Path-Distribution Dependent SDEs with Singular Coefficients*

Xing Huang [a)]

a)Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

xinghuang@tju.edu.cn

May 7, 2021

Abstract

In this paper, existence and uniqueness are proved for path-dependent McKean-Vlasov type SDEs with integrability conditions. Gradient estimates and the Harnack type inequalities are derived in the case that the drifts are Dini continuous in the space variable. These generalize the corresponding results derived for classical functional SDEs with singular coefficients.

AMS subject Classification: 60H1075, 60G44.
Keywords: Path-Distribution dependent SDEs, Krylov's estimate, Zvonkin's transform, Harnack inequality.

1 Introduction

The distribution dependent SDEs can be used to characterize nonlinear Fokker-Planck equations, see [3, 4, 14, 17] and references within for McKean-Vlasov type SDEs, and [2, 5, 6] and references therein for Landau type equations. One can also refer to [8] for the path-distribution dependent SDEs with regular conditions.

Recently, [10] studied the existence and uniqueness of distribution dependent SDEs with singular coefficients. The Harnack inequalities, shift Harnack inequalities and gradient estimates are also investigated in [10]. [16] also obtains the existence and uniqueness, estimate of heat kernel for singular distribution dependent SDEs. For more results on

*Supported in part by NNSFC (11801406).

distribution independent SDEs with singular coefficients, one can see [7, 13, 25, 21] and references therein, where Zvonkin's transform in [26] plays an important role.

The purpose of this paper is to extend results in [10] to path-distribution dependent SDEs with singular drift. Firstly, due to the distribution dependence, Girsanov's transform, which is a useful tool to prove the existence of weak solution for the classical SDEs is unavailable, especially in the case with distribution dependent diffusion coefficients. Thus, compared to the classical SDEs with singular drift, we will pay more attention to the proof of existence of weak solution. More precisely, we will apply an approximation technique similar to that in [10, 16] to obtain existence of weak solution. However, the path-distribution dependent drift will generate new difficulty, see the proof of Theorem 2.1(1) below. Next, by Lemma 3.3, the weak existence together with the strong uniqueness of the de-coupled SDEs implies the strong existence for SDE (1.1). To prove the strong uniqueness, we will use the technique in [10, Section 4.2], i.e. we first identify the distributions of two given solutions from the same initial value, so that these solutions solve the common reduced classical SDE, and thus, the strong uniqueness follows from existing argument developed for the classical SDEs. The essential difficulty lies in identifying the distributions of two solutions of (1.1). Finally, gradient estimates and the Harnack type inequalities can be proved by Zvonkin's transform combined with the existing result in [8].

Let $d \geq 1$ and fix a constant $r > 0$. Define $C([-r,0];\mathbb{R}^d)(C([-r,\infty);\mathbb{R}^d))$ as the set of $\mathbb{R}^d$-valued continuous functions on $[-r,0]([-r,\infty))$. Let $\mathscr{C} = C([-r,0];\mathbb{R}^d)$ be equipped with the uniform norm $\|\xi\|_{\mathscr{C}} =: \sup_{s\in[-r,0]}|\xi(s)|, \xi \in \mathscr{C}$. For any $f \in C([-r,\infty);\mathbb{R}^d)$, $t \geq 0$, define $f_t \in \mathscr{C}$ as $f_t(s) = f(t+s), s \in [-r,0]$, which is called the segment process.

Let $\mathscr{B}(\mathscr{C})$ be the Borelian σ-field on $\mathscr{C}$ and $\mathscr{P}$ be the set of all probability measures on $(\mathscr{C}, \mathscr{B}(\mathscr{C}))$ equipped with the weak topology. Consider the following path-distribution dependent SDE on $\mathbb{R}^d$:

$$(1.1) \qquad \mathrm{d}X(t) = B(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + b(t, X(t), \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X(t), \mathscr{L}_{X_t})\mathrm{d}W(t),$$

where $W(t)$ is a d-dimensional Brownian motion on a complete filtration probability space $(\Omega, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, $\mathscr{L}_{X_t}$ is the law of X_t, and

$$b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}^d, \quad B : \mathbb{R}_+ \times \mathscr{C} \times \mathscr{P} \to \mathbb{R}^d, \quad \sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}^d \otimes \mathbb{R}^d$$

are measurable.

Throughout the paper, we use $\|\cdot\|_\infty$ to denote the uniform norm, and $\mathscr{L}_\xi|\tilde{\mathbb{P}}$ to denote the law of a random variable ξ under the probability $\tilde{\mathbb{P}}$. Let $\mathscr{B}_b^+(\mathscr{C})(\mathscr{B}_b(\mathscr{C}))$denote the set of all bounded and non-negative(bounded) measurable functions on $\mathscr{C}$. We will use the letter C or c to denote a positive constant, and $C(\alpha)$ or $c(\alpha)$ stands for a constant depending on α. The values of the constants may change from one appearance to another.

The remainder of the paper is organized as follows. In Section 2, we summarize the main results of the paper. To prove these results, some preparations are addressed in

Section 3, including a new Krylov's estimate, one lemma on convergence of stochastic processes, and a result on the relationship between existence of strong solutions and weak ones for path-distribution dependent SDEs. Finally, the main results are proved in Sections 4 and 5.

2 Main Results

Let $\theta \in [1, \infty)$. We will consider the SDE (1.1) with initial distributions in the class

$$\mathscr{P}_\theta := \big\{\mu \in \mathscr{P} : \mu(\|\cdot\|_{\mathscr{C}}^\theta) < \infty\big\}.$$

According to [18, Theorem 6.18], $\mathscr{P}_\theta$ is a Polish space under the Wasserstein distance

$$\mathbb{W}_\theta(\mu,\nu) := \inf_{\pi \in \mathbf{C}(\mu,\nu)} \left(\int_{\mathscr{C}\times\mathscr{C}} \|\xi - \eta\|_{\mathscr{C}}^\theta \pi(\mathrm{d}\xi, \mathrm{d}\eta) \right)^{\frac{1}{\theta}}, \quad \mu,\nu \in \mathscr{P}_\theta,$$

where $\mathbf{C}(\mu,\nu)$ is the set of all couplings of μ and ν. Moreover, the topology induced by $\mathbb{W}_\theta$ coincides with the weak topology on $\mathscr{P}_\theta$, see [18, Definition 6.8, Theorem 6.9] for more details, where we can find that the weak convergence in $\mathscr{P}_\theta$ is equivalent to that in $\mathscr{P}$ together with uniform integrability.

In the following three subsections, we state our main results on the existence, uniqueness and Harnack type inequalities respectively for the path-distribution dependent SDE (1.1).

2.1 Existence and Uniqueness

Fix a constant $T > 0$, and we will only consider solutions of (1.1) up to time T. For a measurable function f defined on $[0,T] \times \mathbb{R}^d$, let

$$\|f\|_{L_p^q(s,t)} = \left(\int_s^t \left(\int_{\mathbb{R}^d} |f(v,x)|^p \mathrm{d}x \right)^{\frac{q}{p}} \mathrm{d}v \right)^{\frac{1}{q}}, \quad p,q \geq 1, 0 \leq s \leq t \leq T.$$

When $s = 0$, we simply denote $\|f\|_{L_p^q(0,t)} = \|f\|_{L_p^q(t)}$. A key step in the study of SDEs with integrable drift is to establish the Krylov type estimate (see for instance [10, 13, 25]). For later use we introduce the following class of number pairs (p,q):

$$\mathscr{K} := \Big\{(p,q) \in (1,\infty) \times (1,\infty) : \ \frac{d}{p} + \frac{2}{q} < 2\Big\}.$$

To construct a weak solution of (1.1) by using approximation argument as in [7, 10, 14, 16], we need the following conditions.

(H^θ) The following assumptions hold for some $\theta \geq 1$.

(1) For any $t \in [0,T], x \in \mathbb{R}^d, \xi \in \mathscr{C}$, $b(t,x,\cdot)$, $B(t,\xi,\cdot)$ and $\sigma(t,x,\cdot)$ are continuous in $\mathscr{P}$. There exists a constant $L > 0$ such that

$$\|\sigma(t,x,\mu) - \sigma(t,x,\nu)\| + |b(t,x,\mu) - b(t,x,\nu)| \leq L\,\mathbb{W}_\theta(\mu,\nu), \tag{2.1}$$

and

$$|B(t,\xi,\mu) - B(t,\xi,\nu)| \leq L\mathbb{W}_\theta(\mu,\nu) \tag{2.2}$$

hold for all $\mu,\nu \in \mathscr{P}_\theta$ and $(t,x) \in [0,T] \times \mathbb{R}^d, \xi \in \mathscr{C}$.

(2) $\sigma(t,x,\mu)$ is uniformly continuous in $x \in \mathbb{R}^d$ uniformly with respect to $(t,\mu) \in [0,T] \times \mathscr{P}$. There exist $K > 1$, $(p,q) \in \mathscr{K}$ and nonnegative $F \in L_p^q(T)$ such that

$$|b(t,x,\mu)|^2 \leq F(t,x) + K, \quad K^{-1}I \leq (\sigma\sigma^*)(t,x,\mu) \leq KI$$

for all $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}$.

(3) B is bounded. Moreover, there exists a constant $L_0 > 0$ such that

$$|B(t,\xi,\mu) - B(t,\bar{\xi},\mu)| \leq L_0\|\xi - \bar{\xi}\|_{\mathscr{C}}, \quad t \in [0,T], \xi, \bar{\xi} \in \mathscr{C}, \mu \in \mathscr{P}_\theta. \tag{2.3}$$

Recall that a continuous function f on $\mathbb{R}^d$ is called weakly differentiable, if there exists (hence unique) $h \in L^1_{loc}(\mathbb{R}^d)$ such that

$$\int_{\mathbb{R}^d} (f\Delta g)(x)\mathrm{d}x = -\int_{\mathbb{R}^d} \langle h, \nabla g\rangle(x)\mathrm{d}x, \quad g \in C_0^\infty(\mathbb{R}^d).$$

In this case, we write $h = \nabla f$ and call it the weak gradient of f.

The main result in this part is the following theorem.

Theorem 2.1. *Assume (H^θ) for some constant $\theta \geq 1$. Then the following assertions hold.*

(1) *For any $\mu_0 \in \mathscr{P}_\theta$, the SDE (1.1) has a weak solution $(\tilde{X}, \tilde{W})$ on some complete filtration probability space $\{\tilde{\Omega}, \{\tilde{\mathscr{F}}_t\}_{t\in[0,T]}, \tilde{\mathbb{P}}\}$ with initial distribution μ_0 satisfying $\mathscr{L}_{\tilde{X}_\cdot}|\tilde{\mathbb{P}} \in C([0,T];\mathscr{P}_\theta)$.*

(2) *Let X_0 be an $\mathscr{F}_0$-measurable $\mathscr{C}$-valued random variable with $\mathscr{L}_{X_0} \in \mathscr{P}_\theta$. If in addition, for any $\mu(\cdot) \in C([0,T];\mathscr{P}_\theta)$ and $(t,x) \in [0,T] \times \mathbb{R}^d$, $b^\mu(t,x) := b(t,x,\mu_t)$ and $\sigma^\mu(t,x) := \sigma(t,x,\mu_t)$ satisfy $|b^\mu|^2 + \|\nabla\sigma^\mu\|^2 \in L_p^q(T)$ for some $(p,q) \in \mathscr{K}$, where ∇ is the weak gradient in the space variable $x \in \mathbb{R}^d$, then the SDE (1.1) has a unique strong solution with initial value X_0 satisfying $\mathscr{L}_{X_\cdot} \in C([0,T];\mathscr{P}_\theta)$.*

When B, b and σ do not depend on the distribution, Theorem 2.1 reduces back to the corresponding results derived for classical functional SDEs with singular coefficients and bounded B, see for instance [1] and references within.

2.2 Harnack Inequality

In this subsection, we investigate the dimension-free Harnack inequality introduced in [15] for SDE (1.1), see [20] and references within for general results on these type Harnack inequalities and applications. We establish Harnack inequalities using coupling by change of measures (see for instance [20, §1.1]).

To characterize the singularity of $b(t,x,\mu)$ with respect to x, we introduce

$$\mathscr{D} = \left\{ \phi : [0,\infty) \to [0,\infty) \text{ is increasing}, \phi^2 \text{ is concave}, \int_0^1 \frac{\phi(s)}{s} \mathrm{d}s < \infty \right\}.$$

Remark 2.2. *The condition $\int_0^1 \frac{\phi(s)}{s}\mathrm{d}s < \infty$ is known as the Dini condition. Obviously, $\mathscr{D}$ contains $\phi(s) = s^\alpha$ for any $\alpha \in (0, \frac{1}{2})$. Moreover, it also contains $\phi(s) := \frac{1}{\log^{1+\delta}(c+s^{-1})}$ for constants $\delta > 0$ and large enough $c > 0$ such that ϕ^2 is concave.*

Let $\|\cdot\|_{HS}$ denote the usual Hilbert-Schmidt norm of a matrix. We will need the following assumption.

(H) For any $t \in [0,T], x \in \mathbb{R}^d, \xi \in \mathscr{C}$, $b(t,x,\cdot)$, $B(t,\xi,\cdot)$ and $\sigma(t,x,\cdot)$ are continuous in $\mathscr{P}$. $\|b\|_\infty + \|B\|_\infty < \infty$ and there exist a constant $K > 1$ and $\phi \in \mathscr{D}$ such that

$$K^{-1} I \le (\sigma\sigma^*)(t,x,\mu) \le KI, \quad t \in [0,T], x \in \mathbb{R}^d, \mu \in \mathscr{P},$$

and for any $t \in [0,T]$, $x, y \in \mathbb{R}^d$, and $\mu, \nu \in \mathscr{P}_2$, $\xi, \bar{\xi} \in \mathscr{C}$,

$$\|\sigma(t,x,\mu) - \sigma(t,y,\nu)\|_{HS}^2 \le K(|x-y|^2 + \mathbb{W}_2(\mu,\nu)^2),$$

$$|b(t,x,\mu) - b(t,y,\nu)| \le \phi(|x-y|) + K\mathbb{W}_2(\mu,\nu),$$

$$|B(t,\xi,\mu) - B(t,\bar{\xi},\nu)| \le K(\|\xi - \bar{\xi}\|_{\mathscr{C}} + \mathbb{W}_2(\mu,\nu)).$$

According to Lemma 5.1 below, (1.1) is well-posed under **(H)**. Let $X_t(\mu_0)$ solve (1.1) with $\mathscr{L}_{X_0} = \mu_0$, and $P_t^*\mu_0$ be the distribution of $X_t(\mu_0)$. Define

$$(P_t f)(\mu_0) = \int_{\mathscr{C}} f \mathrm{d}(P_t^*\mu_0) = \mathbb{E} f(X_t(\mu_0)), \quad f \in \mathscr{B}_b(\mathscr{C}), t \in [0,T], \mu_0 \in \mathscr{P}_2.$$

Theorem 2.3. *Assume* **(H)** *and that $\sigma(t,x,\mu)$ does not depend on μ. Then the following assertions hold.*

(1) *There exists a constant* $C > 0$ *such that*

$$(P_t \log f)(\nu_0) \le \log(P_t f)(\mu_0) + \frac{C}{(t-r)\wedge 1}\mathbb{W}_2(\mu_0,\nu_0)^2$$

for any $t \in (r,T], \mu_0, \nu_0 \in \mathscr{P}_2, f \in \mathscr{B}_b^+(\mathscr{C})$ *with* $f \ge 1$. *Consequently, for any different* $\mu_0, \nu_0 \in \mathscr{P}_2$, $t \in (r,T]$, *and any* $f \in \mathscr{B}_b(\mathscr{C})$, *it holds*

$$\frac{|(P_t f)(\mu_0) - (P_t f)(\nu_0)|^2}{\mathbb{W}_2(\mu_0,\nu_0)^2} \le \frac{2C}{(t-r)\wedge 1} \sup_{\nu \in \mathbf{B}(\mu_0, \mathbb{W}_2(\mu_0,\nu_0))} \big\{(P_t f^2)(\nu) - (P_t f)^2(\nu)\big\},$$

where $\mathbf{B}(\mu_0, \mathbb{W}_2(\mu_0,\nu_0)) := \{\nu \in \mathscr{P}_2, \mathbb{W}_2(\mu_0,\nu) < \mathbb{W}_2(\mu_0,\nu_0)\}$.

(2) *There exist constants* $p_0 > 1$, *such that for any* $p > p_0$, $t \in (r,T], f \in \mathscr{B}_b^+(\mathscr{C})$ *and* $\mu_0, \nu_0 \in \mathscr{P}_2$,

$$(P_t f)^p(\nu_0) \le (P_t f^p)(\mu_0)\Big(\mathbb{E}\mathrm{e}^{H_2(p,t)\left(1+\frac{|X(0-Y(0)|^2}{t-r}+\|X_0-Y_0\|_{\mathscr{C}}^2\right)}\Big)^p \tag{2.4}$$

holds for $H_2 : (p_0,\infty)\times(r,T] \to (0,\infty)$ *and* $\mathscr{F}_0$*-measurable* $\mathscr{C}$*-valued random variables* X_0, Y_0 *satisfying* $\mathscr{L}_{X_0} = \mu_0$, $\mathscr{L}_{Y_0} = \nu_0$.

The proof of Theorem 2.3 is given in Section 5.1.

2.3 Shift Harnack Inequality

In this section, we establish the shift Harnack inequality introduced in [19] for P_t. To this end, we assume that $\sigma(t,x,\mu)$ does not depend on x. So SDE (1.1) becomes

$$\mathrm{d}X(t) = B(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + b(t, X(t), \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, \mathscr{L}_{X_t})\mathrm{d}W(t), \quad t \in [0,T]. \tag{2.5}$$

Theorem 2.4. *Assume* **(H)** *and that* $\sigma(t,x,\mu)$ *does not depend on* x. *Then for any* $p > 1, t \in (r,T], \mu_0 \in \mathscr{P}_2, \eta \in C^1([-r,0],\mathbb{R}^d)$ *and* $f \in \mathscr{B}_b^+(\mathscr{C})$, *it holds*

$$(P_t f)^p(\mu_0) \le (P_t f^p(\eta + \cdot))(\mu_0) \times \exp\left[\frac{p}{2(p-1)}\beta(T,\eta,r)\right],$$

where

$$\beta(T,\eta,r) = C\frac{|\eta(-r)|^2}{T-r} + C\int_{-r}^0 |\eta'(s)|^2\mathrm{d}s + CT\phi^2\left(C\|\eta\|_{\mathscr{C}}\right) + CT\|\eta\|_{\mathscr{C}}^2,$$

and $C > 0$ *is a constant. Moreover, for any* $f \in \mathscr{B}_b^+(\mathscr{C})$ *with* $f \ge 1$,

$$(P_t \log f)(\mu_0) \le \log(P_t f(\eta + \cdot))(\mu_0) + \beta(T,\eta,r)$$

holds.

The proof of Theorem 2.4 will be given in Section 5.2.

3 Preparations

We first recall Krylov's estimate of SDEs.

Definition 3.1 (Krylov's Estimate). *An $\mathscr{F}_t$-adapted process $\{X(s)\}_{0\le s\le T}$ is said to satisfy Krylov's estimate, if for any $(p,q)\in\mathscr{K}$, there exist constants $\delta\in(0,1)$ and $C>0$ such that for any non-negative measurable function f on $[0,T]\times\mathbb{R}^d$,*

$$\mathbb{E}\bigg(\int_s^t f(r,X(r))\mathrm{d}r\bigg|\mathscr{F}_s\bigg)\le C(t-s)^\delta\|f\|_{L_p^q(T)},\quad 0\le s\le t\le T. \tag{3.1}$$

We note that (3.1) implies the following Khasminskii type estimate, see for instance [24, Lemma 3.5] and its proof: there exists a constant $c>0$ such that for any $n\ge1$,

$$\mathbb{E}\bigg(\bigg(\int_s^t f(r,X(r))\mathrm{d}r\bigg)^n\bigg|\mathscr{F}_s\bigg)\le cn!(t-s)^{\delta n}\|f\|_{L_p^q(T)}^n,\quad 0\le s\le t\le T, \tag{3.2}$$

and for any $\lambda>0$ there exists a constant $\Lambda=\Lambda(\lambda,\delta,c)>0$ such that

$$\mathbb{E}\big(\mathrm{e}^{\lambda\int_0^T f(r,X(r))\mathrm{d}r}\big|\mathscr{F}_s\big)\le\mathrm{e}^{\Lambda\big(1+\|f\|_{L_p^q(T)}\big)},\quad s\in[0,T]. \tag{3.3}$$

We first present a new result on Krylov's estimate, then recall one lemma from [12] for the construction of weak solution, and finally introduce another lemma on the relation between existence of strong and weak solutions.

3.1 Krylov's Estimate

Consider the following SDE on $\mathbb{R}^d$:

$$\mathrm{d}X(t)=B(t,X_t)\mathrm{d}t+b(t,X(t))\mathrm{d}t+\sigma(t,X(t))\mathrm{d}W(t),\quad t\in[0,T]. \tag{3.4}$$

Lemma 3.1. *Let $T>0$ and $p,q\in(1,\infty)$ with $\frac{d}{p}+\frac{2}{q}<1$. Assume that $\sigma(t,x)$ is uniformly continuous in $x\in\mathbb{R}^d$ uniformly with respect to $t\in[0,T]$, and that for a constant $K>1$ and some non-negative function $F\in L_p^q(T)$ such that*

$$K^{-1}I\le(\sigma\sigma^*)(t,x)\le KI,\quad (t,x)\in[0,T]\times\mathbb{R}^d,$$

$$|b(t,x)|\le K+F(t,x),\quad (t,x)\in[0,T]\times\mathbb{R}^d.$$

$$|B(t,\xi)|\le K,\quad (t,\xi)\in[0,T]\times\mathscr{C}.$$

Then for any $(\alpha,\beta)\in\mathscr{K}$, there exist constants $C=C(\delta,K,\alpha,\beta,\|F\|_{L_p^q(T)})>0$ and $\delta=\delta(\alpha,\beta)\in(0,1)$, such that for any $s_0\in[0,T)$, $\mathscr{F}_{s_0}$-measurable and $\mathscr{C}$-valued random variable X_{s_0} and any solution $(X(s_0;t))_{t\in[s_0,T]}$ of (3.4) with initial value X_{s_0} and initial time s_0, it holds

$$\mathbb{E}\bigg[\int_s^t|f|(v,X(s_0;v))\mathrm{d}v\bigg|\mathscr{F}_s\bigg]\le C(t-s)^\delta\|f\|_{L_\alpha^\beta(T)},\ s_0\le s<t\le T, f\in L_\alpha^\beta(T).$$

Proof. Let $\{X_s(s_0)\}_{s\in[s_0,T]}$ be the segment process of $X(s_0;s)$ and

$$\tilde{W}(\cdot) = W(\cdot) + \int_{s_0}^{\cdot} B(v, X_v(s_0))\mathrm{d}v.$$

Since B is bounded, by Girsanov's theorem, $\tilde{W}$ is a d-dimensional Brownian motion on $[0,T]$ under $\mathbb{Q} = R(T)\mathbb{P}$, where

$$R(T) = \exp\bigg[-\int_{s_0}^{T} \langle B(v, X_v(s_0)), \mathrm{d}W(v)\rangle - \frac{1}{2}\int_{s_0}^{T} |B(v, X_v(s_0))|^2\mathrm{d}v\bigg].$$

Moreover, the boundedness of B implies $\mathbb{E}R(T)^{-1} < \infty$. Thus, under probability measure $\mathbb{Q}$, $(\{X(s_0;v)\}_{v\in[s_0,T]}, \tilde{W})$ is a weak solution to the SDE

$$\mathrm{d}X(t) = b(t, X(t))\mathrm{d}t + \sigma(t, X(t))\mathrm{d}W(t).$$

By [10, Lemma 3.1], there exist constants $C = C(\delta, K, \alpha, \beta, \|F\|_{L_p^q(T)}) > 0$ and $\delta = \delta(\alpha, \beta) \in (0,1)$ such that

$$\mathbb{E}^{\mathbb{Q}}\bigg[\int_s^t |f|(v, X(s_0;v))\mathrm{d}v\bigg|\mathscr{F}_s\bigg] \le C(t-s)^{\delta}\|f\|_{L_\alpha^\beta(T)},\ s_0 \le s < t \le T, f \in L_\alpha^\beta(T).$$

This together with (3.2) and Hölder's inequality implies

$$\begin{aligned}
&\bigg(\mathbb{E}\bigg[\int_s^t |f|(v, X(s_0;v))\mathrm{d}v\bigg|\mathscr{F}_s\bigg]\bigg)^2 = \mathbb{E}R(T)^{-1} \times \mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_s^t |f|(v, X(s_0;v))\mathrm{d}v\bigg)^2\bigg|\mathscr{F}_s\bigg]\\
&\le C\mathbb{E}^{\mathbb{Q}}\bigg[\bigg(\int_s^t |f|(v, X(s_0;v))\mathrm{d}v\bigg)^2\bigg|\mathscr{F}_s\bigg]\\
&\le C(t-s)^{2\delta}\|f\|^2_{L_\alpha^\beta(T)},\ s_0 \le s < t \le T, , f \in L_\alpha^\beta(T).
\end{aligned}$$

Then the proof is finished. □

3.2 Convergence of Stochastic Processes

To prove Theorem 2.1(1), we will use the following lemma due to [12, Theorem 4.3].

Lemma 3.2. *Let $\{\{\psi_t^n\}_{t\in[0,T]}\}_{n\ge1}$ be a sequence of d-dimensional continuous processes on $(\Omega, \mathscr{F}, \mathbb{P})$. Assume that there exist constants $\alpha, \beta > 0$ such that*

$$\sup_{n\ge1}\sup_{t\in[0,T]} \mathbb{E}|\psi_t^n|^\alpha < \infty, \tag{3.5}$$

and there exists a constant $M_T > 0$ such that

$$\sup_{n\ge1}\mathbb{E}|\psi_t^n - \psi_s^n|^\alpha \le M_T|t-s|^{1+\beta},\quad t, s \in [0,T]. \tag{3.6}$$

Then there exist a subsequence $\{n_k\}_{k\geq 1}$, a probability space $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ and d-dimensional continuous processes $\{X_t\}_{t\in[0,T]}, \{\{X_t^k\}_{t\in[0,T]}\}_{k\geq 1}$, such that $\mathscr{L}_{\psi^{n_k}}|\mathbb{P} = \mathscr{L}_{X^k}|\tilde{\mathbb{P}}$, and $\tilde{\mathbb{P}}$-a.s. X^k converges to X as $k\to\infty$.

Proof. By [12, Theorem 4.2, Theorem 4.3], (3.5) and (3.6) imply that $\{\psi^n\}_{n\geq 1}$ is tight. Then there exists a subsequence $\{m_l\}_{l\geq 1}$ such that $\{\psi^{m_l}\}_{l\geq 1}$ is weakly convergent. For $\{\psi^{m_l}\}_{l\geq 1}$, by [12, Theorem 4.3], there exists a subsequence $\{n_k\}_{k\geq 1}$ of $\{m_l\}_{l\geq 1}$, a probability space $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ and stochastic processes $\{X_t\}_{t\in[0,T]}, \{\{X_t^k\}_{t\in[0,T]}\}_{k\geq 1}$, such that $\mathscr{L}_{\psi^{n_k}}|\mathbb{P} = \mathscr{L}_{X^k}|\tilde{\mathbb{P}}$, and $\tilde{\mathbb{P}}$-a.s. X^k converges to X as $k\to\infty$. The proof is completed. □

3.3 Relation between Existence of Strong and Weak Solutions

We present a result on the existence of strong solutions deduced from weak solutions. Consider the following SDE

$$\mathrm{d}X(t) = \hat B(t, X_t, \mathscr{L}_{X_t})\,\mathrm{d}t + \hat\Sigma(t, X_t, \mathscr{L}_{X_t})\,\mathrm{d}W(t), \quad 0\leq t\leq T, \tag{3.7}$$

where $\hat B : [0,T]\times\mathscr{C}\times\mathscr{P}\to\mathbb{R}^d$ and $\hat\Sigma : [0,T]\times\mathscr{C}\times\mathscr{P}\to\mathbb{R}^d\otimes\mathbb{R}^d$ are measurable.

Lemma 3.3. *Let $(\bar\Omega, \bar{\mathscr{F}}_t, (\bar X_t, \bar W(t)), \bar{\mathbb{P}})$ be a weak solution to (3.7) with $\mu_t := \mathscr{L}_{\bar X_t}|\bar{\mathbb{P}}$. If the SDE*

$$\mathrm{d}X(t) = \hat B(t, X_t, \mu_t)\,\mathrm{d}t + \hat\Sigma(t, X_t, \mu_t)\,\mathrm{d}W(t), \quad 0\leq t\leq T \tag{3.8}$$

has a unique strong solution X_t up to life time with $\mathscr{L}_{X_0} = \mu_0$, then (3.7) has a strong solution with initial value X_0.

Proof. Since $\mu_t = \mathscr{L}_{\bar X_t}|\bar{\mathbb{P}}$, $\bar X_t$ is a weak solution to (3.8). By Yamada-Watanabe principle, the strong uniqueness of (3.8) implies the weak uniqueness, so that X_t is nonexplosive with $\mathscr{L}_{X_t} = \mu_t, t\geq 0$. Therefore, X_t is a strong solution to (3.7). □

With the above preparations in hand, we are now in the position to prove Theorem 2.1.

4 Proof of Theorem 2.1

4.1 Proof of Theorem 2.1(1)

Set $a(t,x,\mu) := (\sigma\sigma^*)(t,x,\mu)$ for $t\in[0,T]$. Define $b(t,x,\mu) := 0$, $a(t,x,\mu) := I$ and $F(t,x) := 0$ for $t\in\mathbb{R}\setminus[0,T]$. Let $0\leq\rho\in C_0^\infty(\mathbb{R}\times\mathbb{R}^d)$ with support contained in $\{(s,x) : |(s,x)|\leq 1\}$ such that $\int_{\mathbb{R}\times\mathbb{R}^d}\rho(s,x)\mathrm{d}s\mathrm{d}x = 1$. For any $n\geq 1$, let $\rho_n(s,x) = n^{d+1}\rho(ns, nx)$ and define

$$a^n(t,x,\mu) = \int_{\mathbb{R}\times\mathbb{R}^d} a(s,x',\mu)\rho_n(t-s, x-x')\mathrm{d}s\mathrm{d}x',$$

$$
\begin{aligned}
(4.1)\qquad b^n(t,x,\mu) &= \int_{\mathbb{R}\times\mathbb{R}^d} b(s,x',\mu)\rho_n(t-s,x-x')\mathrm{d}s\mathrm{d}x',\\
F^n(t,x) &= \int_{\mathbb{R}\times\mathbb{R}^d} F(s,x')\rho_n(t-s,x-x')\mathrm{d}s\mathrm{d}x', \quad (t,x,\mu)\in\mathbb{R}\times\mathbb{R}^d\times\mathscr{P}.
\end{aligned}
$$

Let $\hat{\sigma}^n = \sqrt{a^n}$ and $\hat{\sigma} = \sqrt{a}$. Consider the following SDE:

$$
(4.2)\qquad \mathrm{d}X(t) = b(t,X(t),\mathscr{L}_{X_t})\mathrm{d}t + B(t,X_t,\mathscr{L}_{X_t})\mathrm{d}t + \hat{\sigma}(t,X(t),\mathscr{L}_{X_t})\mathrm{d}W(t).
$$

According to [8, Proof of Theorem 2.1-2.3], if (4.2) has a weak solution $(\tilde{X}_t, \tilde{W}(t))$ on a probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$, then $\mu_t := \mathscr{L}_{\tilde{X}_t}|\tilde{\mathbb{P}}$ is a martingale solution to Fokker–Planck equation

$$
\partial_t\mu(t) = (L^{\hat{\sigma}}_{t,\mu_t})^*\mu_t,
$$

where $\mu(t)$ is the marginal distribution of μ_t at $v=0$; i.e.

$$
\{\mu(t)\}(\mathrm{d}x) := \mu_t(\{\xi\in\mathscr{C} : \xi(0)\in\mathrm{d}x\}),
$$

and for any $f\in C_0^\infty(\mathbb{R}^d)$,

$$
\begin{aligned}
[(L^{\hat{\sigma}}_{t,\mu_t})^*\mu_t](f) := &\int_{\mathscr{C}} \frac{1}{2}\sum_{i,j=1}^d (\hat{\sigma}\hat{\sigma}^*)_{ij}(t,\xi(0),\mu_t)(\partial_i\partial_j f)(\xi(0))\mu_t(\mathrm{d}\xi)\\
&+ \int_{\mathscr{C}} \sum_{i=1}^d [b_i(t,\xi(0),\mu_t) + B_i(t,\xi,\mu_t)](\partial_i f)(\xi(0))\mu_t(\mathrm{d}\xi).
\end{aligned}
$$

Noting that $\sigma\sigma^* = \hat{\sigma}\hat{\sigma}^*$, μ_t is also a martingale solution to

$$
\partial_t\mu(t) = (L^{\sigma}_{t,\mu_t})^*\mu_t.
$$

This together with [8, Proof of Theorem 2.1-2.3] implies that (1.1) has a weak solution. Thus, in order to prove that (1.1) has a weak solution, it is sufficient to prove the same claim for SDE (4.2), which will be completed according to the following procedure.

As in [7, Proof of Theorem 2.1], there exist subsequence $\{n_k\}_{k\geq 1}$ and $G\in L^q_p(T)$ such that

$$
(4.3)\qquad |b^{n_k}|^2 \leq K + G, \quad k\geq 1.
$$

In fact, for any $k\geq 1$, we can choose $n_k\geq 1$ such that

$$
\|F - F^{n_k}\|_{L^q_p(T)} \leq \frac{1}{2^k}.
$$

Taking $G=\sum_{k=1}^\infty |F-F^{n_k}|+F$, we have $\|G\|_{L^q_p(T)}\le 1+\|F\|_{L^q_p(T)}$. Moreover, Jensen's inequality, $(H^\theta)(2)$ and (4.1) imply that

$$\begin{aligned}|b^{n_k}|^2(t,x,\mu)&\le \int_{\mathbb{R}\times\mathbb{R}^d} b^2(s,x',\mu)\rho_{n_k}(t-s,x-x')\mathrm{d}s\mathrm{d}x'\\&\le K+F^{n_k}(t,x,\mu)\le K+G(t,x,\mu),\quad k\ge 1.\end{aligned}$$

Below, we use the subsequence b^{n_k} replacing b^n. For simplicity, we still denote b^{n_k} by b^n. Moreover, it follows from $(H^\theta)(2)$ and (4.1) that $\hat\sigma^n(t,x,\mu)$ is uniformly continuous in $x\in\mathbb{R}^d$ uniformly with respect to $(t,\mu)\in[0,T]\times\mathscr{P}$ and

$$K^{-1}I\le \hat\sigma^n(\hat\sigma^n)^*\le KI,\quad (t,x,\mu)\in[0,T]\times\mathbb{R}^d\times\mathscr{P}. \tag{4.4}$$

According to (4.1), (2.1) and $(H^\theta)(2)$, for any $n\ge 1$ there exists a constant $c_n>0$ such that

$$\begin{aligned}&|b^n(t,x,\mu)-b^n(s,x',\nu)|+\|\hat\sigma^n(t,x,\mu)-\hat\sigma^n(s,x',\nu)\|\\&\le c_n\big(|t-s|+|x-x'|\big)+(KL+L)\,\mathbb{W}_\theta(\mu,\nu)\end{aligned}$$

holds for all $s,t\in\mathbb{R}, x,x'\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_\theta$. This combined with (2.2) and (2.3) implies that the SDE

$$\mathrm{d}X^n(t)=B(t,X^n_t,\mathscr{L}_{X^n_t})\mathrm{d}t+b^n(t,X^n(t),\mathscr{L}_{X^n_t})\mathrm{d}t+\hat\sigma^n(t,X^n(t),\mathscr{L}_{X^n_t})\mathrm{d}W(t) \tag{4.5}$$

with $X^n_0=X_0$ has a unique strong solution $(X^n_t)_{t\in[0,T]}$. In fact, this is standard by repeating the proof of [8, Theorem 3.1(1)], where $\theta=2$ is considered. Applying Lemma 3.1 for X^n, we derive that for any $(\alpha,\beta)\in\mathscr{K}$,

$$\mathbb{E}\left(\int_s^t |f|(v,X^n(v))\mathrm{d}v\middle|\mathscr{F}_s\right)\le C(t-s)^\delta\|f\|_{L^\beta_\alpha(T)},\quad 0\le f\in L^\beta_\alpha(T), n\ge 1 \tag{4.6}$$

holds for some constants $C>0$ and $\delta\in(0,1)$.

We first show that Lemma 3.2 holds for (X^n,W) replacing ψ^n, for which it suffices to verify conditions (3.5) and (3.6) with $\psi^n:=X^n$. By (4.3), (4.4), $(H^\theta)(3)$ and (3.2) for X^n replacing X implied by (4.6), there exist constants $c_1,c_2>0$ such that

$$\begin{aligned}\mathbb{E}|X^n(t)|^\theta\le c_1\Bigg\{&\mathbb{E}|X(0)|^\theta+\mathbb{E}\bigg(\int_0^T|b^n(t,X^n(t),\mathscr{L}_{X^n_t})|\,\mathrm{d}t\bigg)^\theta\\&+\mathbb{E}\bigg(\int_0^T|B(t,X^n_t,\mathscr{L}_{X^n_t})|\,\mathrm{d}t\bigg)^\theta\\&+\mathbb{E}\bigg(\int_0^T\|\hat\sigma^n(t,X^n(t),\mathscr{L}_{X^n_t})\|^2\,\mathrm{d}t\bigg)^{\frac{\theta}{2}}\Bigg\}\end{aligned}$$

$$\le c_2\Big(\mathbb{E}|X(0)|^\theta + T^\theta + T^{\delta\theta}\|G\|^\theta_{L^q_p(T)} + T^{\frac{\theta}{2}}\Big) < \infty, \quad n \ge 1, t \in [0,T].$$

Thus, (3.5) holds for $\psi^n = X^n$.

Next, let $\delta_0 = \frac{2}{\frac{1}{2}\wedge\delta}$ by the same reason, there exists a constant $c_3 > 0$ such that for any $0 \le s \le t \le T$,

$$\begin{aligned}
&\mathbb{E}|X^n(t) - X^n(s)|^{\delta_0}\\
&\le c_3\Bigg\{\mathbb{E}\left(\int_s^t |b^n(v, X^n(v), \mathscr{L}_{X^n_v})|\,\mathrm{d}v\right)^{\delta_0} + \mathbb{E}\left(\int_s^t |B(v, X^n_v, \mathscr{L}_{X^n_v})|\,\mathrm{d}v\right)^{\delta_0}\\
&+ \mathbb{E}\left(\int_s^t \|\hat\sigma^n(v, X^n(v), \mathscr{L}_{X^n_v})\|^2\,\mathrm{d}v\right)^{\frac{\delta_0}{2}}\Bigg\}\\
&\le c_3\big((t-s)^{\delta_0} + (t-s)^{\delta\delta_0}\|G\|^{\delta_0}_{L^q_p(T)} + (t-s)^{\frac{\delta_0}{2}}\big).
\end{aligned}$$

Hence, (3.6) holds for $\psi^n = X^n$. According to Lemma 3.2, there exists a subsequence of $(X^n, W)_{n\ge1}$, denoted again by $(X^n, W)_{n\ge1}$, stochastic processes $(\tilde X^n, \tilde W^n)_{n\ge1}$ and $(\tilde X, \tilde W)$ on a complete probability space $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ such that $\mathscr{L}_{(X^n,W)}|\mathbb{P} = \mathscr{L}_{(\tilde X^n,\tilde W^n)}|\tilde{\mathbb{P}}$ for any $n \ge 1$, and $\tilde{\mathbb{P}}$-a.s. $\lim_{n\to\infty}(\tilde X^n, \tilde W^n) = (\tilde X, \tilde W)$. As in [7, Proof of Theorem 2.2], letting $\tilde{\mathscr{F}}^n_t$ and $\tilde{\mathscr{F}}_t$ be the completions of the σ-algebra generated by $\{\tilde X^n(s), \tilde W^n(s) : s \le t\}$ and $\{\tilde X(s), \tilde W(s) : s \le t\}$ respectively, $\tilde X^n(t)$ is $\tilde{\mathscr{F}}^n_t$-adapted and continuous (since X^n is continuous and $\mathscr{L}_{X^n}|\mathbb{P} = \mathscr{L}_{\tilde X^n}|\tilde{\mathbb{P}}$), $\tilde W^n$ is a d-dimensional Brownian motion on $(\tilde\Omega, \{\tilde{\mathscr{F}}^n_t\}_{t\in[0,T]}, \tilde{\mathbb{P}})$, and due to (4.5), $(\tilde X^n(t), \tilde W^n(t))_{t\in[0,T]}$ solves the SDE

$$\begin{aligned}
\mathrm{d}\tilde X^n(t) &= b^n(t, \tilde X^n(t), \mathscr{L}_{\tilde X^n_t}|\tilde{\mathbb{P}})\,\mathrm{d}t\\
&\quad + B(t, \tilde X^n_t, \mathscr{L}_{\tilde X^n_t}|\tilde{\mathbb{P}})\,\mathrm{d}t + \hat\sigma^n(t, \tilde X^n(t), \mathscr{L}_{\tilde X^n_t}|\tilde{\mathbb{P}})\,\mathrm{d}\tilde W^n(t)
\end{aligned} \tag{4.7}$$

with $\mathscr{L}_{\tilde X^n_0}|\tilde{\mathbb{P}} = \mathscr{L}_{X_0}|\mathbb{P}$. Simply denote $\mathscr{L}_{\tilde X^n_t}|\tilde{\mathbb{P}} = \mathscr{L}_{\tilde X^n_t}$ and $\mathscr{L}_{\tilde X_t}|\tilde{\mathbb{P}} = \mathscr{L}_{\tilde X_t}$ and let $\tilde{\mathbb{E}}$ be the expectation under $\tilde{\mathbb{P}}$.

For any $n \ge 1$ and $s \in [0,T]$, we have

$$\left|\int_0^s B(t, \tilde X^n_t, \mathscr{L}_{\tilde X^n_t})\mathrm{d}t - \int_0^s B(t, \tilde X_t, \mathscr{L}_{\tilde X_t})\mathrm{d}t\right| \le I_1(s) + I_2(s),$$

where

$$\begin{aligned}
I_1(s) &:= \left|\int_0^s B(t, \tilde X^n_t, \mathscr{L}_{\tilde X^n_t})\mathrm{d}t - \int_0^s B(t, \tilde X_t, \mathscr{L}_{\tilde X^n_t})\mathrm{d}t\right|,\\
I_2(s) &:= \left|\int_0^s B(t, \tilde X_t, \mathscr{L}_{\tilde X^n_t})\mathrm{d}t - \int_0^s B(t, \tilde X_t, \mathscr{L}_{\tilde X_t})\mathrm{d}t\right|.
\end{aligned}$$

Below we estimate $I_i(s), i = 1,2$ respectively.

Firstly, for any $\varepsilon > 0$, by Markov's inequality, we arrive at

$$\tilde{\mathbb{P}}(\sup_{s\in[0,T]} I_1(s) \geq \varepsilon) \leq \frac{1}{\varepsilon}\tilde{\mathbb{E}}\int_0^T \left|B(t,\tilde{X}_t^n,\mathscr{L}_{\tilde{X}_t^n}) - B(t,\tilde{X}_t,\mathscr{L}_{\tilde{X}_t^n})\right| \mathrm{d}t.$$

Since $\tilde{\mathbb{P}}$-a.s. $\tilde{X}_t^n$ converges to $\tilde{X}_t$, by (2.3) and the boundedness of B, we may apply the dominated convergence theorem to derive

$$\limsup_{n\to\infty} \tilde{\mathbb{P}}(\sup_{s\in[0,T]} I_1(s) \geq \varepsilon) \leq \frac{1}{\varepsilon}\tilde{\mathbb{E}}\int_0^T \lim_{n\to\infty} L_0\|\tilde{X}_t^n - \tilde{X}_t\|_{\mathscr{C}}\mathrm{d}t = 0.$$

Furthermore, since for any $t\in[0,T], \xi\in\mathscr{C}$, $B(t,\xi,\cdot)$ is continuous in $\mathscr{P}$ due to $(H^\theta)(1)$, and $\tilde{X}_t^n$ converges to $\tilde{X}_t$ weakly in $\mathscr{P}$, it is not difficult to see from Markov's inequality and the dominated convergence theorem that

$$\begin{aligned}&\limsup_{n\to\infty} \tilde{\mathbb{P}}\Big(\sup_{s\in[0,T]} I_2(s) \geq \varepsilon\Big)\\ &\leq \limsup_{n\to\infty} \frac{1}{\varepsilon}\int_0^T \tilde{\mathbb{E}}\left|B(t,\tilde{X}_t,,\mathscr{L}_{\tilde{X}_t^n}) - B(t,\tilde{X}_t,\mathscr{L}_{\tilde{X}_t})\right| \mathrm{d}t = 0.\end{aligned}$$

Thus, for any $\varepsilon > 0$, we have

$$\lim_{n\to\infty} \tilde{\mathbb{P}}\left(\sup_{s\in[0,T]}\left|\int_0^s B(t,\tilde{X}_t^n,\mathscr{L}_{\tilde{X}_t^n})\mathrm{d}t - \int_0^s B(t,\tilde{X}_t,\mathscr{L}_{\tilde{X}_t})\mathrm{d}t\right| \geq \varepsilon\right) = 0.$$

Similarly to the proof of [10, (4.5)-(4.6)], $(H^\theta)(1)$-(2) imply

$$\lim_{n\to\infty} \tilde{\mathbb{P}}\left(\sup_{s\in[0,T]}\int_0^s |b^n(t,\tilde{X}^n(t),\mathscr{L}_{\tilde{X}_t^n}) - b(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t \geq \varepsilon\right) = 0,$$

and

$$\lim_{n\to\infty} \tilde{\mathbb{P}}\left(\sup_{s\in[0,T]}\left|\int_0^s \hat{\sigma}^n(t,\tilde{X}^n(t),\mathscr{L}_{\tilde{X}_t^n})\mathrm{d}\tilde{W}^n(t) - \int_0^s \hat{\sigma}(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})\,\mathrm{d}\tilde{W}(t)\right| \geq \varepsilon\right) = 0.$$

Then $(\tilde{X}(t),\tilde{W}(t))_{t\in[0,T]}$ is a weak solution to (4.2) by taking limit in (4.7). Thus, (1.1) has a weak solution, and for simplicity, we still denote it by $(\tilde{X}(t),\tilde{W}(t))$ on $(\tilde{\Omega},\tilde{\mathscr{F}}_t,\tilde{\mathbb{P}})$. Let $\mu_\cdot = \mathscr{L}_{\tilde{X}_\cdot}|\tilde{\mathbb{P}}$. According to (H^θ) and Lemma 3.1, (4.6) and (3.2) hold for $(\tilde{X},\tilde{\mathbb{E}})$ replacing $(X^n,\mathbb{E})$ and $(X,\mathbb{E})$ respectively. Combining this with $(H^\theta)(2)$, we get

$$\tilde{\mathbb{E}}\Big(\int_0^T |b(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t\Big)^\theta \leq c\left(T^\theta + T^{\delta\theta}\|F\|_{L_p^q(T)}^\theta\right) \tag{4.8}$$

for some $c > 0$ and $\delta \in (0,1)$. This together with (H^θ)(2)-(3) and the Burkerholder-Davis-Gundy inequality leads to

$$\begin{aligned}\tilde{\mathbb{E}}\sup_{v\in[-r,0]}|\tilde{X}(s+v)|^\theta &\le c\tilde{\mathbb{E}}\|\tilde{X}_0\|_{\mathscr{C}}^\theta + c\tilde{\mathbb{E}}\bigg(\int_0^T |b(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t\bigg)^\theta\\ &\quad + c\tilde{\mathbb{E}}\bigg(\int_0^T |B(t,\tilde{X}_t,\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t\bigg)^\theta + c\tilde{\mathbb{E}}\bigg(\int_0^T \|\sigma(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})\|^2\,\mathrm{d}t\bigg)^{\frac{\theta}{2}}\\ &\le c\mu_0(\|\cdot\|_{\mathscr{C}}^\theta) + c\Big(T^\theta + T^{\delta\theta}\|F\|^\theta_{L^q_p(T)} + T^{\frac{\theta}{2}}\Big),\quad s\in[0,T]\end{aligned}$$

for some $c > 0$ and $\delta \in (0,1)$. Thus, $\mu_s \in \mathscr{P}_\theta, s \in [0,T]$. In the following, we will prove $\mu_\cdot \in C([0,T],\mathscr{P}_\theta)$. Noting that $\mathbb{W}_\theta(\mu_{s'},\mu_s)^\theta \le \tilde{\mathbb{E}}\|\tilde{X}_{s'} - \tilde{X}_s\|^\theta_{\mathscr{C}}, s', s \in [0,T]$, it is sufficient to prove $\lim_{s'\to s}\tilde{\mathbb{E}}\|\tilde{X}_{s'} - \tilde{X}_s\|^\theta_{\mathscr{C}} = 0, s \in [0,T]$. Fix $s \in [0,T]$. For any $s' \in [0,T]$ with $|s'-s| < r$, we arrive at

$$\begin{aligned}&\tilde{\mathbb{E}}\sup_{v\in[-r,0]}|\tilde{X}(s'+v)-\tilde{X}(s+v)|^\theta\\ &\le c\tilde{\mathbb{E}}\sup_{v\in[0,s\wedge s']}|\tilde{X}(|s'-s|+v)-\tilde{X}(v)|^\theta + c\tilde{\mathbb{E}}\sup_{v\in[-|s'-s|,0]}|\tilde{X}(|s'-s|+v)-\tilde{X}(0)|^\theta\\ &+ c\tilde{\mathbb{E}}\sup_{v\in[-|s'-s|,0]}|\tilde{X}(0)-\tilde{X}(v)|^\theta + c\tilde{\mathbb{E}}\sup_{v\in[-r,-|s'-s|]}|\tilde{X}(|s'-s|+v)-\tilde{X}(v)|^\theta\\ &=: J_1+J_2+J_3+J_4.\end{aligned}$$

In what follows, we will prove $\lim_{s'\to s} J_i = 0$, $i = 1,2,3,4$ respectively. Firstly, it is not difficult to see that

$$\begin{aligned}J_1+J_2 &\le c_1\tilde{\mathbb{E}}\sup_{v\in[0,s\wedge s']}\bigg(\int_v^{|s'-s|+v}|b(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t\bigg)^\theta\\ &\quad + c_1\tilde{\mathbb{E}}\sup_{v\in[0,s\wedge s']}\bigg(\int_v^{|s'-s|+v}|B(t,\tilde{X}_t,\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t\bigg)^\theta\\ &\quad + c_1\tilde{\mathbb{E}}\sup_{v\in[0,s\wedge s']}\bigg|\int_v^{|s'-s|+v}\sigma(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})\,\mathrm{d}\tilde{W}(t)\bigg|^\theta\\ &\quad + c_1\tilde{\mathbb{E}}\sup_{v\in[0,|s'-s|]}|\tilde{X}(v)-\tilde{X}(0)|^\theta\\ &\le c_1\tilde{\mathbb{E}}\sup_{0\le v\le v'\le T, v'-v\le|s'-s|}\bigg(\int_v^{v'}|b(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t\bigg)^\theta\\ &\quad + c_1\tilde{\mathbb{E}}\sup_{0\le v\le v'\le T, v'-v\le|s'-s|}\bigg(\int_v^{v'}|B(t,\tilde{X}_t,\mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t\bigg)^\theta\\ &\quad + c_1\tilde{\mathbb{E}}\sup_{0\le v\le v'\le T, v'-v\le|s'-s|}\bigg|\int_v^{v'}\sigma(t,\tilde{X}(t),\mathscr{L}_{\tilde{X}_t})\,\mathrm{d}\tilde{W}(t)\bigg|^\theta =: J_{11}+J_{12}+J_{13}.\end{aligned}$$

Since B is bounded, we obtain $\lim_{s'\to s} J_{12}=0$. Noting that $\tilde{\mathbb{P}}$-a.s. $\int_0^{\cdot} \sigma(t,\tilde X(t),\mathscr{L}_{\tilde X_t})\,\mathrm{d}\tilde W(t)$ is continuous and hence uniformly continuous on $[0,T]$, we derive $\tilde{\mathbb{P}}$-a.s.

$$\lim_{s'-s\to 0}\sup_{0\le v\le v'\le T, v'-v\le |s'-s|}\left|\int_v^{v'}\sigma(t,\tilde X(t),\mathscr{L}_{\tilde X_t})\,\mathrm{d}\tilde W(t)\right|=0.$$

It follows from the Burkerholder-Davis-Gundy inequality and $(H^\theta)(2)$ that

$$\begin{aligned}
&\tilde{\mathbb{E}}\sup_{0\le v\le v'\le T, v'-v\le |s'-s|}\left|\int_v^{v'}\sigma(t,\tilde X(t),\mathscr{L}_{\tilde X_t})\,\mathrm{d}\tilde W(t)\right|^\theta\\
&\le 2^{\theta-1}\tilde{\mathbb{E}}\sup_{0\le v\le v'\le T, v'-v\le |s'-s|}\left|\int_0^{v'}\sigma(t,\tilde X(t),\mathscr{L}_{\tilde X_t})\,\mathrm{d}\tilde W(t)\right|^\theta\\
&\quad+2^{\theta-1}\tilde{\mathbb{E}}\sup_{0\le v\le v'\le T, v'-v\le |s'-s|}\left|\int_0^{v}\sigma(t,\tilde X(t),\mathscr{L}_{\tilde X_t})\,\mathrm{d}\tilde W(t)\right|^\theta\\
&\le 2^\theta\tilde{\mathbb{E}}\sup_{0\le v'\le T}\left|\int_0^{v'}\sigma(t,\tilde X(t),\mathscr{L}_{\tilde X_t})\,\mathrm{d}\tilde W(t)\right|^\theta\\
&\le c\tilde{\mathbb{E}}\left(\int_0^T\|\sigma(t,\tilde X(t),\mathscr{L}_{\tilde X_t})\|^2\,\mathrm{d}t\right)^{\frac{\theta}{2}}\le cT^{\frac{\theta}{2}}
\end{aligned}$$

for some constant $c>0$. So the dominated convergence theorem implies that $\lim_{s'\to s} J_{13}=0$. Moreover, (4.8) yields that $\tilde{\mathbb{P}}$-a.s. $\int_0^{\cdot}|b(t,\tilde X(t),\mathscr{L}_{\tilde X_t})|\,\mathrm{d}t$ is continuous and thus uniformly continuous on $[0,T]$. This together with the dominated convergence theorem implies $\lim_{s'\to s} J_{13}=0$. Consequently, it holds $\lim_{s'\to s}(J_1+J_2)=0$. In addition, it is clear that

$$J_3+J_4\le c\tilde{\mathbb{E}}\sup_{v,v'\in[-r,0],|v'-v|\le|s'-s|}|\tilde X(v')-\tilde X(v)|^\theta.$$

This together with $\tilde{\mathbb{E}}\sup_{v\in[-r,0]}|\tilde X(v)|^\theta<\infty$ due to $\mathscr{L}_{\tilde X_0}|\tilde{\mathbb{P}}=\mu_0\in\mathscr{P}_\theta$, the fact that any component in $\mathscr{C}$ is uniformly continuous on $[-r,0]$ and the dominated convergence theorem leads to $\lim_{s'\to s}(J_3+J_4)=0$. Therefore, we conclude that $\mu_\cdot\in C([0,T];\mathscr{P}_\theta)$.

4.2 Proof of Theorem 2.1(2)

According to [1, Theorem 1.4], the SDE (3.8) for $\hat B=B+b$ and $\hat\Sigma=\sigma$ has a unique strong solution under the conditions in Theorem 2.1(2). So, Theorem 2.1(1) and Lemma 3.3 for $\hat B=B+b$ and $\hat\Sigma=\sigma$ imply that for any $\mathscr{F}_0$-measurable $\mathscr{C}$-valued random variable X_0 with $\mathscr{L}_{X_0}\in\mathscr{P}_\theta$, SDE (1.1) has a strong solution $\{X_t\}_{t\in[0,T]}$ satisfying $\mathscr{L}_{X_\cdot}\in C([0,T];\mathscr{P}_\theta)$. As a result, to prove Theorem 2.1(2), it suffices to prove the uniqueness of strong solutions of (1.1), which will be finished in Lemma 4.2 below by Zvonkin's transform.

Consider the maximal operator:

$$\mathscr{M}h(x) := \sup_{v>0} \frac{1}{|\mathbf{B}(x,v)|} \int_{\mathbf{B}(x,v)} h(y)\mathrm{d}y, \quad h \in L^1_{loc}(\mathbb{R}^d), x \in \mathbb{R}^d,$$

where $\mathbf{B}(x,v) := \{y : |y-x| < v\}$. The next result comes from [3, Appendix A].

Lemma 4.1. *There exists a constant $C > 0$ such that for any continuous and weak differentiable function f,*

$$|f(x) - f(y)| \le C|x-y|(\mathscr{M}|\nabla f|(x) + \mathscr{M}|\nabla f|(y)), \quad \text{a.e. } x, y \in \mathbb{R}^d. \tag{4.9}$$

Moreover, for any $p > 1$, there exists a constant $C_p > 0$ such that

$$\|\mathscr{M}f\|_{L^p} \le C_p \|f\|_{L^p}, \quad f \in L^p(\mathbb{R}^d). \tag{4.10}$$

The following lemma gives the uniqueness of strong solutions of (1.1).

Lemma 4.2. *Assume the conditions in Theorem 2.1(2) hold. Let X and Y be two strong solutions to* (1.1) *with $X_0 = Y_0$ and $\mathscr{L}_{X_0} \in \mathscr{P}_\theta$. Then $\mathbb{P}$-a.s. $X = Y$.*

Proof. Set $\mu_t = \mathscr{L}_{X_t}, \nu_t = \mathscr{L}_{Y_t}, t \in [0,T]$. Let

$$b^\mu(t,x) = b(t,x,\mu_t), \quad \sigma^\mu(t,x) = \sigma(t,x,\mu_t), \quad (t,x) \in [0,T] \times \mathbb{R}^d,$$

and define b^ν, σ^ν in the same way using ν_t replacing μ_t. Then it is clear that

$$\begin{aligned} \mathrm{d}X(t) &= b^\mu(t,X(t))\,\mathrm{d}t + B(t,X_t,\mu_t)\,\mathrm{d}t + \sigma^\mu(t,X(t))\,\mathrm{d}W(t), \\ \mathrm{d}Y(t) &= b^\nu(t,Y(t))\,\mathrm{d}t + B(t,Y_t,\nu_t)\,\mathrm{d}t + \sigma^\nu(t,Y(t))\,\mathrm{d}W(t). \end{aligned} \tag{4.11}$$

For any $\lambda > 0$, consider the following partial differential equation for $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$:

$$\frac{\partial u(t,\cdot)}{\partial t} + \frac{1}{2}\mathrm{Tr}(\sigma^\mu(\sigma^\mu)^*\nabla^2 u)(t,\cdot) + (\nabla_{b^\mu} u)(t,\cdot) + b^\mu(t,\cdot) = \lambda u(t,\cdot), \quad u(T,\cdot) = 0, \tag{4.12}$$

here $(\nabla_{b^\mu} u)$ stands for the direction derivative of u along b^μ defined as

$$(\nabla_{b^\mu} u)(t,x) := \lim_{\varepsilon \to 0} \frac{u(t, x + \varepsilon b^\mu(t,x)) - u(t,x)}{\varepsilon}, \quad t \in [0,T], x \in \mathbb{R}^d.$$

By [25, Theorem 5.1] and [23, Theorem 3.1, (2.5), (3.2)], when λ is large enough, (4.12) has a unique solution $\mathbf{u}^{\lambda,\mu}$ satisfying

$$\|\nabla \mathbf{u}^{\lambda,\mu}\|_\infty \le \frac{1}{5}, \tag{4.13}$$

and

$$\|\nabla^2 \mathbf{u}^{\lambda,\mu}\|_{L^{2q}_{2p}(T)} < \infty. \tag{4.14}$$

Let $\theta^{\lambda,\mu}(t,x) = x + \mathbf{u}^{\lambda,\mu}(t,x)$. By (4.11), (4.12), and using the Itô formula and an approximation technique (see [25, Lemma 4.3] for more details), we derive

$$\begin{aligned} \mathrm{d}\theta^{\lambda,\mu}(t,X(t)) = \lambda \mathbf{u}^{\lambda,\mu}(t,X(t))\mathrm{d}t + \nabla\theta^{\lambda,\mu}(t,X(t))B(t,X_t,\mu_t)\mathrm{d}t \\ + (\nabla\theta^{\lambda,\mu}\sigma^\mu)(t,X(t))\,\mathrm{d}W(t), \end{aligned} \tag{4.15}$$

and

$$\begin{aligned} \mathrm{d}\theta^{\lambda,\mu}(t,Y(t)) = {} & \lambda \mathbf{u}^{\lambda,\mu}(t,Y(t))\mathrm{d}t + (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t,Y(t))\,\mathrm{d}W(t) \\ & + \nabla\theta^{\lambda,\mu}(t,Y(t))B(t,Y_t,\nu_t)\mathrm{d}t \\ & + [\nabla\theta^{\lambda,\mu}(b^\nu - b^\mu)](t,Y(t))\mathrm{d}t \\ & + \frac{1}{2}\mathrm{Tr}[(\sigma^\nu(\sigma^\nu)^* - \sigma^\mu(\sigma^\mu)^*)\nabla^2\mathbf{u}^{\lambda,\mu}](t,Y(t))\mathrm{d}t. \end{aligned} \tag{4.16}$$

Let $\xi_t = \theta^{\lambda,\mu}(t,X(t)) - \theta^{\lambda,\mu}(t,Y(t)), t \in [0,T]$. It follows from (4.13) that

$$|X(t) - Y(t)| \le \frac{5}{4}|\xi_t|, \quad t \in [0,T]. \tag{4.17}$$

By (4.15), (4.16) and Itô's formula, we obtain

$$\begin{aligned} \mathrm{d}|\xi_t|^2 = {} & 2\lambda \left\langle \xi_t, \mathbf{u}^{\lambda,\mu}(t,X(t)) - \mathbf{u}^{\lambda,\mu}(t,Y(t)) \right\rangle \mathrm{d}t \\ & + 2\left\langle \xi_t, \nabla\theta^{\lambda,\mu}(t,X(t))B(t,X_t,\mu_t) - \nabla\theta^{\lambda,\mu}(t,Y(t))B(t,Y_t,\nu_t) \right\rangle \mathrm{d}t \\ & + 2\left\langle \xi_t, [(\nabla\theta^{\lambda,\mu}\sigma^\mu)(t,X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t,Y(t))]\mathrm{d}W(t) \right\rangle \\ & + \left\| (\nabla\theta^{\lambda,\mu}\sigma^\mu)(t,X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t,Y(t)) \right\|_{HS}^2 \mathrm{d}t \\ & - 2\left\langle \xi_t, \nabla\theta^{\lambda,\mu}(b^\nu - b^\mu)](t,Y(t)) \right\rangle \mathrm{d}t \\ & - \left\langle \xi_t, \mathrm{Tr}[(\sigma^\nu(\sigma^\nu)^* - \sigma^\mu(\sigma^\mu)^*)\nabla^2\mathbf{u}^{\lambda,\mu}](t,Y(t)) \right\rangle \mathrm{d}t. \end{aligned}$$

So, for any $m \ge 1$, it holds

$$\begin{aligned} \mathrm{d}|\xi_t|^{2m} = {} & 2m\lambda|\xi_t|^{2(m-1)} \left\langle \xi_t, \mathbf{u}^{\lambda,\mu}(t,X(t)) - \mathbf{u}^{\lambda,\mu}(t,Y(t)) \right\rangle \mathrm{d}t \\ & + 2m|\xi_t|^{2(m-1)} \left\langle \xi_t, \nabla\theta^{\lambda,\mu}(t,X(t))B(t,X_t,\mu_t) - \nabla\theta^{\lambda,\mu}(t,Y(t))B(t,Y_t,\nu_t) \right\rangle \mathrm{d}t \\ & + 2m|\xi_t|^{2(m-1)} \left\langle \xi_t, [(\nabla\theta^{\lambda,\mu}\sigma^\mu)(t,X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t,Y(t))]\mathrm{d}W(t) \right\rangle \\ & + m|\xi_t|^{2(m-1)} \left\| (\nabla\theta^{\lambda,\mu}\sigma^\mu)(t,X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t,Y(t)) \right\|_{HS}^2 \mathrm{d}t \\ & + 2m(m-1)|\xi_t|^{2(m-2)} \left| [(\nabla\theta^{\lambda,\mu}\sigma^\mu)(t,X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t,Y(t))]^*\xi_t \right|^2 \mathrm{d}t \\ & - 2m|\xi_t|^{2(m-1)} \left\langle \xi_t, \nabla\theta^{\lambda,\mu}(b^\nu - b^\mu)](t,Y(t)) \right\rangle \mathrm{d}t \\ & - m|\xi_t|^{2(m-1)} \left\langle \xi_t, \mathrm{Tr}[(\sigma^\nu(\sigma^\nu)^* - \sigma^\mu(\sigma^\mu)^*)\nabla^2\mathbf{u}^{\lambda,\mu}](t,Y(t)) \right\rangle \mathrm{d}t. \end{aligned}$$

Firstly, applying (4.9), (4.13), $(H^\theta)(3)$, (2.2), (4.17) and Young's inequality $a^p b^{1-p} \leq pa + (1-p)b, a, b \geq 0, p \in (0,1)$ with $p = \frac{2m-1}{2m}$, there exists a constant $c > 0$ such that

$$
\begin{aligned}
&|\xi_t|^{2(m-1)} \left\langle \xi_t, \nabla\theta^{\lambda,\mu}(t, X(t))B(t, X_t, \mu_t) - \nabla\theta^{\lambda,\mu}(t, Y(t))B(t, Y_t, \nu_t)\right\rangle \\
&\leq c|\xi_t|^{2m}\big(\mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, X(t)) + \mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, Y(t))\big) \\
&\quad + c|\xi_t|^{2m-1}\|X_t - Y_t\|_{\mathscr{C}} + c|\xi_t|^{2m-1}\mathbb{W}_\theta(\mu_t, \nu_t) \\
&\leq c|\xi_t|^{2m}\big(\mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, X(t)) + \mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, Y(t))\big) \\
&\quad + c\sup_{s\in[0,t]}|\xi_s|^{2m} + c\mathbb{W}_\theta(\mu_t, \nu_t)^{2m} + c\|X_0 - Y_0\|_{\mathscr{C}}^{2m}.
\end{aligned}
\tag{4.18}
$$

The other terms can be treated as in [10, (4.19)-(4.22)]. For reader's convenience, we give them one by one. It follows from (4.13) and (4.17) that

$$
|\xi_t|^{2(m-1)} \left\langle \xi_t, \mathbf{u}^{\lambda,\mu}(t, X(t)) - \mathbf{u}^{\lambda,\mu}(t, Y(t))\right\rangle \leq \frac{1}{4}|\xi_t|^{2m}.
$$

Furthermore, applying (2.1), $(H^\theta)(2)$, (4.13), (4.14), (4.17) and Young's inequality, we arrive at

$$
\begin{aligned}
&m|\xi_t|^{2(m-1)} \left\|(\nabla\theta^{\lambda,\mu}\sigma^\mu)(t, X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t, Y(t))\right\|_{HS}^2 \\
&\quad + 2m(m-1)|\xi_t|^{2(m-2)}\left|[(\nabla\theta^{\lambda,\mu}\sigma^\mu)(t, X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t, Y(t))]^*\xi_t\right|^2 \\
&\leq c|\xi_t|^{2m-2}\mathbb{W}_\theta(\mu_t, \nu_t)^2 + c|\xi_t|^{2m}\big(\mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, X(t)) + \mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, Y(t))\big)^2 \\
&\quad + c|\xi_t|^{2m}\big(\mathscr{M}\|\nabla\sigma^\mu\|(t, X(t)) + \mathscr{M}\|\nabla\sigma^\mu\|(t, Y(t))\big)^2 \\
&\leq c\mathbb{W}_\theta(\mu_t, \nu_t)^{2m} + c|\xi_t|^{2m} + c|\xi_t|^{2m}\big(\mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, X(t)) + \mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(t, Y(t))\big)^2 \\
&\qquad + c|\xi_t|^{2m}\big(\mathscr{M}\|\nabla\sigma^\mu\|(t, X(t)) + \mathscr{M}\|\nabla\sigma^\mu\|(t, Y(t))\big)^2.
\end{aligned}
$$

Similarly, (2.1), $(H^\theta)(2)$, (4.13), (4.14), $\frac{2m}{2m-1} \leq 2$ and Young's inequality imply that

$$
\begin{aligned}
&|\xi_t|^{2(m-1)} \left\langle \xi_t, \nabla\theta^{\lambda,\mu}(b^\nu - b^\mu)](t, Y(t))\right\rangle \leq c|\xi_t|^{2m-1}\mathbb{W}_\theta(\mu_t, \nu_t) \leq c|\xi_t|^{2m} + c\mathbb{W}_\theta(\mu_t, \nu_t)^{2m}, \\
&|\xi_t|^{2(m-1)} \left\langle \xi_t, \mathrm{Tr}[(\sigma^\nu(\sigma^\nu)^* - \sigma^\mu(\sigma^\mu)^*)\nabla^2\mathbf{u}^{\lambda,\mu}](t, Y(t))\right\rangle \\
&\quad \leq c|\xi_t|^{2m-1}|\nabla^2\mathbf{u}^{\lambda,\mu}(t, Y(t))|\mathbb{W}_\theta(\mu_t, \nu_t) \\
&\quad \leq c|\xi_t|^{2m}|\nabla^2\mathbf{u}^{\lambda,\mu}(t, Y(t))|^{\frac{2m}{2m-1}} + c\mathbb{W}_\theta(\mu_t, \nu_t)^{2m} \\
&\quad \leq c|\xi_t|^{2m}(1 + |\nabla^2\mathbf{u}^{\lambda,\mu}(t, Y(t))|^2) + c\mathbb{W}_\theta(\mu_t, \nu_t)^{2m}.
\end{aligned}
$$

Thus, it holds

$$
\mathrm{d}|\xi_t|^{2m} \leq c\sup_{s\in[0,t]}|\xi_s|^{2m}\mathrm{d}t + c|\xi_t|^{2m}\mathrm{d}A_t + c(\mathbb{W}_\theta(\mu_t, \nu_t)^{2m} + \|X_0 - Y_0\|_{\mathscr{C}}^{2m})\mathrm{d}t + \mathrm{d}M_t
\tag{4.19}
$$

for some constant $c > 0$, a local martingale M_t defined by

$$
M_s = \int_0^s 2m|\xi_t|^{2(m-1)} \left\langle \xi_t, [(\nabla\theta^{\lambda,\mu}\sigma^\mu)(t, X(t)) - (\nabla\theta^{\lambda,\mu}\sigma^\nu)(t, Y(t))]\mathrm{d}W(t)\right\rangle,
$$

and

$$A_t := \int_0^t \Big\{1 + \mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(s,X(s)) + \mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(s,Y(s)) + |\nabla^2\mathbf{u}^{\lambda,\mu}(s,Y(s))|^2 + \big(\mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(s,X(s)) + \mathscr{M}\|\nabla^2\theta^{\lambda,\mu}\|(s,Y(s))\big)^2 + \big(\mathscr{M}\|\nabla\sigma^{\mu}\|(s,X(s)) + \mathscr{M}\|\nabla\sigma^{\mu}\|(s,Y(s))\big)^2\Big\}\mathrm{d}s.$$

By Itô's formula and $X_0 = Y_0$, we have

$$\mathrm{d}\mathrm{e}^{-A_t}|\xi_t|^{2m} \le c\mathrm{e}^{-A_t}\sup_{s\in[0,t]}|\xi_s|^{2m}\mathrm{d}t + c\mathrm{e}^{-A_t}\mathbb{W}_\theta(\mu_t,\nu_t)^{2m}\mathrm{d}t + \mathrm{e}^{-A_t}\mathrm{d}M_t.$$

When $2m > \theta$, we can take $p \in (0,1)$ such that $2mp > \theta$. By Hölder's inequality, (4.17) and the stochastic Grönwall lemma [1, Lemma A.5], we arrive at

$$\begin{aligned}
\mathbb{W}_\theta(\mu_t,\nu_t)^{2m} &\le c\left(\mathbb{E}\sup_{s\in[0,t]}|\xi_s|^\theta\right)^{\frac{2m}{\theta}} \\
&\le c\left(\mathbb{E}\left(\mathrm{e}^{\frac{\theta}{2m}A_t}\sup_{s\in[0,t]}\left(\mathrm{e}^{-\frac{\theta}{2m}A_s}|\xi_s|^\theta\right)\right)\right)^{\frac{2m}{\theta}} \\
&\le c\left(\mathbb{E}\mathrm{e}^{\frac{2mp}{2mp-\theta}\frac{\theta}{2m}A_t}\right)^{\frac{2mp-\theta}{p\theta}}\left(\mathbb{E}\left(\sup_{s\in[0,t]}\left(\mathrm{e}^{-\frac{\theta}{2m}A_s}|\xi_s|^\theta\right)\right)^{\frac{2mp}{\theta}}\right)^{\frac{1}{p}} \qquad (4.20)\\
&= c\left(\mathbb{E}\mathrm{e}^{\frac{\theta p}{2mp-\theta}A_t}\right)^{\frac{2mp-\theta}{p\theta}}\left(\mathbb{E}\left(\sup_{s\in[0,t]}\left(\mathrm{e}^{-A_s}|\xi_s|^{2m}\right)\right)^{p}\right)^{\frac{1}{p}} \\
&\le c\left(\mathbb{E}\mathrm{e}^{\frac{\theta p}{2mp-\theta}A_t}\right)^{\frac{2mp-\theta}{p\theta}}\int_0^t \mathbb{W}_\theta(\mu_s,\nu_s)^{2m}\mathrm{d}s, \quad t\in[0,T]
\end{aligned}$$

for some constants $c > 0$. By Lemma 3.1, (4.10), (4.14) and the Khasminskii's type estimate (3.3), see for instance [24, Lemma 3.5], we have

$$\mathbb{E}\mathrm{e}^{\frac{\theta p}{2mp-\theta}A_T} < \infty,$$

so that by the Grönwall lemma we prove

$$\mathbb{W}_\theta(\mu_t,\nu_t) = 0, \quad t\in[0,T].$$

Combining this with (4.20), we conclude $\mathbb{P}$-a.s. $\xi_t = 0, t \in [0,T]$. This again together with (4.17) implies $\mathbb{P}$-a.s. $X(t) = Y(t), t\in[0,T]$. □

5 Proofs of Theorem 2.3 and Theorem 2.4

Before giving the proofs of Theorem 2.3 and Theorem 2.4, we present a result on the existence and uniqueness of strong solution to (1.1) under **(H)**.

Lemma 5.1. *Assume* **(H)**. *Then for any $\mathscr{F}_0$-measurable $\mathscr{C}$-valued random variable X_0 with $\mathscr{L}_{X_0} \in \mathscr{P}_2$,* (1.1) *has a unique strong solution with initial value X_0.*

Proof. By [11, Theorem 1.1] with $\mathbb{H} = \mathbb{R}^d$ and **(H)**, SDE (3.8) for $\hat{B} = B + b$ and $\hat{\Sigma} = \sigma$ has a unique strong solution X_t up to life time. Combining this with Theorem 2.1(1) and Lemma 3.3, we conclude that the SDE (1.1) has a strong solution under **(H)**. In the following, we prove the uniqueness of strong solution. Under **(H)**, repeating the proof of Lemma 4.2 and using $\|\nabla^2 \mathbf{u}^{\lambda,\mu}\|_\infty \leq \frac{1}{5}$ from [10, (5.6)] in place of (4.14), we arrive at

$$\mathrm{d}A_t \leq C\mathrm{d}t$$

for some constant $C > 0$. Taking $\theta = 2$ and $m = 1$ in (4.18) and (4.19), it follows from Burkerholder-Davis-Gundy's inequality, **(H)** and $\mathbb{W}_2(\mathscr{L}_{X_t}, \mathscr{L}_{Y_t})^2 \leq \mathbb{E}\|X_t - Y_t\|_{\mathscr{C}}^2$ that

$$\mathbb{E} \sup_{v\in[0,t]} |\xi_v|^2 \leq C(T) \int_0^t \mathbb{E} \sup_{s\in[0,v]} |\xi_s|^2 \mathrm{d}v + C(T)\mathbb{E}\|X_0 - Y_0\|_{\mathscr{C}}^2, \quad t \in [0, T]$$

for some constant $C(T) > 0$. Thus, Grönwall's lemma and (4.17) imply there exists a constant $\Gamma(T) > 0$ such that

$$\begin{aligned} \mathbb{W}_2(\mathscr{L}_{X_t}, \mathscr{L}_{Y_t})^2 \leq \mathbb{E}\|X_t - Y_t\|_{\mathscr{C}}^2 &\leq \frac{25}{8}\mathbb{E} \sup_{v\in[0,t]} |\xi_v|^2 + 2\mathbb{E}\|X_0 - Y_0\|_{\mathscr{C}}^2 \\ &\leq \Gamma(T)\mathbb{E}\|X_0 - Y_0\|_{\mathscr{C}}^2, \quad t \in [0, T]. \end{aligned} \tag{5.1}$$

Thus, we complete the proof. □

5.1 Proof of Theorem 2.3

When $\sigma(t, x, \mu)$ does not depend on μ, (1.1) reduces to

$$\mathrm{d}X(t) = b(t, X(t), \mathscr{L}_{X_t})\mathrm{d}t + B(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X(t))\mathrm{d}W(t). \tag{5.2}$$

Proof of Theorem 2.3. For $\mu_t := P_t^*\mu_0$ and $\nu_t := P_t^*\nu_0$, we may rewrite (5.2) as

$$\mathrm{d}X(t) = \bar{b}(t, X(t))\mathrm{d}t + \bar{B}(t, X_t)\mathrm{d}t + \sigma(t, X(t))\mathrm{d}\bar{W}(t), \quad \mathscr{L}_{X_0} = \mu_0,$$

where

$$\begin{aligned} &\bar{b}(t, x) := b(t, x, \nu_t), \quad \bar{B}(t, \xi) := B(t, \xi, \nu_t), \quad \mathrm{d}\bar{W}(t) := \mathrm{d}W(t) + \bar{\gamma}(t)\mathrm{d}t, \\ &\bar{\gamma}(t) := [\sigma^*(\sigma\sigma^*)^{-1}](t, X(t))[b(t, X(t), \mu_t) - b(t, X(t), \nu_t) + B(t, X_t, \mu_t) - B(t, X_t, \nu_t)]. \end{aligned}$$

Then by (5.1) and **(H)**, we have

$$|\bar{\gamma}(t)| \leq C\mathbb{W}_2(\mu_t, \nu_t)^2 \leq C(T)\mathbb{E}\|X_0 - Y_0\|_{\mathscr{C}}^2, \quad t \in [0, T]. \tag{5.3}$$

Let

$$\bar{R}_t = \exp\left\{-\int_0^t \langle\bar{\gamma}(s), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^t |\bar{\gamma}(s)|^2 \mathrm{d}s\right\}, \quad t \in [0, T]. \tag{5.4}$$

By Girsanov's theorem, $\{\bar{W}(t)\}_{t\in[0,T]}$ is a d-dimensional Brownian motion under the probability measure $\bar{\mathbb{P}}_T := \bar{R}_T\mathbb{P}$.

According to the proof of [9, Lemma 3.2], we can construct an adapted process $\tilde{\gamma}(t)$ on $\mathbb{R}^d$ such that

$$\tilde{R}_t := \exp\left\{-\int_0^t \langle\tilde{\gamma}(s), \mathrm{d}\bar{W}(s)\rangle - \frac{1}{2}\int_0^t |\tilde{\gamma}(s)|^2 \mathrm{d}s\right\}, \quad t \in [0, T]$$

is a martingale under the probability measure $\bar{\mathbb{P}}_T$. Thus, under probability measure $\tilde{\mathbb{P}}_T := \tilde{R}_T\bar{\mathbb{P}}_T = \tilde{R}_T\bar{R}_T\mathbb{P}$,

$$\tilde{W}(t) := \bar{W}(t) + \int_0^t \tilde{\gamma}(s)\mathrm{d}s = W(t) + \int_0^t \big(\bar{\gamma}(s) + \tilde{\gamma}(s)\big)\mathrm{d}s, \quad t \in [0, T]$$

is a d-dimensional Brownian motion. Moreover, there exists $C(T) > 0$ such that

$$\mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T |\tilde{\gamma}(s)|^2 \mathrm{d}s \leq C(T)\mathbb{E}\Big(\frac{|X(0) - Y(0)|^2}{T - r} + \|X_0 - Y_0\|_{\mathscr{C}}^2\Big). \tag{5.5}$$

Meanwhile, we can construct a family of homeomorphism $\{\bar{\theta}_t\}_{t\in[0,T]}$ on $\mathscr{C}$ and an $\mathscr{C}$-valued continuous stochastic process $\{\bar{Y}_t\}_{t\in[0,T]}$ such that $\mathscr{L}_{\bar{\theta}_t^{-1}(\bar{Y}_t)}|\tilde{\mathbb{P}}_T = \nu_t, t \in [0, T]$ and $\tilde{\mathbb{P}}_T$-a.s. $X_T = \bar{\theta}_T^{-1}(\bar{Y}_T)$. Let $Y_T = \bar{\theta}_T^{-1}(\bar{Y}_T)$.

Thus, we obtain

$$(P_T f)(\nu_0) = \mathbb{E}_{\tilde{\mathbb{P}}_T} f(Y_T) = \mathbb{E}_{\tilde{P}_T} f(X_T) = \mathbb{E}[\bar{R}_T\tilde{R}_T f(X_T)], \quad f \in \mathscr{B}_b^+(\mathscr{C}).$$

Letting $R_T = \bar{R}_T\tilde{R}_T$, by Young's inequality and Hölder's inequality respectively, we obtain

$$(P_T \log f)(\nu_0) \leq \mathbb{E}[R_T \log R_T] + \log\mathbb{E}f(X_T) = \mathbb{E}[R_T \log R_T] + \log(P_T f)(\mu_0), \tag{5.6}$$

and

$$(P_T f(\nu_0))^p \leq (\mathbb{E}R_T^{\frac{p}{p-1}})^{p-1}\mathbb{E}f^p(X_T) = (\mathbb{E}R_T^{\frac{p}{p-1}})^{p-1}P_T f^p(\mu_0), \quad p > 1. \tag{5.7}$$

By (5.3) and (5.5),

$$\begin{aligned}\mathbb{E}[R_T\log R_T] &\le \frac{1}{2}\mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T|\bar\gamma(s)+\tilde\gamma(s)|^2\mathrm{d}s\\ &\le \mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T|\tilde\gamma(s)|^2\mathrm{d}s+\mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T|\bar\gamma(s)|^2\mathrm{d}s\\ &\le \mathbb{E}_{\tilde{\mathbb{P}}_T}\int_0^T|\tilde\gamma(s)|^2\mathrm{d}s+\int_0^T C(t)\mathbb{W}_2(\mu_t,\nu_t)^2\mathrm{d}t\\ &\le H_1(T)\mathbb{E}\bigg(\frac{|X(0)-Y(0)|^2}{T-r}+\|X_0-Y_0\|_{\mathscr{C}}^2\bigg),\quad T>r\end{aligned}$$

holds for some $H_1\in C(\mathbb{R}_+;\mathbb{R}_+)$. Combining this with (5.6) we obtain the log-Harnack inequality.

Finally, according to the proof of [8, Theorem 4.1], there exists $p_0>1$ and $H_0\in C([p_0,\infty)\times(r,\infty),\mathbb{R}_+)$ such that for any $p\ge p_0$,

$$(\mathbb{E}_{\bar{\mathbb{P}}_T}\tilde R_T^{\frac{p}{p-1}})^{\frac{p-1}{p}}\le \mathbb{E}\mathrm{e}^{H_0(p,T)\big(1+\frac{|X(0-Y(0)|^2}{T-r}+\|X_0-Y_0\|_{\mathscr{C}}^2\big)},\quad T>r.$$

By applying this estimate for $p_1:=\frac{1}{2}(p+p_0)$ and combining with $R_T=\tilde R_T\bar R_T$, (5.3) and (5.4), we arrive at

$$\begin{aligned}\Big(\mathbb{E}R_T^{\frac{p}{p-1}}\Big)^{\frac{p-1}{p}} &= \Big(\mathbb{E}_{\bar{\mathbb{P}}_T}\Big(\tilde R_T^{\frac{p}{p-1}}\bar R_T^{\frac{1}{p-1}}\Big)\Big)^{\frac{p-1}{p}}\le\Big(\mathbb{E}_{\bar{\mathbb{P}}_T}\tilde R_T^{\frac{p_1}{p_1-1}}\Big)^{\frac{p_1-1}{p_1}}\Big(\mathbb{E}_{\bar{\mathbb{P}}_T}\bar R_T^{\frac{p_1}{p-p_1}}\Big)^{\frac{p-p_1}{pp_1}}\\ &\le \mathbb{E}\mathrm{e}^{H_0(p_1,T)\big(1+\frac{|X(0-Y(0)|^2}{T-r}+\|X_0-Y_0\|_{\mathscr{C}}^2\big)}\Big(\mathbb{E}\bar R_T^{\frac{p}{p-p_1}}\Big)^{\frac{p-p_1}{pp_1}}\\ &\le \mathbb{E}\mathrm{e}^{H_2(p,T)\big(1+\frac{|X(0-Y(0)|^2}{T-r}+\|X_0-Y_0\|_{\mathscr{C}}^2\big)},\quad T>r\end{aligned}$$

for some $H_2\in C([p_0,\infty)\times(r,\infty),\mathbb{R}_+)$. Therefore, (2.4) follows from (5.7). □

5.2 Proof of Theorem 2.4

Proof of Theorem 2.4. By the semigroup property for P_t^* and Jensen's inequality, we only need to consider $T-r\in(0,1]$. Define

$$\gamma(s):=\begin{cases}\frac{s^+}{T-r}\eta(-r), & \text{if } s\in[-r,T-r],\\ \eta(s-T), & \text{if } s\in(T-r,T].\end{cases}$$

Next, we construct coupling by change of measure. For fixed $\mu_0\in\mathscr{P}_2$, let $X(t)$ solve (2.5) with $\mathscr{L}_{X_0}=\mu_0$, and let $\mu_t=\mathscr{L}_{X_t}$. Assume that $\bar X(t)$ solves

$$\mathrm{d}\bar X(t)=\{b(t,X(t),\mu_t)+B(t,X_t,\mu_t)\}\mathrm{d}t+\sigma(t,\mu_t)\mathrm{d}W(t)+\gamma'(t)\mathrm{d}t \tag{5.8}$$

with $\bar{X}_0 = X_0$. Then it is not difficult to see that

$$\bar{X}(s) = X(s) + \gamma(s), \quad s \in [-r, T].$$

In particular, it holds that $\bar{X}_T = X_T + \eta$. By the definition of γ, there exists a constant $C > 0$ such that for any $s \in [0, T]$,

$$\begin{aligned} |\gamma'(s)| &\le C 1_{[0,T-r]}(s) \frac{|\eta(-r)|}{T-r} + C 1_{[T-r,T]}(s) |\eta'(s-T)|, \\ |\gamma(s)| &\le C|\eta(-r)| + C\|\eta\|_{\mathscr{C}} \le C\|\eta\|_{\mathscr{C}}, \quad \|\gamma_s\|_{\mathscr{C}} \le C\|\eta\|_{\mathscr{C}}. \end{aligned} \tag{5.9}$$

Let

$$\begin{aligned} \bar{\Phi}(s) = (\sigma^*(\sigma\sigma^*)^{-1})(s, \mu_s) \big[& b(s, X(s), \mu_s) - b(s, \bar{X}(s), \mu_s) \\ & + B(s, X_s, \mu_s) - B(s, \bar{X}_s, \mu_s) + \gamma'(s) \big], \quad s \in [0, T]. \end{aligned}$$

From **(H)** and (5.9), we obtain

$$\begin{aligned} \int_0^T |\bar{\Phi}(s)|^2 \mathrm{d}s &\le C \int_0^T \left(\phi(|\gamma(s)|) + \|\gamma_s\|_{\mathscr{C}} + |\gamma'(s)| \right)^2 \mathrm{d}s \\ &\le C \frac{|\eta(-r)|^2}{T-r} + C \int_{-r}^0 |\eta'(s)|^2 \mathrm{d}s + CT\phi^2 \left(C\|\eta\|_{\mathscr{C}} \right) + CT\|\eta\|_{\mathscr{C}}^2 \end{aligned} \tag{5.10}$$

for some constant $C > 0$. Set

$$\bar{R}(s) = \exp\left[- \int_0^s \langle \bar{\Phi}(u), \mathrm{d}W(u) \rangle - \frac{1}{2} \int_0^s |\bar{\Phi}(u)|^2 \mathrm{d}u \right],$$

and

$$\bar{W}(s) = W(s) + \int_0^s \bar{\Phi}(u) \mathrm{d}u.$$

Girsanov's theorem implies that $\bar{W}$ is a Brownian motion on $[0, T]$ under $\bar{\mathbb{Q}}_T = \bar{R}(T)\mathbb{P}$. Then (5.8) reduces to

$$\mathrm{d}\bar{X}(t) = \{ b(t, \bar{X}(t), \mu_t) + B(t, \bar{X}_t, \mu_t) \} \mathrm{d}t + \sigma(t, \mu_t) \mathrm{d}\bar{W}(t).$$

Thus, the distribution of $\bar{X}_T$ under $\bar{\mathbb{Q}}_T$ coincides with that of X_T under $\mathbb{P}$.

Furthermore, by Young's inequality and Hölder's inequality, we get

$$\begin{aligned} P_T \log f(\mu_0) &= \mathbb{E}_{\bar{\mathbb{Q}}_T} \log f(\bar{X}_T) \\ &= \mathbb{E}_{\bar{\mathbb{Q}}_T} \log f(X_T + \eta) \\ &\le \log P_T f(\cdot + \eta)(\mu_0) + \mathbb{E}\left(\bar{R}(T) \log \bar{R}(T) \right), \end{aligned}$$

and

$$
\begin{aligned}
P_T f(\mu_0) &= \mathbb{E}_{\bar{\mathbb{Q}}_T} f(\bar{X}_T) \\
&= \mathbb{E}_{\bar{\mathbb{Q}}_T} f(X_T+\eta) \le (P_T f^p(\cdot+\eta))^{\frac{1}{p}}(\mu_0)\{\mathbb{E}\bar{R}(T)^{\frac{p}{p-1}}\}^{\frac{p-1}{p}}.
\end{aligned}
$$

Since $\bar{W}$ is a Brownian motion under $\bar{\mathbb{Q}}_T$, by the definition of $\bar{R}(T)$, it is easy to see that

$$
\mathbb{E}\left(\bar{R}(T)\log\bar{R}(T)\right) = \mathbb{E}_{\bar{\mathbb{Q}}_T}\log\bar{R}(T) = \frac{1}{2}\mathbb{E}_{\bar{\mathbb{Q}}_T}\int_0^T |\bar{\Phi}(u)|^2 \mathrm{d}u \le \beta(T,\eta,r),
$$

and

$$
\begin{aligned}
&\mathbb{E}\bar{R}(T)^{\frac{p}{p-1}} \\
&\le \mathbb{E}\Bigg\{\exp\left[\frac{-p}{p-1}\int_0^T\langle\bar{\Phi}(u),\mathrm{d}W(u)\rangle - \frac{1}{2}\frac{p^2}{(p-1)^2}\int_0^T|\bar{\Phi}(u)|^2\mathrm{d}u\right] \\
&\qquad\times\exp\left[\frac{1}{2}\frac{p^2}{(p-1)^2}\int_0^T|\bar{\Phi}(u)|^2\mathrm{d}u - \frac{1}{2}\frac{p}{p-1}\int_0^T|\bar{\Phi}(u)|^2\mathrm{d}u\right]\Bigg\} \\
&\le \operatorname*{ess\,sup}_{\Omega}\exp\left\{\frac{p}{2(p-1)^2}\int_0^T|\bar{\Phi}(u)|^2\mathrm{d}u\right\}.
\end{aligned}
$$

Combining this with (5.10), the shift Harnack inequality holds. □

References

[1] S. Bachmann, *Well-posedness and stability for a class of stochastic delay differential equations with singular drift,* Stochastics and Dynamics, 18(2018), 27pp.

[2] K. Carrapatoso, *Exponential convergence to equilibrium for the homogeneous Landau equation with hard potentials,* Bull. Sci. Math. 139(2015), 777-805.

[3] G. Crippa, C. De Lellis, *Estimates and regularity results for the DiPerna- Lions flow,* J. Reine Angew. Math. 616(2008), 15-46.

[4] P. E. Chaudru de Raynal, *Strong well-posedness of McKean-Vlasov stochastic differential equation with Hölder drift,* Stoch. Proc. Appl. 130(2020), 79-107.

[5] L. Desvillettes, C. Villani, *On the spatially homogeneous Landau equation for hard potentials, Part I: existence, uniqueness and smothness,* Comm. Part. Diff. Equat. 25(2000), 179-259.

[6] L. Desvillettes, C. Villani, *On the spatially homogeneous Landau equation for hard potentials, Part II: H-Theorem and Applications,* Comm. Part. Diff. Equat. 25(2000), 261-298.

[7] L. Gyongy, T. Martinez, *On stochastic differential equations with locally unbounded drift,* Czechoslovak Math. J. 51(2001), 763-783.

[8] X. Huang, M. Röckner, F.-Y. Wang, *Nonlinear Fokker–Planck equations for probability measures on path space and path-distribution dependent SDEs,* Discrete Contin. Dyn. Syst. 39(2019), 3017-3035.

[9] X. Huang, F.-Y. Wang, *Functional SPDE with Multiplicative Noise and Dini Drift,* Ann. Fac. Sci. Toulouse Math. 6(2017), 519-537.

[10] X. Huang, F.-Y. Wang, *Distribution Dependent SDEs with Singular Coefficients,* Stoch. Proc. Appl. 129(2019), 4747-4770.

[11] X. Huang, S.-Q. Zhang, *Mild Solutions and Harnack Inequality for Functional SPDEs with Dini Drift,* J. Theor. Probab. 32(2019), 303-329.

[12] N. Ikeda, S. Watanabe, *Stochastic Differential Equations and Diffusion Processes,* North-Holland and Kodansha Ltd., Amsterdam-Oxford-New York and Tokyo, 1981.

[13] N. V. Krylov, *Controlled diffusion processes,* Springer, New York, 1980.

[14] Yu. S. Mishura, A. Yu. Veretennikov, *Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations,* arXiv:1603.02212.

[15] M. Röckner, F.-Y. Wang, *Harnack and functional inequalities for generalized Mehler semigroups,* J. Funct. Anal. 203(2007), 237-261.

[16] M. Röckner, X. Zhang, *Well-posedness of distribution dependent SDEs with singular drifts,* arXiv:1809.02216.

[17] A.-S. Sznitman, *Topics in propagation of chaos,* In "École d'Été de Probabilités de Sain-Flour XIX-1989", Lecture Notes in Mathematics 1464, p. 165-251, Springer, Berlin, 1991.

[18] C. Villani, *Optimal transport, old and new,* Springer, New York, 2008.

[19] F.-Y. Wang, *Integration by parts formula and shift Harnack inequality for stochastic equations,* Ann. Probab. 42(2014), 994-1019.

[20] F.-Y. Wang, *Harnack Inequality and Applications for Stochastic Partial Differential Equations,* Springer, New York, 2013.

[21] F.-Y. Wang, *Gradient estimate and applications for SDEs in Hilbert space with multiplicative noise and Dini continuous drift,* J. Differential Equations 260(2016), 2792-2829.

[22] F.-Y. Wang, *Distribution-dependent SDEs for Landau type equations,* Stoch. Proc. Appl. 128(2018), 595-621.

[23] P. Xia, L. Xie, X. Zhang, G. Zhao, *$L^q(L^p)$-theory of stochastic differential equations,* arXiv:1908.01255.

[24] L. Xie, X. Zhang, *Ergodicity of stochastic differential equations with jumps and singular coefficients,* Ann. Inst. H. Poincaré Probab. Statist. 56(2020), 175-229.

[25] X. Zhang, *Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients,* Electron. J. Probab. 16(2011), 1096-1116.

[26] A. K. Zvonkin, *A transformation of the phase space of a diffusion process that removes the drift,* Math. Sb. 93(1974), 129-149.