{t\to\infty}\mathbb{W}_2(P_t^*\delta_0,\hat{\mu})=0. \tag{3.38}$$

Then it follows from (3.34), (3.37) and (3.38) that

$$\begin{aligned}\mathbb{W}_2(P_t^*\hat{\mu},\hat{\mu}) &\leqslant \lim_{s\to\infty}\mathbb{W}_2(P_t^*\hat{\mu}, P_t^*P_s^*\delta_0)+\lim_{s\to\infty}\mathbb{W}_2(P_t^*P_s^*\delta_0, P_s^*\delta_0)\\ &\quad+\lim_{s\to\infty}\mathbb{W}_2(P_s^*\delta_0,\hat{\mu})=0,\end{aligned}$$

which means that $\hat{\mu}$ is an invariant measure for P_t^* indeed.

Step 3 Let ξ be an $\mathscr{F}_0$-measurable random variable with distribution μ. For any $t>1$, by Markov property and Theorem 1.2 we have

$$\begin{aligned}\left|\mathbb{E}g(X_t^\xi)-\int_{\mathbb{R}^d}g(y)\hat{\mu}(\mathrm{d}y)\right| &= \left|\mathbb{E}g(X_t^\xi)-\int_{\mathbb{R}^d}\mathbb{E}g(X_t^y)\hat{\mu}(\mathrm{d}y)\right|\\ &\leqslant \int_{\mathbb{R}^d}\left|\mathbb{E}g(X_t^\xi)-\mathbb{E}g(X_t^y)\right|\hat{\mu}(\mathrm{d}y)\\ &\leqslant \int_{\mathbb{R}^d}\left|\mathbb{E}\left[\mathbb{E}\left(g(X_1^{X_{t-1}^\xi})-\mathbb{E}g(X_t^{X_{t-1}^y})|\mathscr{F}_{t-1}\right)\right]\right|\hat{\mu}(\mathrm{d}y)\\ &\leqslant C\|g\|_\infty\int_{\mathbb{R}^d}|\mathbb{E}(X_{t-1}^\xi-X_{t-1}^y)|\hat{\mu}(\mathrm{d}y)\\ &\leqslant C\|g\|_\infty\int_{\mathbb{R}^d}\left(\mathbb{E}|\xi-y|^2\right)^{\frac{1}{2}}\hat{\mu}(\mathrm{d}y)\mathrm{e}^{-\frac{1}{2}(C_2-C_1)(t-1)}\\ &\leqslant C\|g\|_\infty\left[1+\left(\int_{\mathbb{R}^d}|x|^2\mu(\mathrm{d}x)\right)^{\frac{1}{2}}\right]\mathrm{e}^{-\frac{1}{2}(C_2-C_1)t}.\end{aligned}$$

Hence,

$$\begin{aligned}\|P_t^*\mu-\hat{\mu}(\cdot)\|_{TV} &= \sup_{\|g\|_\infty\leqslant 1, g\in C_b^1}\left|\mathbb{E}g(X_t^\xi)-\int_{\mathbb{R}^d}g(y)\hat{\mu}(\mathrm{d}y)\right|\\ &\leqslant C\left[1+\left(\int_{\mathbb{R}^d}|x|^2\mu(\mathrm{d}x)\right)^{\frac{1}{2}}\right]\mathrm{e}^{-\frac{1}{2}(C_2-C_1)t},\end{aligned} \tag{3.39}$$

where C is a constant independent of t and μ. It is obvious that for C large enough, (3.39) holds for all $t\in[0,1]$. So we finish the proof.

□

4 An Example

In this section, as an application of our main results, we study the classic McKean–Vlasov equation. Given $b_0 : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma_0 : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $f_0 : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$, we assume:

$(A1)$ b_0 and σ_0 are twice differentiable functions with bounded derivatives; f_0 is a differential function with a bounded derivative.

$(A2)$ For some $C_0 > 0$,

$$|\langle \sigma_0(x)\xi, \xi\rangle| \geqslant C_0, \quad \forall x \in \mathbb{R}^d, \quad \forall \xi \in \mathbb{S}^d.$$

$(A3)$ For some $C_1 > 0$,

$$|\langle f_0(x)\xi, \xi\rangle| \geqslant C_1, \quad \forall x \in \mathbb{R}^d, \quad \forall \xi \in \mathbb{S}^d.$$

$(A4)$ There exists $\lambda > 0$ such that

$$2\langle b_0(y_1) - b_0(y_2), y_1 - y_2\rangle \leqslant -\lambda |y_1 - y_2|^2, \quad \forall y_1, y_2 \in \mathbb{R}^d.$$

Define

$$b(x,\mu) = \int_{\mathbb{R}^d} b_0(x-y)\mu(\mathrm{d}y), \sigma(x,\mu) = \int_{\mathbb{R}^d} \sigma_0(x-y)\mu(\mathrm{d}y), \ \forall x \in \mathbb{R}^d, \mu \in \mathcal{P}_2,$$

and

$$f(\mu) = \int_{\mathbb{R}^d} f_0(y)\mu(\mathrm{d}y), \quad \forall \mu \in \mathcal{P}_2.$$

For $\alpha \in (0,2)$, $\{L_t\}_{t\geqslant 0}$ is a d-dimensional truncated α-stable process with Lévy measure $\frac{I_{B_0}(z)\mathrm{d}z}{|z|^{d+\alpha}}$, while $\{W_t\}_{t\geqslant 0}$ is a d-dimensional Brownian motion independent of L. Now consider the following equation:

$$X_t^x = x + \int_0^t b\big(X_s^x, \mathbb{P}_{X_s^x}\big)\mathrm{d}s + \int_0^t \sigma\big(X_s^x, \mathbb{P}_{X_s^x}\big)\mathrm{d}W_s + \int_0^t f\big(\mathbb{P}_{X_s^x}\big)\mathrm{d}L_s.$$

Then we have the following results:

Theorem 4.1 1. *Assume $(A1)$ and $(A2)$. Then there exists $C > 0$ such that*

$$|\mathbb{E}\nabla g(X_t^x)| \leqslant C\|g\|_\infty t^{-\frac{1}{2}}, \ |\nabla_y \mathbb{E}g(X_t^x)| \leqslant C\|g\|_\infty t^{-\frac{1}{2}}, \ \forall g \in C_b^1(\mathbb{R}^d), \ y \in \mathbb{R}^d.$$

2. *Assume $(A1)$ and $(A3)$. Then there exists $C > 0$ such that*

$$|\mathbb{E}\nabla g(X_t^x)| \leqslant C\|g\|_\infty t^{-\frac{1}{\alpha}}, \ |\nabla_y \mathbb{E}g(X_t^x)| \leqslant C\|g\|_\infty t^{-\frac{1}{\alpha}}, \ \forall g \in C_b^1(\mathbb{R}^d), \ y \in \mathbb{R}^d.$$

3. *Assume (A1), (A2)(or (A3)) and (A4) hold. Let* $\|\nabla\sigma_0\|_{HS,\infty} := \sup_{|v|=1, x\in\mathbb{R}^d} \|\nabla_v\sigma_0(x)\|_{HS} < \infty$. *If*

$$\lambda_0 := \lambda - 1 - \|\nabla b_0\|_\infty^2 - 4\|\nabla\sigma_0\|_{HS,\infty}^2 - \int_{B_0} |z|^2\nu(\mathrm{d}z)\|\nabla f_0\|_\infty^2 > 0, \quad (4.1)$$

then there exists a unique invariant measure Ξ *such that for any* $\mu_0 \in \mathcal{P}_2$,

$$\|P_t^*\mu_0 - \Xi\|_{TV} \leqslant C\left(1 + \left(\int_{B_0} |x|^2\mu_0(\mathrm{d}x)\right)^{\frac{1}{2}}\right)e^{-\frac{1}{2}\lambda_0 t}.$$

Proof We divide the proof into two steps.

Step 1 In this part, we prove the statements (1) and (2). It suffices for us to verify the conditions required in Theorems 1.2 and 1.5. In fact, due to $(A1)$ it is easy to see that b and σ are twice differentiable with respect to the first variable x and

$$\|\partial_x^i b\|_\infty := \sup_{x\in\mathbb{R}^d, \mu\in\mathcal{P}_2} |\partial_x^i b(x,\mu)| < \infty,$$
$$\|\partial_x^i \sigma\|_\infty := \sup_{x\in\mathbb{R}^d, \mu\in\mathcal{P}_2} |\partial_x^i \sigma(x,\mu)| < \infty, i = 1, 2.$$

Moreover, for all $x \in \mathbb{R}^d$ and $\mu_1, \mu_2 \in \mathcal{P}_2$,

$$|b(x,\mu_1) - b(x,\mu_2)| \leqslant \int_{\mathbb{R}^d} |b_0(x-y) - b_0(x-z)|\pi(\mathrm{d}y,\mathrm{d}z)$$
$$\leqslant \|\nabla b_0\|_\infty \left(\int_{\mathbb{R}^d} |y-z|^2\pi(\mathrm{d}y,\mathrm{d}z)\right)^{\frac{1}{2}},$$

where π is an arbitrary coupling of μ_1 and μ_2. Hence,

$$|b(x,\mu_1) - b(x,\mu_2)| \leqslant \|\nabla b_0\|_\infty \mathbb{W}_2(\mu_1,\mu_2).$$

By similar arguments, we can prove σ, ∇b and f are all Lipschitz continuous with respect to the second variable μ.
For any $x \in \mathbb{R}^d$, due to Example 2.4 we have

$$\partial_\mu b(x,\mu)(y) = -\nabla b_0(x-y), \quad \forall y \in \mathbb{R}^d,$$

which is Lipschitz continuous with respect to both of y and μ. So $b(x,\cdot)$ is in $C_b^{1,1}(\mathcal{P}_2)$. Furthermore,

$$\partial_x\partial_\mu b(x,\mu)(y) = -\nabla^2 b_0(x-y)$$

is bounded on $\mathbb{R}^d \times \mathcal{P}_2 \times \mathbb{R}^d$. The same argument can derive $\sigma(x,\cdot)$, $f(\cdot) \in C_b^{1,1}(\mathcal{P}_2)$ with $\partial_\mu\sigma$, $\partial_\mu f$, and $\partial_x\partial_\mu\sigma$ bounded.

By $(A2)$ and the continuity of map $(x,\xi)\mapsto\langle\sigma_0(x)\xi,\xi\rangle$ we have either

$$\langle\sigma_0(x)\xi,\xi\rangle\geqslant C_0,\quad\forall x\in\mathbb{R}^d,\xi\in\mathbb{S}^d. \tag{4.2}$$

or

$$\langle\sigma_0(x)\xi,\xi\rangle\leqslant -C_0,\quad\forall x\in\mathbb{R}^d,\xi\in\mathbb{S}^d. \tag{4.3}$$

Without loss of generality we assume (4.2) holds, then

$$\langle\sigma(x,\mu)\xi,\xi\rangle=\int_{\mathbb{R}^d}\langle\sigma_0(x-y)\xi,\xi\rangle\mu(\mathrm{d}y)\geqslant C_0,\quad\forall x\in\mathbb{R}^d,\mu\in\mathcal{P}_2,\xi\in\mathbb{S}^d,$$

which implies $\sigma^{-1}(x,\mu)$ exists for all $x\in\mathbb{R}^d$ and $\mu\in\mathcal{P}_2$. Moreover, we have $\|\sigma^{-1}\|_\infty<\infty$. Similarly, by $(A1)$ and $(A3)$ we can obtain $\|f^{-1}\|_\infty<\infty$. Now, according to Theorem 1.2 we have proved the statements of (1) and (2).

Step 2 For any $x_1,x_2\in\mathbb{R}^d$ and $\mu_1,\mu_2\in\mathcal{P}_2$, we have

$$\begin{aligned}
&2\langle b(x_1,\mu_1)-b(x_2,\mu_2),x_1-x_2\rangle\\
&\quad=2\int_{\mathbb{R}^d}\langle b_0(x_1-y)-b_0(x_2-y),x_1-x_2\rangle\mu_1(\mathrm{d}y)\\
&\qquad+2\int_{\mathbb{R}^d}\langle b_0(x_2-z_1)-b_0(x_2-z_2),x_1-x_2\rangle\pi(\mathrm{d}z_1,\mathrm{d}z_2)\\
&\quad\leqslant-\lambda|x_1-x_2|^2+2\|\nabla b_0\|_\infty\left(\int_{\mathbb{R}^d}|z_1-z_2|^2\pi(\mathrm{d}z_1,\mathrm{d}z_2)\right)^{\frac12}|x_1-x_2|,
\end{aligned}$$

where π is a coupling of μ_1 and μ_2. Thus,

$$\begin{aligned}
&2\langle b(x_1,\mu_1)-b(x_2,\mu_2),x_1-x_2\rangle\\
&\quad\leqslant-\lambda|x_1-x_2|^2+2\|\nabla b_0\|_\infty\mathbb{W}_2(\mu_1,\mu_2)|x_1-x_2|\\
&\quad\leqslant-(\lambda-1)|x_1-x_2|^2+\|\nabla b_0\|_\infty^2\mathbb{W}_2(\mu_1,\mu_2)^2.
\end{aligned}\tag{4.4}$$

Meanwhile, the same arguments can derive

$$\begin{aligned}
&\|\sigma(x_1,\mu_1)-\sigma(x_2,\mu_2)\|_{HS}^2+\int_{B_0}|z|^2\nu(\mathrm{d}z)|f(\mu_1)-f(\mu_2)|^2\\
&\quad\leqslant2\|\nabla\sigma_0\|_{HS,\infty}^2|x_1-x_2|^2+2\|\nabla\sigma_0\|_{HS,\infty}^2\mathbb{W}_2(\mu_1,\mu_2)^2\\
&\qquad+\int_{B_0}|z|^2\nu(\mathrm{d}z)\|\nabla f_0\|_\infty^2\mathbb{W}_2(\mu_1,\mu_2)^2.
\end{aligned}\tag{4.5}$$

By (4.1), (4.4), (4.5) and Theorem 1.5, we immediately obtain the claim (3).

□

Acknowledgements The author is very grateful to the editor and referee for detailed reports and corrections. He also would like to thank Professors Zhao Dong, Renming Song, and Fengyu Wang for their valuable discussions. This work is supported by National Natural Science Foundation of China (Nos. 11501286, 11790272), Natural Science Foundation of Jiangsu Province (No. BK20150564) and CSC.

References

1. Barczy, M., Li, Z.H., Pap, M.: Yamada–Watanabe results for stochastic differential equations with jumps. Int. J. Stoch. Anal. Art. ID 460472 (2015)
2. Benachour, S., Roynette, B., Talay, D., Vallois, P.: Nonlinear selfstabilizing processes. I. Existence, invariant probability, propagation of chaos. Stoch. Process. Appl. **75**, 173–201 (1998)
3. Bismut, J.M.: Calcul des variations stochastiques et processus de sauts. Z. Wahrsch. Verw. Gebiete **63**, 147–235 (1983)
4. Bitchtler, K., Jacod, J.J.: Gravereaux, Malliavin Calculus for Processes with Jumps. Gordan and Breach Science Publishers, New York (1987)
5. Buckdahn, R., Li, J., Peng, S., Rainer, C.: Mean-field stochastic differential equations and associated PDEs. Ann. Prob. **45**, 824–878 (2017)
6. Butkovsky, O.A.: On ergodic properties of nonlinear Markov chains and stochastic Mckean–Vlasov equations. Theory Probab. Appl. **58**, 661–674 (2014)
7. Cattiaux, P., Guillin, A., Malrieu, F.: Probability approach for granular media equations in the non-uniformly convex case. Probab. Theory Relat. Fields **140**, 19–40 (2008)
8. Cardaliaguet, P.: Notes on mean filed games (from P.L. Lions' lectures at Collège de France) (2013)
9. Carmona, R., Delarue, F.: Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics. Ann. Probab. **43**, 2647–2700 (2015)
10. Chan, T.: Dynamics of the McKean–Vlasov equation. Ann. Probab. **22**, 431–441 (1994)
11. Dawson, D.: Stochastic McKean–Vlasov equations. Nonlinear Differ. Equ. Appl. **2**, 199–229 (1995)
12. Graham, C.: McKean–Vlasov Ito–Skorohod equations, and nonlinear diffusions with discrete jump sets. Stoch. Process. Appl. **40**, 69–82 (1992)
13. Hao, T., Li, J.: Mean-field SDEs with jumps and nonlocal integral-PDEs. Nonlinear Differ. Equ. Appl. **23**, 17 (2016)
14. Huang, X., Röckner, M., Wang, F.-Y.: Nonlinear Fokker–Planck equations for probability measures on path space and path-distribution dependent SDEs. arXiv:1709.00556v1 (2017)
15. Kac, M.: Foundations of kinetic theory. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. III, pp. 171–197. University California Press, Berkeley (1956)
16. Kolokoltsov, V.N.: Nonlinear Markov Processes and Kinetic Equations. Cambridge University Press, Cambridge (2010)
17. Kotelenez, P., Kurtz, T.: Macroscopic limits for stochastic partial differential equations of McKean–Vlasov type. Probab. Theory Relat. Fields **146**, 189–222 (2010)
18. Lions, P.: Cours au Collège de France: Théorie des jeuàchamps moyens. Available at http://www.college-de-france.fr/default/EN/all/equ[1]der/audiovideo.jsp
19. McKean, H.P.: Propagation of chaos for a class of nonlinear parabolic equations. In: Lecture Series in Differential Equations, vol. 7, pp. 41–57 (1967)
20. Protter, E.: Stochastic Integration and Differential Equations, 2nd edn. Springer, Berlin (2004)
21. Nualart, D.: The Malliavin Calsulus and Related Topics. Springer, New York (2006)
22. Song, Y.: Gradient estimates and coupling property for semilinear SDEs driven by jump processes. Sci. China Math. **58**, 447–458 (2015)
23. Song, Y., Xie, Y.: Existence of density functions for the running maximum of a Lévy–Itô diffusion. Potential Anal. **48**(1), 35–48 (2018)
24. Song, Y., Zhang, X.: Regularity of density for SDEs driven by degenrate Lévy noises. Electron. J. Probab. **20**(21), 1–27 (2015)
25. Wang, F.-Y.: Gradient estimate for Ornstein–Uhlenbeck jump processes. Stoch. Process. Appl. **121**, 466–478 (2011)
26. Wang, F.-Y.: Distribution dependent SDEs for Landau type equations. Stoch. Process. Appl. **128**, 595–621 (2018)

27. Wang, F.-Y., Xu, L., Zhang, X.: Gradient estimates for SDEs driven by multiplicative Lévy noise. J. Funct. Anal. **269**, 3195–3219 (2015)
28. Wang, L., Xie, L., Zhang, X.: Derivative formulae for SDEs driven by multiplicative α-stable-like processes. Stoch. Process. Appl. **125**, 867–885 (2015)
29. Veretennikov, A.Y.: On ergodic measures for Mckean-Vlasov stochastic equations. In: Harald, N., Denis, T. (eds.) Monte-Carlo and Quasi-Monte Carlo Methods, pp. 471–486. Springer, Berlin (2006)