Distribution Dependent SDEs with Singular Coefficients*

Xing Huang [a)], **Feng-Yu Wang** [a),b)]

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

xinghuang@tju.edu.cn

b)Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, United Kingdom

wangfy@tju.edu.cn

December 17, 2018

Abstract

Under integrability conditions on distribution dependent coefficients, existence and uniqueness are proved for distribution dependent SDEs with non-degenerate noise. When the coefficients are Dini continuous in the space variable, gradient estimates and Harnack type inequalities are derived. These generalize the corresponding results derived for classical SDEs, and are new in the distribution dependent setting.

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

1 Introduction

In order to characterize nonlinear Fokker-Planck equations using SDEs, distribution dependent SDEs (DDSDEs for short) have been intensively investigated, see [9, 11] and references within for McKean-Vlasov type SDEs, and [1, 4, 5] and references within for Landau type equations. To ensure the existence and uniqueness of these type SDEs, growth/regularity conditions are used.

*Supported in part by NNSFC (11771326, 11431014, 11801406,11726627).

Next, since the second named author introduced the dimension-free Harnack inequality in [12] to derive the log-Sobolev inequality on Riemannian manifolds, this new type Harnack inequality has been intensively investigated and applied for SDEs and SPDEs. Under a very general framework, the dimension-free Harnack inequality implies the strong Feller property, Gradient estimates, contractivity properties and heat kernel estimates for the associated Markov semigroups. As a dual inequality, the shift-Harnack inequality has been developed and applied in [14]. We may refer to the monograph [15] for a general theory on dimension-free (shift-)Harack inequalities and applications.

Recently, in [17] the second named author established the (shift-)Harnack inequalities and gradient estimates for DDSDEs with regular coefficients, see also [7] for the study on path-distribution dependent SDEs. On the other hand, by means of Krylov's estimate and Zvonkin's transform [21], the well-posedness of classical SDEs has been derived in [8, 20] under integrability conditions allowing the drift unbounded on compact sets, while the new type Harnack inequality and gradient estimates have been established in [16] when the drift is merely Dinni continuous. The purpose of this paper is to extend results in [8, 16, 20] for singular SDEs to singular distribution dependent SDEs.

Let $\mathscr{P}$ be the set of all probability measures on $\mathbb{R}^d$. Consider the following DDSDE on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b_t(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma_t(X_t, \mathscr{L}_{X_t})\mathrm{d}W_t, \tag{1.1}$$

where W_t is the 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

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

are measurable. When a different probability measure $\tilde{\mathbb{P}}$ is concerned, we use $\mathscr{L}_\xi|\tilde{\mathbb{P}}$ to denote the law of a random variable ξ under the probability $\tilde{\mathbb{P}}$.

By using a priori Krylov's estimate, a weak solution can be constructed for (1.1) by using an approximation argument as in the classical setting, see [6] and references within. To prove the existence of strong solution, we use a fixed distribution μ_t to replace the law of solution $\mathscr{L}_{X_t}$, so that the distribution SDE (1.1) reduces to the classical one. We prove that when the reduced SDE has strong uniqueness, the weak solution of (1.1) also provides a strong solution. We will then use Zvonkin's transform to investigate the uniqueness, for which we first identify the distributions of given two solutions, so that these solutions solve the common reduced SDE, and thus, the path-wise uniqueness follows from existing argument developed for the classical SDEs. However, there is essential difficulty to identify the distributions of two solutions of (1.1). Once we have constructed the desired Zvonkin's transform for (1.1) with singular coefficients, gradient estimates and Harnack type inequalities can be proved as in the regular situation considered in [17].

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, two lemmas on weak convergence of stochastic processes, and a result on the existence of strong solutions for distribution dependent SDEs. Finally, the main results are proved in Sections 4 and 5.

2 Main Results

We first recall Krylov's estimate in the study of SDEs. We will fix a constant $T > 0$, and only consider solutions of (1.1) up to time T. For a measurable function f defined on $[0, T] \times \mathbb{R}^d$, let

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

When $s = 0$, we simply denote $\|f\|_{L_p^q(0,t)} = \|f\|_{L_p^q(t)}$. A key step in the study of singular SDEs is to establish Krylov type estimate (see for instance [8]). For later use we introduce the following notion of K-estimate. We consider the following class of number pairs (p, q):

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

Definition 2.1 (Krylov's Estimate). *An $\mathscr{F}_t$-adapted process $\{X_s\}_{0 \leq s \leq T}$ is said to satisfy K-estimate, if for any $(p, q) \in \mathscr{K}$, there exist constants $\delta \in (0, 1)$ and $C > 0$ such that for any nonnegative measurable function f on $[0, T] \times \mathbb{R}^d$,*

$$\mathbb{E}\left(\int_s^t f_r(X_r) \mathrm{d}r \middle| \mathscr{F}_s \right) \leq C(t - s)^\delta \|f\|_{L_p^q(T)}, \quad 0 \leq s \leq t \leq T. \tag{2.1}$$

We note that (2.1) implies the following Khasminskii type estimate, see for instance [19, Lemma 3.5] and its proof: there exists a constant $c > 0$ such that

$$\mathbb{E}\left(\left(\int_s^t f_r(X_r) \mathrm{d}r \right)^n \middle| \mathscr{F}_s \right) \leq cn!(t - s)^{\delta n} \|f\|_{L_p^q(T)}^n, \quad 0 \leq s \leq t \leq T, \tag{2.2}$$

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

$$\mathbb{E}\left(\mathrm{e}^{\lambda \int_0^T f_r(X_r) \mathrm{d}r} \middle| \mathscr{F}_s \right) \leq \mathrm{e}^{\Lambda\left(1 + \|f\|_{L_p^q(T)}\right)}, \quad s \in [0, T]. \tag{2.3}$$

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

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

It is well known that $\mathscr{P}_\theta$ is a Polish space under the Warsserstein distance

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

where $\mathscr{C}(\mu,\nu)$ is the set of all couplings of μ and ν. Moreover, the topology induced by $\mathbb{W}_\theta$ on $\mathscr{P}_\theta$ coincides with the weak topology.

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

2.1 Existence and Uniqueness

Let

$$\mathscr{P}_\theta^a = \big\{\mu \in \mathscr{P}_\theta : \mu \text{ is absolutely continuous with respect to the Lebesgue measure } \big\}.$$

To construct a weak solution of (1.1) by using approximation argument as in [6, 9], we need the following assumptions for some $\theta \geq 1$.

(H^θ) There exists a sequence $(b^n, \sigma^n)_{n\geq 1}$, where

$$b^n : [0,T] \times \mathbb{R}^d \times \mathscr{P}_\theta \to \mathbb{R}^d, \quad \sigma^n : [0,T] \times \mathbb{R}^d \times \mathscr{P}_\theta \to \mathbb{R}^d \otimes \mathbb{R}^d$$

are measurable, such that the following conditions hold:

(1) For $\mu \in \mathscr{P}_\theta^a$ and $\mu^n \to \mu$ in $\mathscr{P}_\theta$,

$$\lim_{n\to\infty} \big\{|b_t^n(x,\mu^n) - b_t(x,\mu)| + \|\sigma_t^n(x,\mu^n) - \sigma_t(x,\mu)\|\big\} = 0, \quad \text{a.e. } (t,x) \in [0,T] \times \mathbb{R}^d.$$

(2) There exist $K > 1$, $(p,q) \in \mathscr{K}$ and nonnegative $G \in L_p^q(T)$ such that for any $n \geq 1$,

$$|b_t^n(x,\mu)|^2 \leq G(t,x) + K, \quad K^{-1} I \leq (\sigma_t^n(\sigma_t^n)^*)(x,\mu) \leq KI$$

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

(3) For each $n \geq 1$, there exists a constant $K_n > 0$ such that $\|b^n\|_\infty \leq K_n$ and

$$\begin{aligned} &|b_t^n(x,\mu) - b_t^n(y,\nu)| + \|\sigma_t^n(x,\mu) - \sigma_t^n(y,\nu)\| \\ &\leq K_n\big\{|x-y| + \mathbb{W}_\theta(\mu,\nu)\big\}, \quad (t,x,y) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d,\ \mu,\nu \in \mathscr{P}_\theta. \end{aligned} \tag{2.4}$$

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

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

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

The main result in this part is the following.

Theorem 2.1. *Assume* (H^θ) *for some constant* $\theta \geq 1$. *Let* X_0 *be an* $\mathscr{F}_0$*-measurable random variable on* $\mathbb{R}^d$ *with* $\mu_0 := \mathscr{L}_{X_0} \in \mathscr{P}_\theta$. *Then the following assertions hold.*

(1) *The SDE* (1.1) *has a weak solution with initial distribution* μ_0 *satisfying* $\mathscr{L}_{X_\cdot} \in C([0,T]; \mathscr{P}_\theta)$ *and the* K*-estimate.*

(2) *If* σ *is uniformly continuous in* $x \in \mathbb{R}^d$ *uniformly with respect to* $(t,\mu) \in [0,T] \times \mathscr{P}_\theta$, *and for any* $\mu_\cdot \in C([0,T]; \mathscr{P}_\theta)$, $b_t^\mu(x) := b_t(x,\mu_t)$ *and* $\sigma_t^\mu(x) := \sigma_t(x,\mu_t)$ *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 strong solution satisfying* $\mathscr{L}_{X_\cdot} \in C([0,T]; \mathscr{P}_\theta)$ *and the* K*-estimate.*

(3) *If, in addition to the condition in* (2), *there exists a constant* $L > 0$ *such that*

$$\|\sigma_t(x,\mu) - \sigma_t(x,\nu)\| + |b_t(x,\mu) - b_t(x,\nu)| \leq L\,\mathbb{W}_\theta(\mu,\nu) \tag{2.5}$$

holds for all $\mu,\nu \in \mathscr{P}_\theta$ *and* $(t,x) \in [0,T] \times \mathbb{R}^d$, *then the strong solution is unique.*

When b and σ do not depend on the distribution, Theorem 2.1 reduces back to the corresponding results derived for classical SDEs with singular coefficients, see for instance [20] and references within.

To compare Theorem 2.1 with recent results on the existence and uniqueness of DDSDEs derived in [2, 9], we consider a specific class of coefficients where the dependence on distributions is of integral type. For $\mu \in \mathscr{P}$ and a (possibly multidimensional valued) real function $f \in L^1(\mu)$, let $\mu(f) = \int_{\mathbb{R}^d} f\,\mathrm{d}\mu$. Let

$$b_t(x,\mu) := B_t(x, \mu(\psi_b(t,x,\cdot)), \quad \sigma_t(x,\mu) := \Sigma_t(x, \mu(\psi_\sigma(t,x,\cdot)) \tag{2.6}$$

for $(t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathscr{P}_\theta$, where for some $k \in \mathbb{N}$,

$$\psi_b, \psi_\sigma : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^k$$

are measurable and bounded such that for some constant $\delta > 0$,

$$|\psi_b(t,x,y) - \psi_b(t,x,y')| + |\psi_\sigma(t,x,y) - \psi_\sigma(t,x,y')| \leq \delta|y-y'| \tag{2.7}$$

holds for all $(t,x) \in [0,T] \times \mathbb{R}^d$ and $y,y' \in \mathbb{R}^d$, and

$$B : [0,T] \times \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d, \quad \Sigma : [0,T] \times \mathbb{R}^d \times \mathbb{R}^k \to \mathbb{R}^d \otimes \mathbb{R}^d$$

are measurable and continuous in the third variable in $\mathbb{R}^k$. We make the following assumption.

(A) Let (b,σ) in (2.6) for (B,Σ) such that (2.7) holds, $B_t(x,\cdot)$ and $\Sigma_t(x,\cdot)$ are continuous for any $(t,x)\in[0,T]\times\mathbb{R}^d$. Moreover, there exist constant $K>1$, $(p,q)\in\mathscr{K}$ and nonnegative $F\in L_p^q(T)$ such that

$$|b_t(x,\mu)|^2 \le F(t,x)+K, \quad K^{-1}I \le \sigma_t(x,\mu)\sigma_t(x,\mu)^* \le KI \tag{2.8}$$

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

Corollary 2.2. *Assume* **(A)***. Then the following assertions hold.*

(1) Assertion (1) *in Theorem* 2.1 *holds.*

(2) If moreover, σ is uniformly continuous in $x\in\mathbb{R}^d$ uniformly with respect to $(t,\mu)\in[0,T]\times\mathscr{P}_\theta$, and for any $\mu_\cdot\in C([0,T];\mathscr{P}_\theta)$, $b_t^\mu(x):=b_t(x,\mu_t)$ and $\sigma_t^\mu(x):=\sigma_t(x,\mu_t)$ 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 assertion (2) *in Theorem* 2.1 *hold.*

(3) Besides the conditions in (2), if there exists a constant $c>0$ such that

$$|B_t(x,y)-B_t(x,y')|+\|\Sigma_t(x,y)-\Sigma_t(x,y')\| \le c|y-y'|, \quad (t,x)\in[0,T]\times\mathbb{R}^d, y,y'\in\mathbb{R}^k,$$

then for any $\mathscr{F}_0$-measurable random variable X_0 on $\mathbb{R}^d$ with $\mu_0:=\mathscr{L}_{X_0}\in\mathscr{P}_\theta$ for some $\theta\ge1$, the SDE (1.1) *has a unique strong solution with $\mathscr{L}_{X_\cdot}$ continuous in $\mathscr{P}_\theta$.*

In the next corollary on the existence of weak solution we do not assume (2.6). This result will be used in Section 5.

Corollary 2.3. *Assume that* (2.5), (2.8) *hold. Then the SDE* (1.1) *has a weak solution with initial distribution μ_0 satisfying $\mathscr{L}_{X_\cdot}\in C([0,T];\mathscr{P}_\theta)$ and the K-estimate.*

We now explain that results in Corollary 2.2 and Corollary 2.3 are new comparing with existing results on McKean-Vlasov SDEs. We first consider the model in [2] where ψ_b and ψ_σ are $\mathbb{R}$-valued functions such that

$$\|B\|_\infty + \sup_{(t,x,r)\in[0,T]\times\mathbb{R}^d\times\mathbb{R}} |\partial_r B_t(x,r)| < \infty,$$

ψ_b is Hölder continuous, ψ_σ is Lipschitz continuous, and for some constants $C>1$, $\theta\in(0,1]$,

$$\begin{aligned}
&C^{-1}I \le \Sigma\Sigma^* \le CI,\\
&\|\Sigma_t(x,r)-\Sigma_t(x',r')\| \le C(|x-x'|+|r-r'|),\\
&\|\partial_r\Sigma_t(x,r)-\partial_r\Sigma_t(x',r)\| \le C|x-x'|^\theta.
\end{aligned}$$

Then [2, Theorem 1] says that when $\mathscr{L}_{X_0}\in\mathscr{P}_2$ the SDE (1.1) has a unique strong solution. Obviously, the above conditions imply $\|b\|_\infty+\|\nabla\sigma\|_\infty<\infty$, but this is not necessary in Corollary 2.2 and Corollary 2.3, since the integrability conditions used in these two results allow b and $\nabla\sigma$ unbounded on compact sets.

Next, [9] considers (1.1) with

$$b_t(x,\mu) := \int_{\mathbb{R}^d} \tilde{b}_t(x,y)\mu(\mathrm{d}y), \quad \sigma_t(x,\mu) := \int_{\mathbb{R}^d} \tilde{\sigma}_t(x,y)\mu(\mathrm{d}y)$$

for measurable functions

$$\tilde{b} : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d, \quad \tilde{\sigma} : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d\otimes\mathbb{R}^d$$

satisfying

$$\|\tilde{\sigma}_t(x,y)\| + |\tilde{b}_t(x,y)| \le C(1+|x|), \quad \tilde{\sigma}\tilde{\sigma}^* \ge C^{-1}I$$

for some constant $C > 1$. Then [9, Theorem 1] says that when $\mathscr{L}_{X_0} \in \mathscr{P}_4$, (1.1) has a weak solution. If moreover σ does not depend on the distribution and $\|\nabla\sigma\|_\infty < \infty$, then [9, Theorem 2] shows that when $\mathbb{E}\mathrm{e}^{r|X_0|^2} < \infty$ for some $r > 0$, the SDE (1.1) has a unique strong solution. Obviously, to apply these results it is necessary that b and $\nabla\sigma$ are (locally) bounded, which is however not necessary for the condition in Corollary 2.2 and Corollary 2.3, since as mentioned above that the integrability conditions used in these two results allow b and $\nabla\sigma$ unbounded on compact sets.

2.2 Harnack Inequality

In this subsection, we investigate the dimension-free log-Harnack inequality introduced in [10] for (1.1), see [15] and references within for general results on these type Harnack inequalities and applications. We establish Harnack inequalities for $P_t f$ using coupling by change of measures (see for instance [15, §1.1]). To this end, we need to assume that the noise part is distribution-free; that is, we consider the following special version of (1.1):

$$\mathrm{d}X_t = b_t(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t, \quad t\in[0,T]. \tag{2.9}$$

As in [17], we define $P_t f(\mu_0)$ and $P_t^*\mu_0$ as follows:

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

where $X_t(\mu_0)$ solves (2.9) with $\mathscr{L}_{X_0} = \mu_0$. Let

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

Remark 2.4. *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.*

We will need the following assumption.

(H) $\|b\|_\infty < \infty$ and there exist a constant $K > 1$ and $\phi \in \mathscr{D}$ such that for any $t \in [0,T]$, $x, y \in \mathbb{R}^d$, and $\mu, \nu \in \mathscr{P}_2$,

$$K^{-1}I \le (\sigma_t\sigma_t^*)(x) \le KI, \ \|\sigma_t(x) - \sigma_t(y)\|_{\mathrm{HS}}^2 \le K|x-y|^2, \tag{2.10}$$

$$|b_t(x,\mu) - b_t(y,\nu)| \le \phi(|x-y|) + K\mathbb{W}_2(\mu,\nu). \tag{2.11}$$

Theorem 2.5. *Assume* **(H)**.

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

$$(P_t \log f)(\nu_0) \le \log(P_t f)(\mu_0) + \frac{C}{t \wedge 1}\mathbb{W}_2(\mu_0,\nu_0)^2 \tag{2.12}$$

for any $t \in (0,T], \mu_0, \nu_0 \in \mathscr{P}_2, f \in \mathscr{B}_b^+(\mathbb{R}^d)$ with $f \ge 1$. Consequently, for any different $\mu_0, \nu_0 \in \mathscr{P}_2$, and any $f \in \mathscr{B}_b(\mathbb{R}^d)$,

$$\frac{|(P_t f)(\mu_0) - (P_t f)(\nu_0)|^2}{\mathbb{W}_2(\mu_0,\nu_0)^2} \le \frac{2C}{t \wedge 1} \sup_{\nu \in B(\mu_0, \mathbb{W}_2(\mu_0,\nu_0))} \big\{(P_t f^2)(\nu) - (P_t f)^2(\nu)\big\}. \tag{2.13}$$

(2) *There exist constants $p_0 > 1$ and $c_1, c_2 > 0$, such that for any $p > p_0$, $t \in (0,T], f \in \mathscr{B}_b^+(\mathbb{R}^d)$ and $\mu_0, \nu_0 \in \mathscr{P}_2$,*

$$(P_t f)^p(\nu_0) \le (P_t f^p)(\mu_0) \exp\Big[\frac{c_1 p}{t \wedge 1}\mathbb{W}_2(\mu_0,\nu_0)^2\Big]\Big(\mathbb{E}\big[\mathrm{e}^{\frac{c_1|X_0 - Y_0|^2}{1 - \mathrm{e}^{-c_2 t}}}\big]\Big)^p \tag{2.14}$$

holds for $\mathscr{F}_0$-measurable random variables X_0, Y_0 satisfying $\mathscr{L}_{X_0} = \mu_0$, $\mathscr{L}_{Y_0} = \nu_0$.

2.3 Shift Harnack Inequality

In this section, we establish the shift Harnack inequality for P_t introduced in [14]. To this end, we assume that $\sigma_t(x,\mu)$ does not depend on x. So SDE (1.1) becomes

$$\mathrm{d}X_t = b_t(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma_t(\mathscr{L}_{X_t})\mathrm{d}W_t, \quad t \in [0,T]. \tag{2.15}$$

Theorem 2.6. *Let $\sigma : [0,T] \times \mathscr{P}_2 \to \mathbb{R}^d \otimes \mathbb{R}^d$ and $b : [0,\infty) \times \mathbb{R}^d \times \mathscr{P}_2 \to \mathbb{R}^d$ be measurable such that σ is invertible with $\|\sigma_t\|_\infty + \|\sigma_t^{-1}\|_\infty$ is bounded in $t \in [0,T]$, and b satisfies the corresponding conditions in* **(H)**.

(1) *For any $p > 1, t \in [0,T], \mu_0 \in \mathscr{P}_2, v \in \mathbb{R}^d$ and $f \in \mathscr{B}_b^+(\mathbb{R}^d)$,*

$$\begin{aligned}(P_t f)^p(\mu_0) \le &(P_t f^p(v + \cdot))(\mu_0)\\ &\times \exp\Bigg[\frac{p\int_0^t \|\sigma_s^{-1}\|_\infty^2 \big\{|v|/t + \phi(s|v|/t)\big\}^2 \mathrm{d}s}{2(p-1)}\Bigg].\end{aligned}$$

Moreover, for any $f \in \mathscr{B}_b^+(\mathbb{R}^d)$ with $f \ge 1$,

$$(P_t \log f)(\mu_0) \le \log(P_t f(v + \cdot))(\mu_0) + \frac{1}{2}\int_0^t \|\sigma_s^{-1}\|_\infty^2 \big\{|v|/t + \phi(s|v|/t)\big\}^2 \mathrm{d}s.$$

3 Preparations

We first present a new result on Krylov's estimate, then recall two lemmas from [6] for the construction of weak solution, and finally introduce two lemmas on the existence and uniqueness of strong solutions.

3.1 Krylov's Estimate

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

$$\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t, \quad t \in [0,T]. \tag{3.1}$$

Lemma 3.1. *Let $T > 0$, and let $p, q \in (1, \infty)$ 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 nonnegative function $F \in L_p^q(T)$ such that*

$$K^{-1}I \le \sigma_t(x)\sigma_t(x)^* \le KI, \quad (t,x) \in [0,T] \times \mathbb{R}^d, \tag{3.2}$$

$$|b_t(x)| \le K + F(t,x), \quad (t,x) \in [0,T] \times \mathbb{R}^d. \tag{3.3}$$

Then for any $(\alpha, \beta) \in \mathscr{K}$, there exist constants $C = C(\delta, K, \alpha, \beta, \|F\|_{L_p^q(T)}) > 0$ and $\delta = \delta(\alpha, \beta) > 0$, such that for any $s_0 \in [0,T)$ and any solution $(X_{s_0,t})_{t \in [s_0,T]}$ of (3.1) from time s_0,

$$\mathbb{E}\bigg[\int_s^t |f|(r, X_{s_0,r})\mathrm{d}r\bigg|\mathscr{F}_s\bigg] \le C(t-s)^\delta \|f\|_{L_\alpha^\beta(T)}, \ s_0 \le s < t \le T, f \in L_\alpha^\beta(T). \tag{3.4}$$

Proof. When b is bounded, the assertion is due to [20, Theorem 2.1]. If $|b| \le K + F$ for some constant $K > 0$ and $0 \le F \in L_p^q(T)$, then we have a decomposition $b = b^{(1)} + b^{(2)}$ with $\|b^{(1)}\|_\infty \le K$ and $|b^{(2)}| \le F$, for instance, $b^{(1)} = \frac{b}{1 \vee (|b|/K)}$. Let $u : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ solve the PDE

$$\frac{\partial u_t}{\partial t} + \frac{1}{2}\mathrm{Tr}(\sigma_t\sigma_t^*\nabla^2 u_t) + \nabla_{b_t^{(2)}} u_t + b_t^{(2)} = 0, \quad u_T = 0, \tag{3.5}$$

and let $\theta_t(x) = x + u_t(x)$. As in [20, Lemma 4.3], where $(\theta_t, b_t^{(2)})$ is denoted by (Φ_t, b_t), we see that $Y_{s_0,t} = \theta_t(X_{s_0,t})$ for $t \ge s_0$ solves

$$\mathrm{d}Y_t = \bar{b}_t(Y_t)\mathrm{d}t + \bar{\sigma}_t(Y_t)\mathrm{d}W_t, \quad t \in [s_0, T], \tag{3.6}$$

where $\bar{b}$ is bounded, and $\bar{\sigma}$ is uniformly continuous in $x \in \mathbb{R}^d$ uniformly with respect to $t \in [0,T]$. Moreover, there exists a constant $\bar{K} > 1$ depending on K and $\|F\|_{L_p^q(T)}$ such that

$$\bar{K}^{-1}I \le \bar{\sigma}_t(x)\bar{\sigma}_t(x)^* \le \bar{K}I, \quad (t,x) \in [0,T] \times \mathbb{R}^d, \tag{3.7}$$

and

$$\|\bar{b}\|_\infty + \|\nabla\theta\|_\infty + \|\nabla\theta^{-1}\|_\infty \le \bar{K}.$$

Again by [20, Theorem 2.1], there exists a constant $C = C(\delta, \bar{K}, \alpha, \beta) > 0$ and $\delta = \delta(\alpha, \beta) > 0$ such that

$$\mathbb{E}\left[\int_s^t |f|(r, Y_{s_0,r})\mathrm{d}r\middle|\mathscr{F}_s\right] \le C(t-s)^\delta \|f\|_{L^\beta_\alpha(T)},\ s_0 \le s < t \le T, f \in L^\beta_\alpha(T). \tag{3.8}$$

This together with $\|\nabla\theta\|_\infty < \bar{K}$ implies that

$$\begin{aligned}
&\mathbb{E}\left[\int_s^t |f|(r, X_{s_0,r})\mathrm{d}r\middle|\mathscr{F}_s\right] = \mathbb{E}\left[\int_s^t |f|(r, \theta_r^{-1}(Y_{s_0,r}))\mathrm{d}r\middle|\mathscr{F}_s\right] \\
&\le C(t-s)^\delta \left(\int_0^T \left(\int_{\mathbb{R}^d} |f(r, \theta_r^{-1}(x))|^\alpha \mathrm{d}x\right)^{\frac{\beta}{\alpha}} \mathrm{d}r\right)^{\frac{1}{\beta}} \\
&= C(t-s)^\delta \left(\int_0^T \left(\int_{\mathbb{R}^d} |f(r, y)|^\alpha |\mathrm{det}\nabla\theta_r|(y)\mathrm{d}y\right)^{\frac{\beta}{\alpha}} \mathrm{d}r\right)^{\frac{1}{\beta}} \\
&\le C(t-s)^\delta \|f\|_{L^\beta_\alpha(T)},\ s_0 \le s < t \le T, , f \in L^\beta_\alpha(T).
\end{aligned}$$

Then the proof is finished. □

3.2 Convergence of Stochastic Processes

To prove Theorem 2.1(1), we will use the following two lemmas due to [6, Lemma 5.1, 5.2].

Lemma 3.2. *Let $\{\psi^n\}_{n\ge1}$ be a sequence of d-dimensional processes defined on some probability space. Assume that*

$$\lim_{R\to\infty} \sup_{n\ge1} \sup_{t\in[0,T]} \mathbb{P}(|\psi^n_t| > R) = 0, \tag{3.9}$$

and for any $\varepsilon > 0$,

$$\lim_{\theta\to0} \sup_{n\ge1} \sup_{s,t\in[0,T],|t-s|\le\theta} \{\mathbb{P}(|\psi^n_t - \psi^n_s| > \varepsilon)\} = 0. \tag{3.10}$$

Then there exist a sequence $\{n_k\}_{k\ge1}$, a probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ and stochastic processes $\{X_t, X^k_t\}_{t\in[0,T]}(k \ge 1)$, such that for every $t \in [0,T]$, $\mathscr{L}_{\psi^{n_k}_t}|\mathbb{P} = \mathscr{L}_{X^k_t}|\tilde{\mathbb{P}}$, and X^k_t converges to X_t in probability $\tilde{\mathbb{P}}$ as $k \to \infty$.

Lemma 3.3. *Let $\{\eta^n\}_{n\geq 1}$ and η be uniformly bounded $\mathbb{R}^d\otimes\mathbb{R}^k$-valued stochastic processes, and let W_t^n and W_t for $t\in[0,T]$ be Wiener processes such that the stochastic Itô integrals*

$$I_t^n := \int_0^t \eta_s^n \mathrm{d}W_s^n,\quad I_t := \int_0^t \eta_s \mathrm{d}W_s,\quad t\in[0,T]$$

are well-defined. Assume that $\eta_t^n\to\eta_t$ and $W_t^n\to W_t$ in probability for every $t\in[0,T]$. Then

$$\lim_{n\to\infty}\mathbb{P}\left(\sup_{t\in[0,T]}|I_t^n-I_t|\geq\varepsilon\right)=0,\quad \varepsilon>0.$$

3.3 Existence and Uniqueness on Strong Solutions

We first present a result on the existence of strong solutions deduced from weak solutions, then introduce a result on the existence and uniqueness of strong solutions under a Lipschitz type condition.

Lemma 3.4. *Let $(\bar\Omega,\bar{\mathscr{F}}_t,\bar W_t,\bar{\mathbb{P}})$ and $\bar X_t$ be a weak solution to* (1.1) *with $\mu_t:=\mathscr{L}_{\bar X_t}|\bar{\mathbb{P}}$. If the SDE*

$$\mathrm{d}X_t=b_t(X_t,\mu_t)\,\mathrm{d}t+\sigma_t(X_t,\mu_t)\,\mathrm{d}W_t,\quad 0\leq t\leq T \tag{3.11}$$

has a unique strong solution X_t up to life time with $\mathscr{L}_{X_0}=\mu_0$, then (1.1) *has a strong solution.*

Proof. Since $\mu_t=\mathscr{L}_{\bar X_t}|\bar{\mathbb{P}}$, $\bar X_t$ is a weak solution to (3.11). By Yamada-Watanabe principle, the strong uniqueness of (3.11) 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 (1.1). □

Lemma 3.5. *Let $\theta\geq 1$ and δ_0 be the Dirac measure at point 0. If $b_t(0,\delta_0)$ is bounded in $t\in[0,T]$, and there exists a constant $L>0$ such that*

$$\begin{aligned}&\|\sigma_t(x,\mu)-\sigma_t(y,\nu)\|+|b_t(x,\mu)-b_t(y,\nu)|\\&\leq L\big\{|x-y|+\mathbb{W}_\theta(\mu,\nu)\big\},\quad x,y\in\mathbb{R}^d,\mu,\nu\in\mathscr{P}_\theta,t\in[0,T],\end{aligned} \tag{3.12}$$

then for any X_0 with $\mathbb{E}|X_0|^\theta<\infty$, (1.1) *has a unique strong solution $(X_t)_{t\in[0,T]}$.*

Proof. When $\theta\geq 2$ the assertion follows from [17, Theorem 2.1]. So we only consider $\theta<2$. As explained in [17, Proof of Theorem 2.1 (1)], it suffices to find a constant $t_0\in(0,T)$ independent of X_0 such that (1.1) has a unique strong solution up to time t_0 and $\sup_{t\in[0,t_0]}\mathbb{E}|X_t|^\theta<\infty$.

Let $X_t^{(0)}=X_0$ and $\mu_t^{(0)}=\mu_0$ for $t\in[0,T]$. For any $n\geq 1$, consider the SDE

$$\mathrm{d}X_t^{(n)}=b_t(X_t^{(n)},\mu_t^{(n-1)})\mathrm{d}t+\sigma_t(X_t^{(n)},\mu_t^{(n-1)})\mathrm{d}W_t,\quad X_0^{(n)}=X_0,$$

where $\mu_t^{(n-1)} = \mathscr{L}_{X_t^{(n-1)}}, 0 \le t \le T$. By [17, Lemma 2.3(1)], for any $n \ge 1$ this SDE has a unique solution and

$$\sup_{s\in[0,T]} \mathbb{E}|X_s^{(n)}|^\theta < \infty, \quad n \ge 1. \tag{3.13}$$

Moreover, letting

$$\xi_t^{(n)} := X_t^{(n+1)} - X_t^{(n)}, \quad \Lambda_t^{(n)} := \sigma_t(X_t^{(n+1)}, \mu_t^{(n)}) - \sigma_t(X_t^{(n)}, \mu_t^{(n-1)}),$$

[17, (2.11)] implies

$$\mathrm{d}|\xi_t^{(n)}|^2 \le 2\langle \Lambda_t^{(n)} \mathrm{d}W_t, \xi_t^{(n)}\rangle + K_0\big\{|\xi_t^{(n)}|^2 + \mathbb{W}_\theta(\mu_t^{(n)}, \mu_t^{(n-1)})^2\big\}\mathrm{d}t, \quad n \ge 1, t \in [0,T]$$

for some constant $K_0 > 0$. Since $\xi_0^{(n)} = 0$, it follows that

$$\begin{aligned}\mathbb{E}|\xi_t^{(n)}|^2 &\le \int_0^t K_0 \mathrm{e}^{K_0(t-s)} \mathbb{W}_\theta(\mu_s^{(n)}, \mu_s^{(n-1)})^2 \mathrm{d}s \\ &\le tK_0\mathrm{e}^{K_0T} \sup_{s\in[0,t]} \big(\mathbb{E}|\xi_s^{(n-1)}|^\theta\big)^{\frac{2}{\theta}}, \quad t \in [0,T], n \ge 1.\end{aligned}$$

Since $\theta < 2$, by Jensen's inequality we may find out a constant $K_1 > 0$ such that

$$\sup_{s\in[0,t]} \mathbb{E}|\xi_s^{(n)}|^\theta \le K_1 t^{\frac{\theta}{2}} \sup_{s\in[0,t]} \mathbb{E}|\xi_s^{(n-1)}|^\theta, \quad n \ge 1, t \in [0,T].$$

So, taking $t_0 \in (0, T \wedge K_1^{-\frac{2}{\theta}})$, we may find a constant $\varepsilon \in (0,1)$ such that

$$\sup_{s\in[0,t_0]} \mathbb{E}|\xi_s^{(n)}|^\theta \le \varepsilon^n \sup_{s\in[0,t_0]} \mathbb{E}|X_s^{(1)} - X_0|^\theta < \infty, \quad n \ge 1.$$

Therefore, for any $t \in [0,t_0]$ there exists an $\mathscr{F}_t$-measurable random variable X_t on $\mathbb{R}^d$ such that

$$\lim_{n\to\infty} \sup_{t\in[0,t_0]} \mathbb{W}_\theta(\mu_t^{(n)}, \mu_t)^\theta \le \lim_{n\to\infty} \sup_{t\in[0,t_0]} \mathbb{E}|X_t^{(n)} - X_t|^\theta = 0,$$

where $\mu_t := \mathscr{L}_{X_t}$. Combining this with (3.12) and letting $n \to \infty$ in the equation

$$X_t^{(n)} = \int_0^t b_s(X_s^{(n)}, \mu_s^{(n-1)})\mathrm{d}s + \int_0^t \sigma_s(X_s^{(n)}, \mu_s^{(n-1)})\mathrm{d}W_s, \quad n \ge 1, t \in [0,t_0],$$

we derive for every $t \in [0,t_0]$,

$$X_t = \int_0^t b_s(X_s, \mu_s)\mathrm{d}s + \int_0^t \sigma_s(X_s, \mu_s)\mathrm{d}W_s.$$

Thus, $(X_s)_{s\in[0,t_0]}$ has a continuous version which is a strong solution of (1.1) up to time t_0. The uniqueness is trivial by using condition (3.12) and the Itô formula. $\square$

4 Proofs of Theorem 2.1 and Corollary 2.2

4.1 Proof of Theorem 2.1(1)-(2)

According to [20, Theorem 1.1], the condition in Theorem 2.1(2) implies that the SDE (3.11) has a unique strong solution. So, by Lemma 3.4, Theorem 2.1(2) follows from Theorem 2.1(1). Below we only prove the existence of weak solution.

By Lemma 3.5, condition (3) in (H^θ) implies that the SDE

$$\mathrm{d}X_t^n = b_t^n(X_t^n, \mathscr{L}_{X_t^n})\mathrm{d}t + \sigma_t^n(X_t^n, \mathscr{L}_{X_t^n})\mathrm{d}W_t, \quad X_0^n = X_0 \tag{4.1}$$

has a unique strong solution $(X_t^n)_{t\in[0,T]}$. So, Lemma 3.1, (2.4) and condition (2) in (H^θ) imply that for any $(p,q) \in \mathscr{K}$,

$$\mathbb{E}\int_s^t f(r, X_r^n)\mathrm{d}r \le C(t-s)^\delta \|f\|_{L_p^q(T)}, \quad 0 \le f \in L_p^q(T), n \ge 1 \tag{4.2}$$

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

We first show that Lemma 3.2 applies to (X^n, W) replacing ψ_n, for which it suffices to verify conditions (3.9) and (3.10) with $\psi_n := X^n$. By condition (2) in (H^θ) and (2.2) implied by (3.4), there exist constants $c_1, c_2 > 0$ such that

$$\begin{aligned}\mathbb{E}|X_t^n|^\theta &\le c_1\bigg\{\mathbb{E}|X_0|^\theta + \mathbb{E}\bigg(\int_0^T |b_t^n(X_t^n, \mathscr{L}_{X_t^n})|\,\mathrm{d}t\bigg)^\theta \\ &\qquad + \mathbb{E}\left(\int_0^T \|\sigma_t^n(X_t^n, \mathscr{L}_{X_t^n})\|^2\,\mathrm{d}t\right)^{\frac{\theta}{2}}\bigg\} \\ &\le c_2\Big(\mathbb{E}|X_0|^\theta + T^\theta + \|G\|_{L_p^q(T)}^\theta + T^{\frac{\theta}{2}}\Big) < \infty, \quad n \ge 1, t \in [0,T].\end{aligned} \tag{4.3}$$

Thus, by the Markov inequality, (3.9) holds for $\psi_n = X^n$.

Next, by the same reason, there exists a constant $c_3 > 0$ such that for any $0 \le s \le t \le T$,

$$\begin{aligned}\mathbb{E}|X_t^n - X_s^n| &\le \mathbb{E}\int_s^t |b_r^n(X_r^n, \mathscr{L}_{X_r^n})|\,\mathrm{d}r + \mathbb{E}\left(\int_s^t \|\sigma_r^n(X_r^n, \mathscr{L}_{X_r^n})\|^2\,\mathrm{d}r\right)^{\frac{1}{2}} \\ &\le c_3\big(t - s + (t-s)^\delta \|G\|_{L_p^q(T)} + (t-s)^{\frac{1}{2}}\big).\end{aligned}$$

Hence, again by the Markov inequality, (3.10) 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 for any $t \in [0,T]$, $\lim_{n\to\infty}(\tilde{X}_t^n, \tilde{W}_t^n) = (\tilde{X}_t, \tilde{W}_t)$ in the probability $\tilde{\mathbb{P}}$. As in [6], let $\tilde{\mathscr{F}}_t^n$ be the completion of the σ-algebra generated by the $\{\tilde{X}_s^n, \tilde{W}_s^n : s \le t\}$. Then as shown in [6], $\tilde{X}_t^n$ is $\tilde{\mathscr{F}}_t^n$-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 $(\tilde{\Omega}, \{\tilde{\mathscr{F}}_t^n\}_{t\in[0,T]}, \tilde{\mathbb{P}})$, and $(\tilde{X}_t^n, \tilde{W}_t^n)_{t\in[0,T]}$ solves the SDE

$$\mathrm{d}\tilde{X}_t^n = b_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}|\tilde{\mathbb{P}})\,\mathrm{d}t + \sigma_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}|\tilde{\mathbb{P}})\,\mathrm{d}\tilde{W}_t^n, \quad \mathscr{L}_{\tilde{X}_0^n}|\tilde{\mathbb{P}} = \mathscr{L}_{X_0}|\mathbb{P}. \tag{4.4}$$

Simply denote $\mathscr{L}_{\tilde{X}_t^n}|\tilde{\mathbb{P}} = \mathscr{L}_{\tilde{X}_t^n}$ and $\mathscr{L}_{\tilde{X}_t}|\tilde{\mathbb{P}} = \mathscr{L}_{\tilde{X}_t}$. Then $(\tilde{X}_t, \tilde{W}_t)_{t\in[0,T]}$ is a weak solution to (1.1) provided for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \tilde{\mathbb{P}}\left(\sup_{s\in[0,T]} \int_0^s |b_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}) - b_t(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t \geq \varepsilon\right) = 0, \tag{4.5}$$

and

$$\lim_{n\to\infty} \tilde{\mathbb{P}}\left(\sup_{s\in[0,T]} \left|\int_0^s \sigma_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n})\mathrm{d}\tilde{W}_t^n - \int_0^s \sigma_t(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t})\,\mathrm{d}\tilde{W}_t\right| \geq \varepsilon\right) = 0. \tag{4.6}$$

In the following we prove these two limits respectively.

Proof of (4.5). For any $n \geq m \geq 1$, we have

$$\int_0^s |b_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}) - b_t(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t \leq I_1(s) + I_2(s) + I_3(s),$$

where

$$\begin{aligned}
I_1(s) &:= \int_0^s |b_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}) - b_t^m(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t,\\
I_2(s) &:= \int_0^s |b_t^m(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t}) - b_t^m(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t,\\
I_3(s) &:= \int_0^s |b_t^m(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t}) - b_t(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t})|\,\mathrm{d}t.
\end{aligned}$$

Below we estimate these $I_i(s)$ respectively.

Firstly, by Chebyshev's inequality, $(H^\theta)(2)$ and (4.2), we arrive at

$$\begin{aligned}
\tilde{\mathbb{P}}(\sup_{s\in[0,T]} I_1(s) \geq \frac{\varepsilon}{3}) &\leq \frac{9}{\varepsilon^2}\mathbb{E}\int_0^T 1_{\{|\tilde{X}_t^n|\leq R\}}|b_t^n(\tilde{X}_t^n, \tilde{\mu}_t^n) - b_t^m(\tilde{X}_t^n, \tilde{\mu}_t)|^2\,\mathrm{d}t\\
&\quad + \frac{9}{\varepsilon^2}\mathbb{E}\int_0^T 1_{\{|\tilde{X}_t^n|>R\}}|b_t^n(\tilde{X}_t^n, \tilde{\mu}_t^n) - b_t^m(\tilde{X}_t^n, \tilde{\mu}_t)|^2\,\mathrm{d}t\\
&\leq \frac{9C}{\varepsilon^2}\left(\int_0^T \left(\int_{|x|\leq R} |b_t^n(x, \tilde{\mu}_t^n) - b_t^m(x, \tilde{\mu}_t)|^{2p}\mathrm{d}x\right)^{q/p}\mathrm{d}t\right)^{\frac{1}{q}}
\end{aligned}$$

$$+\frac{36K}{\varepsilon^2}\int_0^T \tilde{\mathbb{P}}(|\tilde X_t^n|>R)\mathrm{d}t+\frac{36C}{\varepsilon^2}\|G1_{\{|\cdot|>R\}}\|_{L_p^q(T)}.$$

Since $\tilde X_t^n$ converges to $\tilde X_t$ in probability, (4.3) implies

$$\lim_{n\to\infty}\mathbb{W}_\theta(\tilde\mu_t^n,\mu_t)=0,$$

and

$$\lim_{n\to\infty}\tilde{\mathbb{P}}(|\tilde X_t^n|>R)\le\tilde{\mathbb{P}}(|\tilde X_t|\ge R).$$

Then it follows from (H^θ) (1) and (3) that

$$\lim_{n\to\infty}|b_t^n(x,\tilde\mu_t^n)-b_t(x,\tilde\mu_t)|=0,\quad a.e.\ \ t\in[0,T],x\in\mathbb{R}^d.$$

So, by condition (2) in (H^θ), we may apply the dominated convergence theorem to derive

$$\begin{aligned}&\limsup_{n\to\infty}\tilde{\mathbb{P}}(\sup_{s\in[0,T]}I_1(s)\ge\frac{\varepsilon}{3})\\ &\le\frac{9C}{\varepsilon^2}\left(\int_0^T\left(\int_{|x|\le R}|b_t(x,\tilde\mu_t)-b_t^m(x,\tilde\mu_t)|^{2p}\mathrm{d}x\right)^{q/p}\mathrm{d}t\right)^{\frac1q}\\ &+\frac{36K}{\varepsilon^2}\int_0^T\tilde{\mathbb{P}}(|\tilde X_t|\ge R)\mathrm{d}t+\frac{36C}{\varepsilon^2}\|G1_{\{|\cdot|>R\}}\|_{L_p^q(T)}.\end{aligned}\tag{4.7}$$

Since b^m is bounded and continuous, it follows that

$$\limsup_{n\to\infty}\tilde{\mathbb{P}}\Big(\sup_{s\in[0,T]}I_2(s)\ge\frac{\varepsilon}{3}\Big)\le\limsup_{n\to\infty}\frac{3}{\varepsilon}\mathbb{E}\int_0^T|b_t^m(\tilde X_t^n,\mathscr{L}_{\tilde X_t})-b_t^m(\tilde X_t,\mathscr{L}_{\tilde X_t})|\,\mathrm{d}t=0.$$

Finally, since $\tilde X_t^n\to\tilde X_t$ in probability, estimate (4.2) also holds for $\tilde X$ replacing $\tilde X^n$. Therefore, inequality (4.7) holds for I_3 replacing I_1. In conclusion, we arrive at

$$\begin{aligned}&\limsup_{n\to\infty}\tilde{\mathbb{P}}\Big(\sup_{s\in[0,T]}\int_0^s|b_t^n(\tilde X_t^n,\mathscr{L}_{\tilde X_t^n})-b_t(\tilde X_t,\mathscr{L}_{\tilde X_t})|\,\mathrm{d}t\ge\varepsilon\Big)\\ &\le\limsup_{n\to\infty}\sum_{i=1}^3\tilde{\mathbb{P}}\Big(\sup_{s\in[0,T]}I_i(s)\ge\frac{\varepsilon}{3}\Big)\\ &\le\frac{18C}{\varepsilon^2}\left(\int_0^T\left(\int_{|x|\le R}|b_t(x,\tilde\mu_t)-b_t^m(x,\tilde\mu_t)|^{2p}\mathrm{d}x\right)^{q/p}\mathrm{d}t\right)^{\frac1q}\\ &+\frac{72K}{\varepsilon^2}\int_0^T\tilde{\mathbb{P}}(|\tilde X_t|\ge R)\mathrm{d}t+\frac{72C}{\varepsilon^2}\|G1_{\{|\cdot|>R\}}\|_{L_p^q(T)}\end{aligned}$$

for any $m > 0$ and $R > 0$. Then letting first $m \to \infty$ and then $R \to \infty$, due to (1) and (2) in (H^θ), we obtain from the dominated convergence theorem that

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

□

Proof of (4.6). For any $n \geq m \geq 1$ we have

$$\begin{aligned}
&\left| \int_0^s \sigma_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}) \mathrm{d}\tilde{W}_t^n - \int_0^s \sigma_t(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t}) \,\mathrm{d}\tilde{W}_t \right| \\
&\leq \left| \int_0^s \sigma_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}) \mathrm{d}\tilde{W}_t^n - \int_0^s \sigma_t^m(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^m}) \,\mathrm{d}\tilde{W}_t^n \right| \\
&+ \left| \int_0^s \sigma_t^m(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^m}) \mathrm{d}\tilde{W}_t^n - \int_0^s \sigma_t^m(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t^m}) \,\mathrm{d}\tilde{W}_t \right| \\
&+ \left| \int_0^s \sigma_t^m(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t^m}) \mathrm{d}\tilde{W}_t - \int_0^s \sigma_t(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t}) \,\mathrm{d}\tilde{W}_t \right| \\
&=: J_1(s) + J_2(s) + J_3(s).
\end{aligned}$$

By Chebyshev's inequality, BDG inequality and (4.2), we have

$$\begin{aligned}
\tilde{\mathbb{P}}\Big(\sup_{s\in[0,T]} J_1(s) \geq \frac{\varepsilon}{3} \Big) &\leq \frac{9}{\varepsilon^2} \mathbb{E} \int_0^T 1_{\{|\tilde{X}_t^n| \leq R\}} \|\sigma_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}) - \sigma_t^m(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^m})\|_{HS}^2 \,\mathrm{d}t \\
&+ \frac{9}{\varepsilon^2} \mathbb{E} \int_0^T 1_{\{|\tilde{X}_t^n| > R\}} \|\sigma_t^n(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^n}) - \sigma_t^m(\tilde{X}_t^n, \mathscr{L}_{\tilde{X}_t^m})\|_{HS}^2 \,\mathrm{d}t \\
&\leq \frac{9C}{\varepsilon^2} \left(\int_0^T \left(\int_{|x|\leq R} \|\sigma_t^n(x, \tilde{\mu}_t^n) - \sigma_t^m(x, \tilde{\mu}_t^m)\|_{HS}^{2p} \mathrm{d}x \right)^{\frac{q}{p}} \mathrm{d}t \right)^{\frac{1}{q}} \\
&+ \frac{18dK}{\varepsilon^2} \int_0^T \tilde{\mathbb{P}}(|\tilde{X}_t^n| > R) \mathrm{d}t.
\end{aligned}$$

By condition (1) in (H^θ), and $\tilde{\mu}_t^n \to \tilde{\mu}_t$ in $\mathscr{P}_\theta$ as observed above, we have

$$\lim_{n\to\infty} \|\sigma_t^n(x, \tilde{\mu}_t^n) - \sigma_t(x, \tilde{\mu}_t)\| = 0,$$

and

$$\lim_{n\to\infty} \tilde{\mathbb{P}}(|\tilde{X}_t^n| > R) \leq \tilde{\mathbb{P}}(|\tilde{X}_t| \geq R).$$

So, the dominated convergence theorem gives

$$
\begin{aligned}
&\limsup_{n\to\infty}\tilde{\mathbb{P}}\Big(\sup_{s\in[0,T]} J_1(s)\geq\frac{\varepsilon}{3}\Big)\\
&\leq\frac{9C}{\varepsilon^2}\left(\int_0^T\left(\int_{|x|\leq R}\|\sigma_t(x,\tilde{\mu}_t)-\sigma_t^m(x,\tilde{\mu}_t^m)\|_{HS}^{2p}\mathrm{d}x\right)^{\frac{q}{p}}\mathrm{d}t\right)^{\frac{1}{q}}\\
&+\frac{18dK}{\varepsilon^2}\int_0^T\tilde{\mathbb{P}}(|\tilde{X}_t|>R)\mathrm{d}t.
\end{aligned}
\tag{4.8}
$$

Similarly,

$$
\begin{aligned}
&\tilde{\mathbb{P}}\Big(\sup_{s\in[0,T]} J_3(s)\geq\frac{\varepsilon}{3}\Big)\\
&\leq\frac{9C}{\varepsilon^2}\left(\int_0^T\left(\int_{|x|\leq R}\|\sigma_t(x,\tilde{\mu}_t)-\sigma_t^m(x,\tilde{\mu}_t^m)\|_{HS}^{2p}\mathrm{d}x\right)^{\frac{q}{p}}\mathrm{d}t\right)^{\frac{1}{q}}\\
&+\frac{18dK}{\varepsilon^2}\int_0^T\tilde{\mathbb{P}}(|\tilde{X}_t|>R)\mathrm{d}t.
\end{aligned}
$$

To deal with $J_2(s)$, applying Lemma 3.3 to

$$\eta_n(t):=\sigma_t^m(\tilde{X}_t^n,\tilde{\mu}_t^m),\quad \eta(t):=\sigma_t^m(\tilde{X}_t,\tilde{\mu}_t^m),$$

we conclude that when $n\to\infty$,

$$\int_0^s\sigma_t^m(\tilde{X}_t^n,\mathscr{L}_{\tilde{X}_t^m})\mathrm{d}\tilde{W}_t^n\to\int_0^s\sigma_t^m(\tilde{X}_t,\mathscr{L}_{\tilde{X}_t^m})\,\mathrm{d}\tilde{W}_t$$

in probability $\tilde{\mathbb{P}}$, uniformly in $s\in[0,T]$. Hence,

$$
\begin{aligned}
&\lim_{n\to\infty}\tilde{\mathbb{P}}\left(\sup_{s\in[0,T]}\left|\int_0^s\sigma_t^n(\tilde{X}_t^n,\mathscr{L}_{\tilde{X}_t^n})\mathrm{d}\tilde{W}_t^n-\int_0^s\sigma_t(\tilde{X}_t,\mathscr{L}_{\tilde{X}_t})\,\mathrm{d}\tilde{W}_t\right|\geq\varepsilon\right)\\
&\leq\frac{18C}{\varepsilon^2}\left(\int_0^T\left(\int_{|x|\leq R}\|\sigma_t(x,\tilde{\mu}_t)-\sigma_t^m(x,\tilde{\mu}_t^m)\|_{HS}^{2p}\mathrm{d}x\right)^{\frac{q}{p}}\mathrm{d}t\right)^{\frac{1}{q}}\\
&+\frac{36dK}{\varepsilon^2}\int_0^T\tilde{\mathbb{P}}(|\tilde{X}_t|>R)\mathrm{d}t.
\end{aligned}
$$

Letting first $m\to\infty$ and then $R\to\infty$, we prove that when $n\to\infty$,

$$\int_0^s\sigma_t^n(\tilde{X}_t^n,\mathscr{L}_{\tilde{X}_t^n})\mathrm{d}\tilde{W}_t^n\to\int_0^s\sigma_t(\tilde{X}_t,\mathscr{L}_{\tilde{X}_t})\,\mathrm{d}\tilde{W}_t$$

in probability $\tilde{\mathbb{P}}$, uniformly in $s\in[0,T]$. $\square$

4.2 Proof of Theorem 2.1(3)

We will use the following result for the maximal operator:

$$\mathscr{M} h(x) := \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} h(y)\mathrm{d}y, \quad h \in L^1_{loc}(\mathbb{R}^d), x \in \mathbb{R}^d, \tag{4.9}$$

where $B(x,r) := \{y : |x-y| < r\}$, see [3, Appendix A].

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

$$|f(x) - f(y)| \le C|x-y|(\mathscr{M}|\nabla f|(x) + \mathscr{M}|\nabla f|(y)), \quad \text{a.e. } x, y \in \mathbb{R}^d. \tag{4.10}$$

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

$$\|\mathscr{M} f\|_{L^p} \le C_p \|f\|_{L^p}, \quad f \in L^p(\mathbb{R}^d). \tag{4.11}$$

Let X and Y be two solutions to (1.1) with $X_0 = Y_0$, and let $\mu_t = \mathscr{L}_{X_t}, \nu_t = \mathscr{L}_{Y_t}, t \in [0,T]$. Then $\mu_0 = \nu_0$. Let

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

and define b_t^ν, σ_t^ν in the same way using ν_t replacing μ_t. Then

$$\begin{aligned} \mathrm{d}X_t &= b_t^\mu(X_t)\,\mathrm{d}t + \sigma_t^\mu(X_t)\,\mathrm{d}W_t, \\ \mathrm{d}Y_t &= b_t^\nu(Y_t)\mathrm{d}t + \sigma_t^\nu(Y_t)\mathrm{d}W_t. \end{aligned} \tag{4.12}$$

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

$$\frac{\partial u_t}{\partial t} + \frac{1}{2}\mathrm{Tr}(\sigma_t^\mu (\sigma_t^\mu)^* \nabla^2 u_t) + \nabla_{b_t^\mu} u_t + b_t^\mu = \lambda u_t, \quad u_T = 0. \tag{4.13}$$

By [20, Theorem 5.1], when λ is large enough, (4.13) has a unique solution $\mathbf{u}^{\lambda,\mu}$ satisfying

$$\|\nabla \mathbf{u}^{\lambda,\mu}\|_\infty \le \frac{1}{5}, \tag{4.14}$$

and

$$\|\nabla^2 \mathbf{u}^{\lambda,\mu}\|_{L^{2q}_{2p}(T)} < \infty. \tag{4.15}$$

Let $\theta_t^{\lambda,\mu}(x) = x + \mathbf{u}_t^{\lambda,\mu}(x)$. By (4.12), (4.13), and using the Itô formula and an approximation technique (see [20, Lemma 4.3] for more details), we derive

$$\mathrm{d}\theta_t^{\lambda,\mu}(X_t) = \lambda \mathbf{u}_t^{\lambda,\mu}(X_t)\mathrm{d}t + (\nabla \theta_t^{\lambda,\mu} \sigma_t^\mu)(X_t)\,\mathrm{d}W_t, \tag{4.16}$$

and

$$\begin{aligned} \mathrm{d}\theta_t^{\lambda,\mu}(Y_t) = {} & \lambda \mathbf{u}_t^{\lambda,\mu}(Y_t)\mathrm{d}t + (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\nu})(Y_t)\,\mathrm{d}W_t + [\nabla\theta_t^{\lambda,\mu}(b_t^{\nu} - b_t^{\mu})](Y_t)\mathrm{d}t \\ & + \frac{1}{2}\mathrm{Tr}[(\sigma_t^{\nu}(\sigma_t^{\nu})^* - \sigma_t^{\mu}(\sigma_t^{\mu})^*)\nabla^2\mathbf{u}_t^{\lambda,\mu}](Y_t)\mathrm{d}t. \end{aligned} \tag{4.17}$$

Let $\xi_t = \theta_t^{\lambda,\mu}(X_t) - \theta_t^{\lambda,\mu}(Y_t)$. By (4.16), (4.17) and Itô formula, we obtain

$$\begin{aligned} \mathrm{d}|\xi_t|^2 = {} & 2\lambda \left\langle \xi_t, \mathbf{u}_t^{\lambda,\mu}(X_t) - \mathbf{u}_t^{\lambda,\mu}(Y_t) \right\rangle \mathrm{d}t \\ & + 2\left\langle \xi_t, [(\nabla\theta_t^{\lambda,\mu}\sigma_t^{\mu})(X_t) - (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\nu})(Y_t)]\mathrm{d}W_t \right\rangle \\ & + \left\| (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\mu})(X_t) - (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\nu})(Y_t) \right\|_{HS}^2 \mathrm{d}t \\ & - 2\left\langle \xi_t, [\nabla\theta_t^{\lambda,\mu}(b_t^{\nu} - b_t^{\mu})](Y_t) \right\rangle \mathrm{d}t \\ & - \left\langle \xi_t, \mathrm{Tr}[(\sigma_t^{\nu}(\sigma_t^{\nu})^* - \sigma_t^{\mu}(\sigma_t^{\mu})^*)\nabla^2\mathbf{u}_t^{\lambda,\mu}](Y_t) \right\rangle \mathrm{d}t. \end{aligned}$$

So, for any $m \geq 1$,

$$\begin{aligned} \mathrm{d}|\xi_t|^{2m} = {} & 2m\lambda|\xi_t|^{2(m-1)} \left\langle \xi_t, \mathbf{u}_t^{\lambda,\mu}(X_t) - \mathbf{u}_t^{\lambda,\mu}(Y_t) \right\rangle \mathrm{d}t \\ & + 2m|\xi_t|^{2(m-1)} \left\langle \xi_t, [(\nabla\theta_t^{\lambda,\mu}\sigma_t^{\mu})(X_t) - (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\nu})(Y_t)]\mathrm{d}W_t \right\rangle \\ & + m|\xi_t|^{2(m-1)} \left\| (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\mu})(X_t) - (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\nu})(Y_t) \right\|_{HS}^2 \mathrm{d}t \\ & + 2m(m-1)|\xi_t|^{2(m-2)} \left| [(\nabla\theta_t^{\lambda,\mu}\sigma_t^{\mu})(X_t) - (\nabla\theta_t^{\lambda,\mu}\sigma_t^{\nu})(Y_t)]^*\xi_t \right|^2 \mathrm{d}t \\ & - 2m|\xi_t|^{2(m-1)} \left\langle \xi_t, [\nabla\theta_t^{\lambda,\mu}(b_t^{\nu} - b_t^{\mu})](Y_t) \right\rangle \mathrm{d}t \\ & - m|\xi_t|^{2(m-1)} \left\langle \xi_t, \mathrm{Tr}[(\sigma_t^{\nu}(\sigma_t^{\nu})^* - \sigma_t^{\mu}(\sigma_t^{\mu})^*)\nabla^2\mathbf{u}_t^{\lambda,\mu}](Y_t) \right\rangle \mathrm{d}t. \end{aligned} \tag{4.18}$$

By (4.14), we may find out a constant $c_0 > 0$ such that

$$|\xi_t|^{2(m-1)}|\xi_t| \cdot |\mathbf{u}_t^{\lambda,\mu}(X_t) - \mathbf{u}_t^{\lambda,\mu}(Y_t)| \leq c_0|\xi_t|^{2m}. \tag{4.19}$$

According to (2.5), (4.14), the boundedness of σ due to (H^θ)(1)-(2), Lemma 4.1, and noting that the distributions of X_t and Y_t are absolutely continuous with respect to the

Lebesgue measure, for large enough constant $c_1 > 0$ we have

$$
\begin{aligned}
&|\xi_t|^{2(m-2)}\Big|[(\nabla\theta_t^{\lambda,\mu}\sigma_t^\mu)(X_t)-(\nabla\theta_t^{\lambda,\mu}\sigma_t^\nu)(Y_t)]^*\xi_t\Big|^2\\
&\le |\xi_t|^{2(m-1)}\Big\|(\nabla\theta_t^{\lambda,\mu}\sigma_t^\mu)(X_t)-(\nabla\theta_t^{\lambda,\mu}\sigma_t^\nu)(Y_t)\Big\|_{HS}^2\\
&\le |\xi_t|^{2(m-1)}\Big\{C|\xi_t|\mathscr{M}\big(\|\nabla^2\theta_t^{\lambda,\mu}\|+\|\nabla\sigma_t^\mu\|\big)(X_t)\\
&\qquad\qquad + C|\xi_t|\mathscr{M}\big(\|\nabla^2\theta_t^{\lambda,\mu}\|+\|\nabla\sigma_t^\mu\|\big)(Y_t)+\mathbb{W}_\theta(\mu_t,\nu_t)\Big\}^2\\
&\le c_1|\xi_t|^{2m}\big\{\mathscr{M}\big(\|\nabla^2\theta_t^{\lambda,\mu}\|+\|\nabla\sigma_t^\mu\|\big)(X_t)+\mathscr{M}\big(\|\nabla^2\theta_t^{\lambda,\mu}\|+\|\nabla\sigma_t^\mu\|\big)(Y_t)\big\}^2\\
&\quad + c_1|\xi_t|^{2m}+c_1\mathbb{W}_\theta(\mu_t,\nu_t)^{2m},
\end{aligned}
\tag{4.20}
$$

$$
\begin{aligned}
&|\xi_t|^{2(m-1)}|\xi_t|\cdot|\{\nabla\theta_t^{\lambda,\mu}(b_t^\nu-b_t^\mu)\}(Y_t)|\\
&\le L\|\nabla\theta^{\lambda,\mu}\|_{T,\infty}|\xi_t|^{2(m-1)}|\xi_t|\mathbb{W}_\theta(\mu_t,\nu_t)\le c_1\big(|\xi_t|^{2m}+\mathbb{W}_\theta(\mu_t,\nu_t)^{2m}\big),
\end{aligned}
\tag{4.21}
$$

and

$$
\begin{aligned}
&|\xi_t|^{2(m-1)}|\xi_t|\cdot\big|\mathrm{Tr}[(\sigma_t^\nu(\sigma_t^\nu)^*-\sigma_t^\mu(\sigma_t^\mu)^*)\nabla^2\mathbf{u}_t^{\lambda,\mu}](Y_t)\big|\\
&\le c_1|\xi_t|^{2m}|\|\nabla^2\mathbf{u}_t^{\lambda,\mu}\|^{\frac{2m}{2m-1}}(Y_t)+c_1\mathbb{W}_\theta(\mu_t,\nu_t)^{2m}.
\end{aligned}
\tag{4.22}
$$

Substituting (4.19)-(4.22) into (4.18), and noting that $\frac{2m}{2m-1}\le 2$, we arrive at

$$
\mathrm{d}|\xi_t|^{2m}\le c_2|\xi_t|^{2m}\mathrm{d}A_t+c_2\mathbb{W}_\theta(\mu_t,\nu_t)^{2m}\mathrm{d}t+\mathrm{d}M_t
\tag{4.23}
$$

for some constant $c_2 > 0$, a local martingale M_t, and

$$
\begin{aligned}
A_t := \int_0^t \Big\{&1+|\nabla^2\mathbf{u}_s^{\lambda,\mu}(Y_s)|^2+\big(\mathscr{M}\big(\|\nabla^2\theta_s^{\lambda,\mu}\|+\|\nabla\sigma_s^\mu\|\big)(X_s)\\
&+\mathscr{M}\big(\|\nabla^2\theta_s^{\lambda,\mu}\|+\|\nabla\sigma_s^\mu\|\big)(Y_s)\big)^2\Big\}\mathrm{d}s.
\end{aligned}
$$

By the stochastic Grönwall lemma due to [19, Lemma 3.8], when $2m > \theta$ this implies

$$
\mathbb{W}_\theta(\mu_t,\nu_t)^{2m}\le c_3(\mathbb{E}|\xi_t|^\theta)^{\frac{2m}{\theta}}\le c_4\big(\mathbb{E}\mathrm{e}^{\frac{c_2\theta}{2m-\theta}A_t}\big)^{\frac{2m-\theta}{\theta}}\int_0^t\mathbb{W}_\theta(\mu_s,\nu_s)^{2m}\mathrm{d}s,\quad t\in[0,T]
\tag{4.24}
$$

for some constants $c_3, c_4 > 0$. Since by Lemma 3.1, (4.11), (4.15) and the Khasminskii type estimate, see for instance [19, Lemma 3.5], we have

$$
\mathbb{E}\mathrm{e}^{\frac{c_2\theta}{2m-\theta}A_T}<\infty,
$$

so that by the Grönwall lemma we prove $\mathbb{W}_\theta(\mu_t,\nu_t)=0$ for all $t\in[0,T]$. Then by (4.12) both X_t and Y_t solve the same SDE with coefficients b_t^μ and σ_t^μ, and due to [20, Theorem 1.3], the condition $1_D(|b_t^\mu|^2+|\nabla\sigma_t^\mu|^2)\in L_p^q(T)$ for compact $D\subset\mathbb{R}^d$ implies the pathwise uniqueness of this SDE, so we conclude that $X_t=Y_t$ for all $t\in[0,T]$.

Remark 4.2. *As an essential difference between the present argument and that in [20], the PDE* (4.13) *we considered depends on the distribution* μ*. In [20, Theorem 5.1] the PDE*

$$\frac{\partial u_t}{\partial t} + \frac{1}{2}\mathrm{Tr}(\sigma_t\sigma_t^*\nabla^2 u_t) + \nabla_{b_t}u_t + b_t = 0, \quad u_T = 0 \tag{4.25}$$

is used for small enough T *ensuring* $\sup_{t\in[0,T_0],x\in\mathbb{R}^d}\|\nabla u_t(x)\| < 1$*. This is equivalent to taking large enough* $\lambda > 0$ *ensuring* (4.14) *in the present situation.*

4.3 Proof of Corollary 2.2 and Corollary 2.3

Proof of Corollary 2.2. We set $a_t(x,\mu) := (\sigma\sigma^*)_t(x,\mu)$ for $t \in [0,T]$, and $b_t(x,\mu) := 0$, $a_t(x,\mu) := I$ for $t \in \mathbb{R}\setminus[0,T]$. Let $0 \le \rho \in C_0^\infty(\mathbb{R}\times\mathbb{R}^d)$ with support contained in $\{(r,x) : |(r,x)| \le 1\}$ such that $\int_{\mathbb{R}\times\mathbb{R}^d}\rho(r,x)\mathrm{d}r\mathrm{d}x = 1$. For any $n \ge 1$, let $\rho_n(r,x) = n^{d+1}\rho(nr,nx)$ and define

$$\begin{aligned} a_t^n(x,\mu) &= \int_{\mathbb{R}\times\mathbb{R}^d}\sigma_s\sigma_s^*(x',\mu)\rho_n(t-s,x-x')\mathrm{d}s\mathrm{d}x', \\ b_t^n(x,\mu) &= \int_{\mathbb{R}\times\mathbb{R}^d} b_s(x',\mu)\rho_n(t-s,x-x')\mathrm{d}s\mathrm{d}x', \quad (t,x,\mu) \in \mathbb{R}\times\mathbb{R}^d\times\mathscr{P}. \end{aligned} \tag{4.26}$$

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

$$\mathrm{d}X_t = b_t(X_t,\mathscr{L}_{X_t})\mathrm{d}t + \hat\sigma_t(X_t,\mathscr{L}_{X_t})\mathrm{d}W_t. \tag{4.27}$$

We first show that $(b,\hat\sigma)$ satisfies assumption (H^θ). Firstly, (2.6)-(2.7) and the continuity in the third variable of B and Σ imply that b and σ are continuous in the third variable $\mu \in \mathscr{P}_\theta$. Thus, (1) in (H^θ) holds. As to (H^θ) (2), since by [20], it holds that

$$\lim_{n\to\infty}\|F - F*\rho_n\|_{L_p^q(T)} = 0,$$

there exists a subsequence n_k such that

$$\|F - F*\rho_{n_k}\|_{L_p^q(T)} < 2^{-k}.$$

Letting

$$G = \sum_{k=1}^\infty |F - F*\rho_{n_k}| + F,$$

then $\|G\|_{L_p^q(T)} \le 1 + \|F\|_{L_p^q(T)}$ and noting $|b^{n_k}|^2 \le K + F*\rho_{n_k}$, we have $|b^{n_k}|^2 \le K + G$. So, using the subsequence b^{n_k} replacing b^n, we verify condition (2) in (H^θ). Finally, by (2.6), for any $n \ge 1$ there exists a constant $c_n > 0$ such that

$$|b_t^n(x,\mu) - b_s^n(x',\nu)| + \|\hat\sigma_t^n(x,\mu) - \hat\sigma_s^n(x',\nu)\| \le c_n\big(|t-s| + |x-x'| + \mathbb{W}_1(\mu,\nu)\big)$$

holds for all $s,t\in\mathbb{R}, x,x'\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_1$. So, for any $\theta\geq 1$, condition (3) in (H^θ) holds. By Theorem 2.1 (1), SDE (4.27) has a weak solution. Noting that $\sigma\sigma^*=\hat\sigma\hat\sigma^*$, the SDE (1.1) also has a weak solution. Finally, the strong existence and uniqueness follow from Theorem 2.1 (2) and (3). □

Proof of Corollary 2.3. Let b_t^n and a_t^n be in (4.26), and let $\hat\sigma_t^n=\sqrt{a_t^n}$ and $\hat\sigma_t=\sqrt{a_t}$. Then (2.5) and (4.26) imply $(b,\hat\sigma)$ satisfy H^θ. Then we may complete the proof as in the proof of Corollary 2.2 (1). □

5 Proofs of Theorems 2.5-2.6

5.1 Proof of Theorem 2.5

By [18, Theorem 1.2 (2)] with $d_1=0$, we know that (3.11) has a unique strong solution X_t up to life time. Combining this with Corollary 2.3, Lemma 3.4 and **(H)**, we see that the SDE (1.1) has strong existence and uniqueness. For any $\mu\in\mathscr{P}_2$ we let $\mu_t=P_t^*\mu$ be the distribution of X_t which solves (2.9) with $\mathscr{L}_{X_0}=\mu$.

We first figure out the outline of proof using coupling by change of measure as in [13, 15]. From now on, we fix $t_0\in(0,T]$ and $\mu_0,\nu_0\in\mathscr{P}_2$, and take $\mathscr{F}_0$-measurable variables X_0 and Y_0 in $\mathbb{R}^d$ such that $\mathscr{L}_{X_0}=\mu_0,\mathscr{L}_{Y_0}=\nu_0$ and

$$\mathbb{E}|X_0-Y_0|^2=\mathbb{W}_2(\mu_0,\nu_0)^2. \tag{5.1}$$

Let X_t with $\mathscr{L}_{X_0}=\mu_0$ solve (2.9), we have

$$\mathrm{d}X_t=b_t(X_t,\mu_t)\mathrm{d}t+\sigma_t(X_t)\mathrm{d}W_t. \tag{5.2}$$

To establish the log-Harnack inequality, We construct a process Y_t such that for a weighted probability measure $\mathbb{Q}:=R\mathbb{P}$

$$X_{t_0}=Y_{t_0}\ \mathbb{Q}\text{-a.s.},\quad \text{and } \mathscr{L}_{Y_{t_0}}|\mathbb{Q}=P_{t_0}^*\nu_0=:\nu_{t_0}. \tag{5.3}$$

Then

$$(P_{t_0}f)(\nu_0)=\mathbb{E}_{\mathbb{Q}}[f(Y_{t_0})]=\mathbb{E}[R_{t_0}f(X_{t_0})],\quad f\in\mathscr{B}_b(\mathbb{R}^d).$$

So, by Young's inequality we obtain the log-Harnack inequality

$$\begin{aligned}(P_{t_0}\log f)(\nu_0)&\leq\mathbb{E}[R_{t_0}\log R_{t_0}]+\log\mathbb{E}[f(X_{t_0})]\\&=\log(P_{t_0}f)(\mu_0)+\mathbb{E}[R_{t_0}\log R_{t_0}],\quad f\in\mathscr{B}_b^+(\mathbb{R}^d), f\geq 1.\end{aligned} \tag{5.4}$$

Moreover, by the Hölder inequality, for any $p>1$ we have

$$\begin{aligned}(P_{t_0}f)^p(\nu_0)&=\{\mathbb{E}[R_{t_0}f(X_{t_0})]\}^p\\&\leq(P_{t_0}f^p)(\mu_0)\times\{\mathbb{E}[R_{t_0}^{\frac{p}{p-1}}]\}^{p-1},\quad f\in\mathscr{B}_b^+(\mathbb{R}^d).\end{aligned} \tag{5.5}$$

To construct the desired Y_t, we follow the line of [18] using Zvonkin's transform. As shown in [18, Theorem 3.10] for $d_1 = 0$ that Assumption **(H)** implies that for large enough $\lambda > 0$, the PDE (4.13) has a unique solution $\mathbf{u}^{\lambda,\mu}$ satisfying

$$\|\mathbf{u}^{\lambda,\mu}\|_\infty + \|\nabla \mathbf{u}^{\lambda,\mu}\|_\infty + \|\nabla^2 \mathbf{u}^{\lambda,\mu}\|_\infty \le \frac{1}{5}. \tag{5.6}$$

By combining $\|\nabla^2 \mathbf{u}^{\lambda,\mu}\|_\infty < \infty$ with the Lipschitzian continuity of σ and (4.9), we see that the increasing process A_t in (4.23) satisfies

$$\mathrm{d}A_t \le c\mathrm{d}t$$

for some constant $c > 0$. Moreover, $\mathbb{E}|\xi_t|^2 \ge c'\mathbb{W}_2(\mu_t, \nu_t)^2$ holds for some constant $c' > 0$. So, with $m = 1, \theta = 2, \mathscr{L}_{X_0} = \mu_0$ and $\mathscr{L}_{Y_0} = \nu_0$, the inequality (4.23) gives

$$\mathbb{W}_2(\mu_t, \nu_t) \le \kappa \mathbb{W}_2(\mu_0, \nu_0), \quad t \in [0, T] \tag{5.7}$$

for some constant $\kappa > 0$.

As in [13, §2], let $\gamma = \frac{72}{25}K + \frac{2d}{25\delta} + \frac{12\lambda}{25}$ and take

$$\zeta_t = \frac{12}{25\gamma}\Big(1 - \mathrm{e}^{\frac{25\gamma}{16}(t-t_0)}\Big), \quad t \in [0, t_0], \tag{5.8}$$

and let Y_t solve the modified SDE

$$\mathrm{d}Y_t = \Big\{b_t(Y_t, \nu_t) + \frac{1}{\zeta_t}\sigma_t(Y_t)\sigma_t(X_t)^{-1}(X_t - Y_t)\Big\}\mathrm{d}t + \sigma_t(Y_t)\mathrm{d}W_t, \quad t \in [0, t_0). \tag{5.9}$$

Since $\sup_{t\in[0,T]} \nu_t(|\cdot|^2) < \infty$, this SDE has a unique solution $(Y_t)_{t\in[0,t_0)}$. Let

$$\tau_n := t_0 \wedge \inf\{t \in [0, t_0) : |X_t| + |Y_t| \ge n\}, \quad n \ge 1,$$

where $\inf \emptyset := \infty$ by convention. We have $\tau_n \uparrow t_0$ as $n \uparrow \infty$. To see that the process Y meets the above requirement, we first prove that

$$R_s := \exp\bigg[\int_0^s \frac{1}{\zeta_t}\big\langle \sigma_t(X_t)^{-1}(Y_t - X_t), \mathrm{d}W_t\big\rangle - \frac{1}{2}\int_0^s \frac{|\sigma_t(X_t)^{-1}(Y_t - X_t)|^2}{\zeta_t^2}\mathrm{d}t\bigg] \tag{5.10}$$

for $s \in [0, t_0)$ is a uniformly integrable martingale, and hence extends also to time t_0.

Lemma 5.1. *Assume* **(H)** *and let X_0, Y_0 be two $\mathscr{F}_0$-measurable random variables such that $\mathscr{L}_{X_0} = \mu_0, \mathscr{L}_{Y_0} = \nu_0$, and*

$$\mathbb{E}|X_0 - Y_0|^2 = \mathbb{W}_2(\mu_0, \nu_0)^2. \tag{5.11}$$

Then there exists a constant $c > 0$ uniformly in $t_0 \in (0,T)$ such that

$$\sup_{t\in[0,t_0)} \mathbb{E}[R_t \log R_t] \le \frac{c}{t_0}\mathbb{W}_2(\mu_0,\nu_0)^2. \tag{5.12}$$

Consequently, R_t extends to $t = t_0$, $\mathbb{Q} := R_{t_0}\mathbb{P}$ is a probability measure under which (5.9) *has a unique solution $(Y_t)_{t\in[0,t_0]}$ satisfying*

$$\mathbb{Q}(X_{t_0} = Y_{t_0}) = 1. \tag{5.13}$$

Proof. By **(H)**, for any $n \ge 1$ and $t \in (0,t_0)$, the process $(R_{s\wedge\tau_n})_{s\in[0,t]}$ is a uniformly integrable continuous martingale. So, for the first assertion it suffices to find out a constant $c > 0$ uniformly in $t_0 \in (0,T)$ such that

$$\sup_{n\ge1} \mathbb{E}[R_{t\wedge\tau_n} \log R_{t\wedge\tau_n}] \le \frac{c}{t_0}\mathbb{W}_2(\mu_0,\nu_0)^2, \quad t \in [0,t_0). \tag{5.14}$$

To this end, for fixed $t \in (0,T)$ and $n \ge 1$, we consider the weighted probability $\mathbb{Q}_{t,n} := R_{t\wedge\tau_n}\mathbb{P}$. By Girsanov's theorem $(\tilde W_s)_{s\in[0,t\wedge\tau_n]}$ is a d-dimensional Brownian motion under $\mathbb{Q}_{t,n}$. Reformulating (5.2) and (5.9) as

$$\begin{aligned} \mathrm{d}X_s &= b_s(X_s,\mu_s) - \frac{X_s - Y_s}{\zeta_s}\mathrm{d}s + \sigma_s(X_s)\mathrm{d}\tilde W_s,\\ \mathrm{d}Y_s &= b_s(Y_s,\nu_s) + \sigma_s(Y_s)\mathrm{d}\tilde W_s, \quad s \in [0, t\wedge\tau_n], \end{aligned}$$

where

$$\tilde W_s = W_s + \int_0^s \frac{1}{\zeta_r}\sigma_r(X_r)^{-1}(X_r - Y_r)\mathrm{d}W_r.$$

Let $\mathbf{u}^{\lambda,\mu}$ solve (4.13) and take $\theta_s^{\lambda,\mu}(x) = x + \mathbf{u}_s^{\lambda,\mu}(x)$. By the Itô formula, we have

$$\mathrm{d}\theta_s^{\lambda,\mu}(X_s) = \lambda\mathbf{u}_s^{\lambda,\mu}(X_s)\mathrm{d}s + (\nabla\theta_s^{\lambda,\mu}\sigma_s)(X_s)\,\mathrm{d}\tilde W_s - \nabla\theta_s^{\lambda,\mu}(X_s)\frac{X_s - Y_s}{\zeta_s}\mathrm{d}s, \tag{5.15}$$

and

$$\mathrm{d}\theta_s^{\lambda,\mu}(Y_s) = \lambda\mathbf{u}_s^{\lambda,\mu}(Y_s)\mathrm{d}s + (\nabla\theta_s^{\lambda,\mu}\sigma_s)(Y_s)\,\mathrm{d}\tilde W_s + [\nabla\theta_s^{\lambda,\mu}(b_s^\nu - b_s^\mu)](Y_s)\mathrm{d}s. \tag{5.16}$$

Next, using the Itô formula under the probability $\mathbb{Q}_{t,n}$, we obtain

$$\begin{aligned} &\mathrm{d}|\theta_s^{\lambda,\mu}(Y_s) - \theta_s^{\lambda,\mu}(X_s)|^2\\ &= 2\langle\theta_s^{\lambda,\mu}(X_s) - \theta_s^{\lambda,\mu}(Y_s), \lambda\mathbf{u}_s^{\lambda,\mu}(X_s) - \lambda\mathbf{u}_s^{\lambda,\mu}(Y_s)\rangle\mathrm{d}s\\ &+ 2\langle\theta_s^{\lambda,\mu}(X_s) - \theta_s^{\lambda,\mu}(Y_s), (\nabla\theta_s^{\lambda,\mu}\sigma_s)(X_s)\mathrm{d}\tilde W_s - (\nabla\theta_s^{\lambda,\mu}\sigma_s)(Y_s)\mathrm{d}\tilde W_s\rangle\\ &+ \|\nabla\theta_s^{\lambda,\mu}\sigma_s)(X_s) - \nabla\theta_s^{\lambda,\mu}\sigma_s)(Y_s)\|_{HS}^2\mathrm{d}s\\ &- 2\langle\theta_s^{\lambda,\mu}(X_s) - \theta_s^{\lambda,\mu}(Y_s), [\nabla\theta_s^{\lambda,\mu}(b_s^\nu - b_s^\mu)](Y_s)\mathrm{d}s\rangle\\ &- 2\Big\langle\theta_s^{\lambda,\mu}(X_s) - \theta_s^{\lambda,\mu}(Y_s), \nabla\theta_s^{\lambda,\mu}(X_s)\frac{X_s - Y_s}{\zeta_s}\mathrm{d}s\Big\rangle, \quad s \in [0, t\wedge\tau_n]. \end{aligned} \tag{5.17}$$

Moreover, (5.6) implies

$$
\begin{aligned}
&-\Big\langle \theta_s^{\lambda,\mu}(X_s)-\theta_s^{\lambda,\mu}(Y_s), \nabla\theta_s^{\lambda,\mu}(X_s)\frac{X_s-Y_s}{\zeta_s}\Big\rangle\\
&=-\Big\langle X_s-Y_s+\mathbf{u}_s^{\lambda,\mu}(X_s)-\mathbf{u}_s^{\lambda,\mu}(Y_s), \frac{X_s-Y_s}{\zeta_s}+\nabla\mathbf{u}_s^{\lambda,\mu}(X_s)\frac{X_s-Y_s}{\zeta_s}\Big\rangle\\
&=-\Big\langle X_s-Y_s, \frac{X_s-Y_s}{\zeta_s}\Big\rangle-\Big\langle \mathbf{u}_s^{\lambda,\mu}(X_s)-\mathbf{u}_s^{\lambda,\mu}(Y_s), \frac{X_s-Y_s}{\zeta_s}\Big\rangle\\
&-\Big\langle X_s-Y_s, \nabla\mathbf{u}_s^{\lambda,\mu}(X_s)\frac{X_s-Y_s}{\zeta_s}\Big\rangle-\Big\langle \mathbf{u}_s^{\lambda,\mu}(X_s)-\mathbf{u}_s^{\lambda,\mu}(Y_s), \nabla\mathbf{u}_s^{\lambda,\mu}(X_s)\frac{X_s-Y_s}{\zeta_s}\Big\rangle\\
&\le -\frac{14}{25}\frac{|X_s-Y_s|^2}{\zeta_s}, \quad s\in[0,t\wedge\tau_n].
\end{aligned}
$$

Substituting into (5.17) leads to

$$
\begin{aligned}
&\mathrm{d}|\theta_s^{\lambda,\mu}(Y_s)-\theta_s^{\lambda,\mu}(X_s)|^2\\
&\le \Big\{\gamma|X_s-Y_s|^2+\frac{72}{25}\kappa_2(T)|X_s-Y_s|\mathbb{W}_2(\mu_s,\nu_s)-\frac{4}{5}\frac{|X_s-Y_s|^2}{\zeta_s}\Big\}\mathrm{d}s\\
&\quad+\mathrm{d}M_s, \quad s\in[0,t\wedge\tau_n]
\end{aligned}
\tag{5.18}
$$

for the $\mathbb{Q}_{t,n}$-martingale

$$
M_s:=2\int_0^s\langle\theta_r^{\lambda,\mu}(X_r)-\theta_r^{\lambda,\mu}(Y_r), (\nabla\theta_r^{\lambda,\mu}\sigma_r)(X_r)\mathrm{d}\tilde{W}_r-(\nabla\theta_r^{\lambda,\mu}\sigma_r)(Y_r)\mathrm{d}\tilde{W}_r\rangle.
$$

On the other hand, (5.8) implies

$$
\frac{4}{5}-\gamma\zeta_s+\frac{16}{25}\zeta_s'=\frac{8}{25}.
$$

Combining this with (5.18) and using the Itô formula, we may find out a constant $c_0>0$ such that

$$
\begin{aligned}
&\mathrm{d}\frac{|\theta_s^{\lambda,\mu}(Y_s)-\theta_s^{\lambda,\mu}(X_s)|^2}{\zeta_s}\\
&\le\frac{\mathrm{d}M_s}{\zeta_s}+c_0\mathbb{W}_2(\mu_s,\nu_s)^2\mathrm{d}s-\frac{|X_s-Y_s|^2}{\zeta_s^2}\Big\{\frac{4}{5}-\gamma\zeta_s+\frac{16}{25}\zeta_s'-\frac{1}{25}\Big\}\mathrm{d}s\\
&\le\frac{\mathrm{d}M_s}{\zeta_s}+c_0\mathbb{W}_2(\mu_s,\nu_s)^2\mathrm{d}s-\frac{7|X_s-Y_s|^2}{25\zeta_s^2}, \quad s\in[0,t\wedge\tau_n].
\end{aligned}
\tag{5.19}
$$

Combining this with (5.7) and (5.1), we arrive at

$$
\mathbb{E}_{\mathbb{Q}_{t,n}}\int_0^{t\wedge\tau_n}\frac{|X_s-Y_s|^2}{\zeta_s^2}\mathrm{d}s\le\frac{c_1}{t_0}\mathbb{W}_2(\mu_0,\nu_0)^2, \quad t\in[0,t_0)
\tag{5.20}
$$

for some constant $c_1 > 0$. Therefore, there exists a constant $C > 0$ such that

$$\begin{aligned}\mathbb{E}[R_{t\wedge\tau_n}\log R_{t\wedge\tau_n}] &= \frac{1}{2}\mathbb{E}_{\mathbb{Q}_{t,n}}\int_0^{t\wedge\tau_n}\frac{|\sigma_s(X_s)^{-1}(Y_s-X_s)|^2}{\zeta_s^2}\mathrm{d}s\\ &\le \frac{C}{t_0}\mathbb{W}_2(\mu_0,\nu_0)^2,\quad t\in(0,t_0).\end{aligned}$$

Thus, (5.12) holds.

By (5.12) and the martingale convergence theorem, $(R_t)_{t\in[0,t_0]}$ is a uniformly integrable martingale, so $\mathbb{Q} := R_{t_0}\mathbb{P}$ is a probability measure. By Girsanov theorem, we can reformulate (5.9) as

$$\mathrm{d}Y_t = b_t(Y_t,\nu_t)\mathrm{d}t + \sigma_t(Y_t)\mathrm{d}\tilde{W}_t, \tag{5.21}$$

which has a unique solution $(Y_t)_{t\in[0,t_0]}$. By (5.12),

$$\mathbb{E}_{\mathbb{Q}}\int_0^{t_0}\frac{|X_t-Y_t|^2}{\zeta_t^2}\mathrm{d}t < \infty.$$

Since $X_t - Y_t$ is continuous and $\int_0^{t_0}\frac{1}{\zeta_t}\mathrm{d}t = \infty$, this implies $\mathbb{Q}(X_{t_0} = Y_{t_0}) = 1$. □

Proof of Theorem 2.5. Consider the DDSDE

$$\mathrm{d}\tilde{X}_t = b_t(\tilde{X}_t, \mathscr{L}_{\tilde{X}_t}|\tilde{\mathbb{P}})\mathrm{d}t + \sigma_t(\tilde{X}_t)\mathrm{d}\tilde{W}_t,\quad \tilde{X}_0 = Y_0.$$

By the weak uniqueness we have $\mathscr{L}_{\tilde{X}_t}|\tilde{\mathbb{P}} = P_t^*\nu_0 = \nu_t$ for $t\in[0,t_0]$. Combining this with (5.21) and the strong uniqueness, we conclude that $\tilde{X}_t = Y_t$ for $t\in[0,T]$. Therefore, (5.4) and Lemma 5.1 lead to

$$(P_{t_0}\log f)(\nu_0) \le \log(P_{t_0}f)(\mu_0) + \frac{C}{t_0}\mathbb{W}_2(\mu_0,\nu_0)^2,\quad t_0\in(0,T],$$

which implies (2.13) due to [17, Theorem 4.1 (4.11)].

Finally, by repeating the proof of [15, Lemma 3.4.3] and [15, Proof of Theorem 3.4.1(2)], we may find out constants $p_0 > 1$ and $c_1, c_2 > 0$ such that

$$\big(\mathbb{E}R_{t_0}^{\frac{p_0}{p_0-1}}\big)^{\frac{p_0-1}{p_0}} \le \exp\left\{\frac{c}{t_0\wedge 1}\mathbb{W}_2(\mu_0,\nu_0)^2\right\}\cdot\mathbb{E}\left[\exp\left\{\frac{c|X_0-Y_0|^2}{1-\mathrm{e}^{-c_2 t_0}}\right\}\right]. \tag{5.22}$$

This together with (5.5) and Jense's inequality proves (2.14). □

5.2 Proof of Theorem 2.6

Proof. Fix $t_0 > 0$. Denote $\mu_t = P_t^* \mu_0 = \mathscr{L}_{X_t}, t \in [0, t_0]$. Then (2.15) becomes

$$\mathrm{d}X_t = b_t(X_t, \mu_t)\mathrm{d}t + \sigma_t(\mu_t)\mathrm{d}W_t, \quad \mathscr{L}_{X_0} = \mu_0. \tag{5.23}$$

Let $Y_t = X_t + \frac{tv}{t_0}$, $t \in [0, t_0]$. Then

$$\mathrm{d}Y_t = b_t(Y_t, \mu_t)\mathrm{d}t + \sigma_t(\mu_t)\mathrm{d}\tilde{W}_t, \quad \mathscr{L}_{Y_0} = \mu_0, t \in [0, t_0],$$

where

$$\begin{aligned} \tilde{W}_t &:= W_t + \int_0^t \eta_s \mathrm{d}s, \\ \eta_t &:= \sigma_t^{-1}\Big\{\frac{v}{t_0} + b_t(X_t, \mu_t) - b_t\Big(X_t + \frac{tv}{t_0}, \mu_t\Big)\Big\}. \end{aligned}$$

Let $R_{t_0} = \exp[-\int_0^{t_0} \langle \eta_t, \mathrm{d}W_t\rangle - \frac{1}{2}\int_0^{t_0} |\eta_s|^2 \mathrm{d}s]$. By the Girsanov theorem we obtain

$$(P_{t_0} f)(\mu_0) = \mathbb{E}[R_{t_0} f(Y_{t_0})] = \mathbb{E}[R_{t_0} f(X_{t_0} + v)] \leq (P_{t_0} f^p(v + \cdot))^{\frac{1}{p}}(\mu_0)\big(\mathbb{E}R_{t_0}^{\frac{p}{p-1}}\big)^{\frac{p-1}{p}},$$

and by Young's inequality, we obtain

$$\begin{aligned} (P_{t_0} \log f)(\mu_0) &= \mathbb{E}[R_{t_0} \log f(Y_{t_0})] \\ &= \mathbb{E}[R_{t_0} \log f(X_{t_0} + v)] \leq \log P_{t_0} f(v + \cdot)(\mu_0) + \mathbb{E}R_{t_0} \log R_{t_0}. \end{aligned}$$

Then we have

$$\begin{aligned} \mathbb{E}R_{t_0}^{\frac{p}{p-1}} &\leq \sup_\Omega \mathrm{e}^{\frac{p}{2(p-1)^2}\int_0^{t_0} |\eta_s|^2 \mathrm{d}s} \\ &\leq \exp\left[\frac{p\int_0^{t_0} \|\sigma_t^{-1}\|_\infty^2 \big\{|v|/t_0 + \phi(t|v|/t_0)\big\}^2 \mathrm{d}t}{2(p-1)^2}\right], \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}R_{t_0} \log R_{t_0} = \mathbb{E}_{\mathbb{Q}} \log R_{t_0} \leq \frac{1}{2}\mathbb{E}_{\mathbb{Q}}\int_0^{t_0} |\eta_s|^2 \mathrm{d}s \\ &\leq \frac{1}{2}\int_0^{t_0} \|\sigma_t^{-1}\|_\infty^2 \big\{|v|/t_0 + \phi(t|v|/t_0)\big\}^2 \mathrm{d}t. \end{aligned}$$

□

Acknowledgement. Financial support by the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications" is acknowledged.

References

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

[2] P. E. Chaudru de Raynal, *Strong well-posedness of McKean-Vlasov stochastic differential equation with Hölder drift,* arXiv: 1512.08096.

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

[4] 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.

[5] 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.

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

[7] X. Huang, M. Röckner, F.-Y. Wang, *Nonlinear Fokker–Planck equations for probability measures on path space and path-distribution dependent SDEs,* arXiv:1709.00556.

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

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

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

[11] 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.

[12] F.-Y. Wang, *Logarithmic Sobolev inequalities on noncompact Riemannian manifolds,* Probab. Theory Related Fields 109(1997), 417-424.

[13] F.-Y. Wang, *Harnack inequality for SDE with multiplicative noise and extension to Neumann semigroup on nonconvex manifolds,* Ann. Probab. 39(2011), 1449-1467.

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

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

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

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

[18] F.-Y. Wang, X. C. Zhang, *Degenerate SDE with Hölder-Dini Drift and Non-Lipschitz Noise Coefficient,* SIAM J. Math. Anal. 48(2016), 2189-2226.

[19] L. Xie, X. Zhang, *Ergodicity of stochastic differential equations with jumps and singular coefficients,* arXiv:1705.07402.

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

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