

# Probability Distance Estimates Between Diffusion Processes and Applications to Singular McKean-Vlasov SDEs\*

Xing Huang <sup>a)</sup>, Panpan Ren <sup>b)</sup>, Feng-Yu Wang <sup>a)</sup>

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

b)Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Hong Kong, China

xinghuang@tju.edu.cn, wangfy@tju.edu.cn

## Abstract

The  $L^k$ -Wasserstein distance  $\mathbb{W}_k(k \geq 1)$  and the probability distance  $\mathbb{W}_\psi$  induced by a concave function  $\psi$ , are estimated between different diffusion processes with singular coefficients. As applications, the well-posedness, probability distance estimates and the log-Harnack inequality are derived for McKean-Vlasov SDEs with multiplicative distribution dependent noise, where the coefficients are singular in time-space variables and  $(\mathbb{W}_k + \mathbb{W}_\psi)$ -Lipschitz continuous in the distribution variable. This improves existing results derived in the literature under the  $\mathbb{W}_k$ -Lipschitz or derivative conditions in the distribution variable.

AMS subject Classification: 60H10, 60H15.

Keywords: Probability distance, Diffusion processes, Log-Harnack inequality,

## 1 Introduction

Let  $T > 0$ , and let  $\Xi$  be the space of  $(a, b)$ , where

$$b : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad a : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$$

---

\*Supported in part by the National Key R&D Program of China (2022YFA1006000, 2020YFA0712900) and NNSFC (12271398, 11921001).

are measurable, and for any  $(t, x) \in [0, T] \times \mathbb{R}^d$ ,  $a(t, x)$  is positive definite. For any  $(a, b) \in \Xi$ , consider the time dependent second order differential operator on  $\mathbb{R}^d$ :

$$L_t^{a,b} := \text{tr}\{a(t, \cdot) \nabla^2\} + b(t, \cdot) \cdot \nabla, \quad t \in [0, T].$$

Let  $(a_i, b_i) \in \Xi$ ,  $i = 1, 2$ , such that for any  $s \in [0, T]$ , each  $(L_t^{a_i, b_i})_{t \in [s, T]}$  generates a unique diffusion process  $(X_{s,t}^{i,x})_{(t,x) \in [s, T] \times \mathbb{R}^d}$  on  $\mathbb{R}^d$  with  $X_{s,s}^{i,x} = x$ . Let

$$P_{s,t}^{i,x} := \mathcal{L}_{X_{s,t}^{i,x}}$$

be the distribution of  $X_{s,t}^{i,x}$ . When  $s = 0$ , we simply denote

$$X_{0,t}^{i,x} = X_t^{i,x}, \quad P_{0,t}^{i,x} = P_t^{i,x}.$$

If the initial value is random with distributions  $\gamma \in \mathcal{P}$ , where  $\mathcal{P}$  is the set of all probability measures on  $\mathbb{R}^d$ , we denote the diffusion process by  $X_{s,t}^{i,\gamma}$ , which has distribution

$$(1.1) \quad P_{s,t}^{i,\gamma} = \int_{\mathbb{R}^d} P_{s,t}^{i,x} \gamma(dx), \quad i = 1, 2, \quad 0 \leq s \leq t \leq T.$$

By developing the bi-coupling argument and using an entropy inequality due to [1], the relative entropy

$$\text{Ent}(P_{s,t}^{1,\gamma} | P_{s,t}^{2,\tilde{\gamma}}) := \int_{\mathbb{R}^d} \left( \log \frac{dP_{s,t}^{1,\gamma}}{dP_{s,t}^{2,\tilde{\gamma}}} \right) dP_{s,t}^{1,\gamma}, \quad 0 \leq s < t \leq T, \gamma, \tilde{\gamma} \in \mathcal{P}$$

is estimated in [13], and as an application, the log-Haranck inequality is established for McKean-Vlasov SDEs with multiplicative distribution dependent noise, where the drift is Dini continuous in the spatial variable  $x$ , and the diffusion coefficient is Lipschitz continuous in  $x$  and the distribution variable with respect to  $\mathbb{W}_2$ .

In this paper, we estimate a weighted variational distance between  $P_t^{1,\gamma}$  and  $P_t^{2,\tilde{\gamma}}$  for diffusion processes with singular coefficients, and apply to the study of singular McKean-Vlasov SDEs with multiplicative distribution dependent noise, so that existing results in the literature are considerably extended.

Consider the class

$$\mathcal{A} := \left\{ \psi : [0, \infty) \rightarrow [0, \infty) \text{ is increasing and concave, } \psi(r) > 0 \text{ for } r > 0 \right\}.$$

For any  $\psi \in \mathcal{A}$ , the  $\psi$ -continuity modulus of a function  $f$  on  $\mathbb{R}^d$  is

$$[f]_\psi := \sup_{x \neq y} \frac{|f(x) - f(y)|}{\psi(|x - y|)}.$$

Then

$$\mathcal{P}_\psi := \left\{ \mu \in \mathcal{P} : \|\mu\|_\psi := \int_{\mathbb{R}^d} \psi(|x|) \mu(dx) < \infty \right\}$$

is a complete metric space under the distance  $\mathbb{W}_\psi$  induced by  $\psi$ :

$$\mathbb{W}_\psi(\mu, \nu) := \sup_{[f]_\psi \leq 1} |\mu(f) - \nu(f)|,$$

where  $\mu(f) := \int_{\mathbb{R}^d} f d\mu$  for  $f \in L^1(\mu)$ . In particular,  $\mathbb{W}_\psi = \mathbb{W}_1$  is the  $L^1$ -Wasserstein distance if  $\psi(r) = r$ , while  $\mathbb{W}_\psi$  with  $\psi \equiv 2$  reduces to the total variational distance

$$\|\mu - \nu\|_{var} := \sup_{|f| \leq 1} |\mu(f) - \nu(f)|.$$

For any  $k > 0$ , the  $L^k$ -Wasserstein distance is

$$\mathbb{W}_k(\mu, \nu) := \inf_{\pi \in \mathcal{C}(\mu, \nu)} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^k \pi(dx, dy) \right)^{\frac{1}{1 \vee k}},$$

where  $\mathcal{C}(\mu, \nu)$  is the set of couplings for  $\mu$  and  $\nu$ . Then

$$\mathcal{P}_k := \{\mu \in \mathcal{P} : \mu(|\cdot|^k) < \infty\}$$

is a Polish space under  $\mathbb{W}_k$ . Since  $\psi$  has at most linear growth, we have  $\mathcal{P}_k \subset \mathcal{P}_\psi$ , and  $\mathcal{P}_k$  is complete under  $\mathbb{W}_\psi + \mathbb{W}_k$ .

To characterize the singularity of coefficients in time-space variables, we recall some functional spaces introduced in [17]. For any  $p \geq 1$ ,  $L^p(\mathbb{R}^d)$  is the class of measurable functions  $f$  on  $\mathbb{R}^d$  such that

$$\|f\|_{L^p(\mathbb{R}^d)} := \left( \int_{\mathbb{R}^d} |f(x)|^p dx \right)^{\frac{1}{p}} < \infty.$$

For any  $p, q > 1$  and a measurable function  $f$  on  $[0, T] \times \mathbb{R}^d$ , let

$$\|f\|_{\tilde{L}_q^p(s, t)} := \sup_{z \in \mathbb{R}^d} \left( \int_s^t \|1_{B(z, 1)} f_r\|_{L^p(\mathbb{R}^d)}^q dr \right)^{\frac{1}{q}},$$

where  $B(z, 1) := \{x \in \mathbb{R}^d : |x - z| \leq 1\}$ . When  $s = 0$ , we simply denote  $\|\cdot\|_{\tilde{L}_q^p(t)} = \|\cdot\|_{\tilde{L}_q^p(0, t)}$ . Let

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

Let  $\|\cdot\|_\infty$  be the uniform norm, and for any function  $f$  on  $[0, T] \times \mathbb{R}^d$ , let

$$\|f\|_{t, \infty} := \sup_{x \in \mathbb{R}^d} |f(t, x)|, \quad \|f\|_{r \rightarrow t, \infty} := \sup_{s \in [r, t]} \|f\|_{s, \infty}, \quad 0 \leq r \leq t \leq T.$$

We make the following assumptions for the coefficients  $(a, b) \in \Xi$ , where  $\nabla$  is the gradient operator on  $\mathbb{R}^d$ .

( $A^{a,b}$ ) There exist constants  $\alpha \in (0, 1]$ ,  $K > 1$ ,  $l \in \mathbb{N}$  and  $\{(p_i, q_i)\}_{0 \leq i \leq l} \subset \mathcal{K}$  such that the following conditions hold.

(1)  $\|a\|_\infty \vee \|a^{-1}\|_\infty \leq K$ , and

$$(1.2) \quad \|a(t, x) - a(t, y)\| \leq K|x - y|^\alpha, \quad t \in [0, T], x, y \in \mathbb{R}^d.$$

Moreover, there exist  $\{1 \leq f_i\}_{1 \leq i \leq l}$  with  $\sum_{i=1}^l \|f_i\|_{\tilde{L}_{q_i}^{p_i}(T)} \leq K$ , such that

$$\|\nabla a\| \leq \sum_{i=1}^l f_i.$$

(2)  $b$  has a decomposition  $b = b^{(0)} + b^{(1)}$  such that

$$\sup_{t \in [0, T]} |b^{(1)}(t, 0)| + \|\nabla b^{(1)}\|_\infty + \|b^{(0)}\|_{\tilde{L}_{q_0}^{p_0}(T)} \leq K.$$

Let  $\sigma(t, x) := \sqrt{2a(t, x)}$ , and let  $W_t$  be a  $d$ -dimensional Brownian motion on a probability basis  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0, T]}, \mathbb{P})$ . By [11, Theorem 2.1] for  $V(x) := 1 + |x|^2$ , see also [17] or [19], under ( $A^{a,b}$ ), for any  $(s, x) \in [0, T] \times \mathbb{R}^d$ , the SDE

$$(1.3) \quad dX_{s,t}^x = b(t, X_{s,t}^x)dt + \sigma(t, X_{s,t}^x)dW_t, \quad t \in [s, T]$$

is well-posed, so that  $(L_t^{a,b})_{t \in [s, T]}$  generates a unique diffusion process. Moreover, for any  $k \geq 1$ , there exists a constant  $c(k) > 0$  such that

$$(1.4) \quad \mathbb{E} \left[ \sup_{t \in [s, T]} |X_{s,t}^x|^k \right] \leq c(k)(1 + |x|^k), \quad (s, x) \in [0, T] \times \mathbb{R}^d.$$

The associated Markov semigroup is given by

$$P_{s,t}^{a,b} f(x) := \mathbb{E}[f(X_{s,t}^x)], \quad 0 \leq s \leq t \leq T, x \in \mathbb{R}^d, f \in \mathcal{B}_b(\mathbb{R}^d).$$

Since  $(p_0, q_0) \in \mathcal{K}$ , we have

$$m_0 := \inf \left\{ m > 1 : \frac{(m-1)p_0}{m} \wedge \frac{(m-1)q_0}{m} > 1, \frac{dm}{p_0(m-1)} + \frac{2m}{q_0(m-1)} < 2 \right\} \in (1, 2).$$

For a  $\mathbb{R}^d \otimes \mathbb{R}^d$  valued differentiable function  $a = (a^{ij})_{1 \leq i, j \leq d}$ , its divergence is an  $\mathbb{R}^d$  valued function defined as

$$(\operatorname{div} a)^i := \sum_{j=1}^d \partial_j a^{ij}, \quad 1 \leq i \leq d.$$

Our first result is the following.

**Theorem 1.1.** Assume  $(A^{a,b})$  for  $(a,b) = (a_i, b_i), i = 1, 2$ . Then for any  $m \in (m_0, 2)$ , there exists a constant  $c > 0$  depending only on  $m, K, d, T$  and  $(p_i, q_i)_{0 \leq i \leq l}$ , such that for any  $\psi \in \mathcal{A}$  and  $\gamma, \tilde{\gamma} \in \mathcal{P}$ ,

$$\begin{aligned}
(1.5) \quad \mathbb{W}_\psi(P_{s,t}^{1,\gamma}, P_{s,t}^{2,\tilde{\gamma}}) &\leq \frac{c\psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}} \mathbb{W}_1(\gamma, \tilde{\gamma}) + c \int_s^t \frac{\psi((t-r)^{\frac{1}{2}}) \|a_1 - a_2\|_{r,\infty}}{\sqrt{(r-s)(t-r)}} dr \\
&+ c \left( \int_s^t \left( \frac{\psi((t-r)^{\frac{1}{2}}) \|a_1 - a_2\|_{r,\infty}}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}} \\
&+ c \int_s^t \frac{\psi((t-r)^{\frac{1}{2}})}{\sqrt{t-r}} \{ \|b_1 - b_2\|_{r,\infty} + \|\operatorname{div}(a_1 - a_2)\|_{r,\infty} \} dr, \quad 0 \leq s < t \leq T, \quad \gamma, \tilde{\gamma} \in \mathcal{P}.
\end{aligned}$$

Moreover, for any  $k \geq 1$ , there exists a constant  $C > 0$  depending only on  $k, K, d, T$  and  $(p_i, q_i)_{0 \leq i \leq l}$ , such that for any  $\gamma, \tilde{\gamma} \in \mathcal{P}$  and  $0 \leq s \leq t \leq T$ ,

$$(1.6) \quad \mathbb{W}_k(P_{s,t}^{1,\gamma}, P_{s,t}^{2,\tilde{\gamma}}) \leq C \left[ \mathbb{W}_k(\gamma, \tilde{\gamma}) + \int_s^t \|b_1 - b_2\|_{r,\infty} dr + \left( \int_s^t \|a_1 - a_2\|_{r,\infty}^2 dr \right)^{\frac{1}{2}} \right].$$

Next, we consider the following distribution dependent SDE on  $\mathbb{R}^d$ :

$$(1.7) \quad dX_t = b_t(X_t, \mathcal{L}_{X_t})dt + \sigma_t(X_t, \mathcal{L}_{X_t})dW_t, \quad t \in [0, T],$$

where  $\mathcal{L}_{X_t}$  is the distribution of  $X_t$ , and for some  $k \geq 1$ ,

$$b : [0, T] \times \mathbb{R}^d \times \mathcal{P}_k \rightarrow \mathbb{R}^d, \quad a : [0, T] \times \mathbb{R}^d \times \mathcal{P}_k \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$$

are measurable, each  $a_t(x, \mu)$  is positive definite and  $\sigma = \sqrt{2a}$ .

Let  $C_b^w([0, T]; \mathcal{P}_k)$  be the set of all weakly continuous maps  $\mu : [0, T] \rightarrow \mathcal{P}_k$  such that

$$\sup_{t \in [0, T]} \mu_t(| \cdot |^k) < \infty.$$

We call the SDE (1.7) well-posed for distributions in  $\mathcal{P}_k$ , if for any initial value  $X_0$  with  $\mathcal{L}_{X_0} \in \mathcal{P}_k$  (correspondingly, any initial distribution  $\nu \in \mathcal{P}_k$ ), the SDE has a unique solution (correspondingly, a unique weak solution) with  $(\mathcal{L}_{X_t})_{t \in [0, T]} \in C_b^w([0, T]; \mathcal{P}_k)$ . In this case, let  $P_t^* \nu := \mathcal{L}_{X_t}$  for the solution with  $\mathcal{L}_{X_0} = \nu$ , and define

$$P_t f(\nu) := \int_{\mathbb{R}^d} f d(P_t^* \nu), \quad \nu \in \mathcal{P}_k, t \in [0, T], f \in \mathcal{B}_b(\mathbb{R}^d).$$

In particular, for  $k = 2$ , the following log-Harnack inequality

$$(1.8) \quad P_t \log f(\gamma) \leq \log P_t f(\tilde{\gamma}) + \frac{c}{t} \mathbb{W}_2(\mu, \nu)^2, \quad f \in \mathcal{B}_b^+(\mathbb{R}^d), t \in (0, T], \mu, \nu \in \mathcal{P}_2$$

for some constant  $c > 0$  has been established and applied in [6, 8, 12, 14, 15] for  $\sigma_t(x, \mu) = \sigma_t(x)$  not dependent on  $\mu$ , see also [4, 5, 16] for extensions to the infinite-dimensional and

reflecting models. When the noise coefficient is also distribution dependent and is  $\mathbb{W}_2$ -Lipschitz continuous, this inequality is established in the recent work [13] by using a bi-coupling method.

In the following, we consider more singular situation where  $\sigma_t(x, \mu)$  may be not  $\mathbb{W}_2$ -Lipschitz continuous in  $\mu$ , and the drift is singular in the time-spatial variables. For any  $\mu \in C_b^w([0, T]; \mathcal{P}_k)$ , let

$$a^\mu(t, x) := a_t(x, \mu_t), \quad b^\mu(t, x) := b_t(x, \mu_t), \quad t \in [0, T], x \in \mathbb{R}^d.$$

Correspondingly to  $(A^{a,b})$ , we make the following assumption.

$(B^{a,b})$  Let  $k \in [1, \infty)$  and  $\psi \in \mathcal{A}$  with  $\lim_{t \rightarrow 0} \psi(t) = 0$ .

- (1)  $(A^{a,b})$  holds for  $(a, b) = (a^\mu, b^\mu)$  uniformly in  $\mu \in C_b^w([0, T]; \mathcal{P}_k)$ , with drift decomposition  $b^\mu = (b^\mu)^{(0)} + (b^\mu)^{(1)}$ .
- (2) There exists a constant  $K > 0$  such that

$$\|a_t(\cdot, \gamma) - a_t(\cdot, \tilde{\gamma})\|_\infty \leq K(\mathbb{W}_\psi + \mathbb{W}_k)(\gamma, \tilde{\gamma}), \quad t \in [0, T], \gamma, \tilde{\gamma} \in \mathcal{P}_k.$$

- (3) There exist  $p \geq 2$  and  $1 \leq \rho \in L^p([0, T])$ , where  $p = 2$  if  $\int_0^1 \frac{\psi(r)^2}{r} dr < \infty$  and  $p > 2$  otherwise, such that for any  $t \in [0, T]$  and  $\gamma, \tilde{\gamma} \in \mathcal{P}_k$ ,

$$\|b_t(\cdot, \gamma) - b_t(\cdot, \tilde{\gamma})\|_\infty + \|\operatorname{div}(a_t(\cdot, \gamma) - a_t(\cdot, \tilde{\gamma}))\|_\infty \leq \rho_t(\mathbb{W}_\psi + \mathbb{W}_k)(\gamma, \tilde{\gamma}).$$

**Remark 1.2.** We give a simple example satisfying  $(B^{a,b})$  for some  $\rho \in L^\infty([0, T])$ , where  $b$  contains a locally integrable term  $b^{(0)}$ , and the dependence of  $b$  and  $\sigma$  in distribution is given by singular integral kernels. Let  $\psi \in \mathcal{A}$  with  $\lim_{t \rightarrow 0} \psi(t) = 0$ , let

$$b_t(\cdot, \mu) = b_t^{(0)} + \int_{\mathbb{R}^d} \tilde{b}_t(\cdot, y) \mu(dy),$$

$$\sigma_t(\cdot, \mu) = \sqrt{\lambda I + \int_{\mathbb{R}^d} (\tilde{\sigma}_t \tilde{\sigma}_t^*)(\cdot, y) \mu(dy)}, \quad (t, \mu) \in [0, T] \times \mathcal{P}_k,$$

where  $\lambda > 0$  is a constant,  $b^{(0)} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  satisfies  $\|b^{(0)}\|_{\tilde{L}_{q_0}^{p_0}(T)} < \infty$  for some  $(p_0, q_0) \in \mathcal{X}$ ,  $\tilde{b} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is measurable such that

$$|\tilde{b}_t^{(1)}(x, y) - \tilde{b}_t^{(1)}(\tilde{x}, \tilde{y})| \leq K(|x - \tilde{x}| + \psi(|y - \tilde{y}|)), \quad x, \tilde{x}, y, \tilde{y} \in \mathbb{R}^d, t \in [0, T]$$

holds for some constant  $K > 0$ , and  $\tilde{\sigma} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{d \otimes d}$  is measurable and bounded such that for some constant  $K > 0$ ,

$$\|\tilde{\sigma}_t(x, y) - \tilde{\sigma}_t(\tilde{x}, \tilde{y})\| \leq K(|x - \tilde{x}| + \psi(|y - \tilde{y}|)),$$

$$|\nabla \tilde{\sigma}_t(\cdot, y)(x) - \nabla \tilde{\sigma}_t(\cdot, \tilde{y})(x)| \leq K\psi(|y - \tilde{y}|), \quad (t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d.$$

We have the following result on the well-posedness and estimates on  $(\mathbb{W}_\psi, \mathbb{W}_k)$  for  $P_t^*$ .

**Theorem 1.3.** *Assume  $(B^{a,b})$ . Then the following assertions hold.*

- (1) *The SDE (1.7) is well-posed for distributions in  $\mathcal{P}_k$ . Moreover, for any  $n \in \mathbb{N}$ , there exists a constant  $c > 0$  such that any solution satisfies*

$$(1.9) \quad \mathbb{E} \left[ \sup_{t \in [0, T]} |X_t|^n \middle| \mathcal{F}_0 \right] \leq c(1 + |X_0|^n).$$

- (2) *If  $\psi$  is a Dini function, i.e.*

$$(1.10) \quad \int_0^1 \frac{\psi(s)}{s} ds < \infty,$$

*then there exists a constant  $c > 0$  such that*

$$(1.11) \quad \begin{aligned} \mathbb{W}_\psi(P_t^* \gamma, P_t^* \tilde{\gamma}) &\leq \frac{c\psi(t^{\frac{1}{2}})}{\sqrt{t}} \mathbb{W}_1(\gamma, \tilde{\gamma}) + c\mathbb{W}_k(\gamma, \tilde{\gamma}), \\ \mathbb{W}_k(P_t^* \gamma, P_t^* \tilde{\gamma}) &\leq c\mathbb{W}_k(\gamma, \tilde{\gamma}), \quad t \in (0, T], \quad \gamma, \tilde{\gamma} \in \mathcal{P}_k. \end{aligned}$$

**Remark 1.4.** *Theorem 1.3(1) improves existing well-posedness results for singular McKean-Vlasov SDEs where the coefficients are either  $(\mathbb{W}_k + \mathbb{W}_\alpha)$ -Lipschitz continuous in distribution for some  $\alpha \in (0, 1]$  and  $k \geq 1$  (see [7, 3] and references therein), or satisfy some derivative conditions in distribution (see for instance [2]).*

To estimate  $\mathbb{W}_\psi(P_t^* \gamma, P_t^* \tilde{\gamma})$  for worse  $\psi$  not satisfying (1.10), and to estimate the relative entropy  $\text{Ent}(P_t^* \gamma | P_t^* \tilde{\gamma})$ , we need the drift to be Dini continuous in the spatial variable.

**Theorem 1.5.** *Assume  $(B^{a,b})$  with  $\|\rho\|_\infty < \infty$  and  $\int_0^1 \frac{\psi(r)^2}{r} dr < \infty$ , and there exists  $\phi \in \mathcal{A}$  satisfying (1.10) such that*

$$\sup_{\mu \in C_b^w([0, T]; \mathcal{P}_k)} \left\{ \|(b^\mu)^{(0)}\|_\infty + [(b^\mu)^0]_\phi + \|\nabla a^\mu\|_\infty \right\} < \infty.$$

*Then the following assertions hold.*

- (1) *If  $\psi(r)^2 \log(1 + r^{-1}) \rightarrow 0$  as  $r \rightarrow 0$ , then there exists a constant  $c > 0$  such that (1.11) holds, and for any  $t \in (0, T]$ ,  $\gamma, \tilde{\gamma} \in \mathcal{P}_k$ ,*

$$(1.12) \quad \begin{aligned} \text{Ent}(P_t^* \gamma | P_t^* \tilde{\gamma}) &\leq \frac{c\mathbb{W}_2(\gamma, \tilde{\gamma})^2}{t} \\ &+ c\mathbb{W}_k(\gamma, \tilde{\gamma})^2 \left( \frac{1}{t} \int_0^t \frac{\psi(r)^2}{r} dr + \frac{\psi(t^{\frac{1}{2}})^2}{t} \log(1 + t^{-1}) \right). \end{aligned}$$

(2) If either  $\|b\|_\infty < \infty$  or

$$(1.13) \quad \sup_{(t,\mu) \in [0,T] \times \mathcal{P}_k} (\|\nabla^i b_t(\cdot, \mu)\|_\infty + \|\nabla^i \sigma_t(\cdot, \mu)\|_\infty) < \infty, \quad i = 1, 2,$$

then there exists a constant  $c > 0$  such that (1.11) holds, and

$$(1.14) \quad \text{Ent}(P_t^* \gamma | P_t^* \tilde{\gamma}) \leq \frac{c \mathbb{W}_2(\gamma, \tilde{\gamma})^2}{t} + \frac{c \mathbb{W}_k(\gamma, \tilde{\gamma})^2}{t} \int_0^t \frac{\psi(r)^2}{r} dr, \quad t \in (0, T], \gamma, \tilde{\gamma} \in \mathcal{P}_k.$$

**Remark 1.6.** When  $k \leq 2$ , (1.8) follows from (1.14) or (1.12). This improves [13, Theorem 1.2], where the  $\mathbb{W}_2$ -Lipschitz condition on the coefficients  $(a, b)$  is relaxed as the  $(\mathbb{W}_\psi + \mathbb{W}_k)$ -Lipschitz condition.

## 2 Proof of Theorem 1.1

We first present a lemma to bound  $\mathbb{W}_\psi$  by the total variation distance and  $\mathbb{W}_1$ .

**Lemma 2.1.** For any  $\psi \in \mathcal{A}$ ,

$$\mathbb{W}_\psi(\gamma, \tilde{\gamma}) \leq \sqrt{d} \psi(\sqrt{t}) \|\gamma - \tilde{\gamma}\|_{\text{var}} + \frac{d\psi(\sqrt{t})}{\sqrt{t}} \mathbb{W}_1(\gamma, \tilde{\gamma}), \quad \gamma, \tilde{\gamma} \in \mathcal{P}_1.$$

*Proof.* Since  $\psi$  is nonnegative and concave, we have

$$(2.1) \quad \psi(Rr) \leq R\psi(r), \quad r \geq 0, R \geq 1.$$

For any function  $f$  on  $\mathbb{R}^d$  with  $[f]_\psi \leq 1$ , let

$$f_t(x) := \mathbb{E}[f(x + B_t)], \quad t \geq 0, x \in \mathbb{R}^d,$$

where  $B_t$  is the standard Brownian motion on  $\mathbb{R}^d$  with  $B_0 = 0$ . We have  $\mathbb{E}[|B_t|^2] = dt$ . By  $[f]_\psi \leq 1$ , Jensen's inequality and (2.1), we obtain

$$|f_t(x) - f(x)| \leq \mathbb{E}[\psi(|B_t|)] \leq \psi(\mathbb{E}|B_t|) \leq \psi((dt)^{\frac{1}{2}}) \leq \sqrt{d} \psi(t^{\frac{1}{2}}), \quad t \geq 0, x \in \mathbb{R}^d.$$

So,

$$(2.2) \quad \sup_{[f]_\psi \leq 1} |\gamma(f_t - f) - \tilde{\gamma}(f_t - f)| \leq \sqrt{d} \psi(t^{\frac{1}{2}}) \|\gamma - \tilde{\gamma}\|_{\text{var}}, \quad t \geq 0.$$

Next, for  $[f]_\psi \leq 1$ , by Jensen's inequality, (2.1),  $\mathbb{E}|B_t|^2 = dt$  and  $\mathbb{E}|B_t| \leq \sqrt{dt}$ , we obtain

$$|\nabla f_t(x)| = \left| \nabla_x \int_{\mathbb{R}^d} (2\pi t)^{-\frac{d}{2}} e^{-\frac{|x-y|^2}{2t}} (f(y) - f(z)) dy \right|_{z=x}$$



$$\begin{aligned}
&\leq (2\pi t)^{-\frac{d}{2}} \int_{\mathbb{R}^d} \frac{|x-y|}{t} |f(y) - f(x)| e^{-\frac{|x-y|^2}{2t}} dy \leq \frac{1}{t} \mathbb{E}[|B_t| \psi(|B_t|)] \\
&\leq \frac{\mathbb{E}|B_t|}{t} \psi\left(\frac{\mathbb{E}|B_t|^2}{\mathbb{E}|B_t|}\right) = \frac{\mathbb{E}|B_t|}{t} \psi\left(\frac{(d\mathbb{E}|B_t|^2)^{\frac{1}{2}}}{\mathbb{E}|B_t|} t^{\frac{1}{2}}\right) \leq dt^{-\frac{1}{2}} \psi(t^{\frac{1}{2}}), \quad t > 0.
\end{aligned}$$

Combining this with (2.2) and noting that

$$\mathbb{W}_1(\gamma, \tilde{\gamma}) = \sup_{\|\nabla g\| \leq 1} |\gamma(g) - \tilde{\gamma}(g)|,$$

we derive that for any  $f$  with  $[f]_\psi \leq 1$ ,

$$\begin{aligned}
|\gamma(f) - \tilde{\gamma}(f)| &\leq |\gamma(f_t - f) - \tilde{\gamma}(f_t - f)| + |\gamma(f_t) - \tilde{\gamma}(f_t)| \\
&\leq \sqrt{d} \psi(t^{\frac{1}{2}}) \|\gamma - \tilde{\gamma}\|_{var} + dt^{-\frac{1}{2}} \psi(t^{\frac{1}{2}}) \mathbb{W}_1(\gamma, \tilde{\gamma}), \quad t > 0.
\end{aligned}$$

Then the proof is finished.  $\square$

Next, we present a gradient estimate on  $P_{s,t}^{a,b}$ . All constants in the following only depend on  $T, K, d$  and  $(p_i, q_i)_{0 \leq i \leq l}$ .

**Lemma 2.2.** *Assume  $(A^{a,b})$  without (1.2). Then there exists a constant  $c > 0$  such that for any  $\psi \in \mathcal{A}$ ,*

$$\sup_{[f]_\psi \leq 1} \|\nabla P_{s,t}^{a,b} f\|_\infty \leq c(t-s)^{-\frac{1}{2}} \psi((t-s)^{\frac{1}{2}}), \quad 0 \leq s < t \leq T.$$

*Proof.* (a) By [17, Theorem 1.1] or [15, Theorem 2.1], there exists a constant  $c_1 > 0$  such that for any  $0 \leq s < t \leq T$  and  $x \in \mathbb{R}^d$ , the Bismut formula

$$(2.3) \quad \nabla P_{s,t}^{a,b} f(x) = \mathbb{E}[f(X_{s,t}^x) M_{s,t}^x]$$

holds for some random variable  $M_{s,t}^x$  on  $\mathbb{R}^d$  with

$$(2.4) \quad \mathbb{E}[M_{s,t}^x] = 0, \quad \mathbb{E}|M_{s,t}^x|^2 \leq c_1^2(t-s)^{-1}.$$

So, for any  $z \in \mathbb{R}^d$  and a function  $f$  with  $[f]_\psi \leq 1$ ,

$$|\nabla P_{s,t}^{a,b} f(x)| = \left| \mathbb{E}[\{f(X_{s,t}^x) - f(z)\} M_{s,t}^x] \right| \leq \mathbb{E}[\psi(|X_{s,t}^x - z|) |M_{s,t}^x|].$$

By Jensen's inequality for the weighted probability  $\frac{|M_{s,t}^x| \mathbb{P}}{\mathbb{E}|M_{s,t}^x|}$ , we obtain

$$\begin{aligned}
|\nabla P_{s,t}^{a,b} f(x)| &\leq \mathbb{E}[|M_{s,t}^x|] \psi\left(\frac{\mathbb{E}[|X_{s,t}^x - z| \cdot |M_{s,t}^x|]}{\mathbb{E}[|M_{s,t}^x|]}\right) \\
&\leq \mathbb{E}[|M_{s,t}^x|] \psi\left(\frac{(\mathbb{E}[|M_{s,t}^x|^2])^{\frac{1}{2}}}{\mathbb{E}[|M_{s,t}^x|]} (\mathbb{E}|X_{s,t}^x - z|^2)^{\frac{1}{2}}\right).
\end{aligned}$$

Combining this with (2.1) and (2.4), we obtain

$$(2.5) \quad \sup_{[f]_\psi \leq 1} |\nabla P_{s,t}^{a,b} f(x)| \leq c_1(t-s)^{-\frac{1}{2}} \inf_{z \in \mathbb{R}^d} \psi \left( \left\{ \mathbb{E} |X_{s,t}^x - z|^2 \right\}^{\frac{1}{2}} \right), \quad 0 \leq s < t \leq T, x \in \mathbb{R}^d.$$

(b) To estimate  $\inf_{z \in \mathbb{R}^d} \mathbb{E} |X_{s,t}^x - z|^2$ , we use Zvonkin's transform. By [19, Theorem 2.1], there exist constants  $\beta \in (0, 1)$  and  $\lambda, C > 0$  such that the PDE

$$(2.6) \quad (\partial_t + L_t^{a,b} - \lambda)u_t = -b^{(0)}(t, \cdot), \quad t \in [0, T], u_T = 0$$

for  $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  has a unique solution satisfying

$$(2.7) \quad \|u\|_\infty + \|\nabla u\|_\infty + \sup_{x \neq y} \frac{|\nabla u_t(x) - \nabla u_t(y)|}{|x - y|^\beta} \leq \frac{1}{2},$$

$$(2.8) \quad \|\nabla^2 u\|_{\tilde{L}_{q_0}^{p_0}(T)} + \|(\partial_t + b^{(1)} \cdot \nabla)u\|_{\tilde{L}_{q_0}^{p_0}(T)} \leq C.$$

By Itô's formula,  $Y_{s,t} := \Theta_t(X_{s,t}^x)$ , where  $\Theta_t(y) := y + u_t(y)$ , solves the SDE

$$dY_{s,t} = \bar{b}(t, Y_{s,t})dt + \bar{\sigma}(t, Y_{s,t})dW_t, \quad t \in [s, T], Y_{s,s} = x + u_s(x),$$

where

$$(2.9) \quad \bar{b}(t, \cdot) := (\lambda u_t + b^{(1)}) \circ \Theta_t^{-1}, \quad \bar{\sigma}(t, \cdot) := \{(\nabla \Theta_t) \sigma_t\} \circ \Theta_t^{-1}.$$

By (2.7), we find a constant  $c_1 > 0$  such that

$$(2.10) \quad |\bar{b}(t, y) - \bar{b}(t, z)| \leq c_1|y - z|, \quad \|\bar{\sigma}(t, y)\| \leq c_1, \quad t \in [s, T], y, z \in \mathbb{R}^d.$$

Let

$$\frac{d}{dt} \theta_{s,t} = \bar{b}(t, \theta_{s,t}), \quad t \in [s, T], \theta_{s,s} = Y_{s,s} = x + u_s(x).$$

By Itô's formula and (2.10), we find a constant  $c_2 > 0$  and a martingale  $M_t$  such that

$$\begin{aligned} d|Y_{s,t} - \theta_{s,t}|^2 &= \left\{ 2\langle Y_{s,t} - \theta_{s,t}, \bar{b}(t, Y_{s,t}) - \bar{b}(t, \theta_{s,t}) \rangle + \|\bar{\sigma}(t, Y_{s,t})\|_{HS}^2 \right\} dt + dM_t \\ &\leq c_2 \left\{ |Y_{s,t} - \theta_{s,t}|^2 + 1 \right\} dt + dM_t, \quad t \in [s, T], |Y_{s,s} - \theta_{s,s}| = 0. \end{aligned}$$

Thus,

$$\mathbb{E}[|Y_{s,t} - \theta_{s,t}|^2] \leq c_2 e^{c_2 T} (t - s), \quad 0 \leq s \leq t \leq T.$$

Taking  $z_{s,t} = \Theta_t^{-1}(\theta_{s,t})$  and noting that  $\|\nabla \Theta^{-1}\|_\infty < \infty$  due to  $\|\nabla u\|_\infty \leq \frac{1}{2}$  in (2.7), we find a constant  $c_3 > 0$  such that

$$\mathbb{E}[|X_{s,t}^x - z_{s,t}|^2] = \mathbb{E}[|\Theta_t^{-1}(Y_{s,t}) - \Theta_t^{-1}(\theta_{s,t})|^2] \leq c_3(t - s), \quad 0 \leq s \leq t \leq T.$$

Combining this with (2.5) and (2.1), we finish the proof.  $\square$

Moreover, we estimate  $\nabla_y p_{s,t}^{a,b}(x, y)$ , where  $\nabla_y$  is the gradient in  $y$  and  $p_{s,t}^{a,b}(x, \cdot)$  is the density function of  $\mathcal{L}_{X_{s,t}^x}$ . For any constant  $\kappa > 0$ , let

$$g_\kappa(r, z) := (\pi\kappa r)^{-\frac{d}{2}} e^{-\frac{|z|^2}{\kappa r}}, \quad r > 0, z \in \mathbb{R}^d$$

be the standard Gaussian heat kernel with parameter  $\kappa$ .

**Lemma 2.3.** *Assume  $(A^{a,b})$ . Then for any  $m \in (m_0, 2)$  there exists a constant  $c(m) > 0$  such that for any  $t \in (0, T]$  and  $0 \leq g_{\cdot, t} \in \mathcal{B}([0, t])$ ,*

$$(2.11) \quad \begin{aligned} & \int_s^t \frac{g_{r,t}}{\sqrt{t-r}} dr \int_{\mathbb{R}^d} |\nabla_y p_{s,r}^{a,b}(x, y)| dy \\ & \leq c(m) \int_s^t \frac{g_{r,t}}{\sqrt{(t-r)(r-s)}} dr + c(m) \left( \int_s^t \left( \frac{g_{r,t}}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}}, \quad s \in [0, t]. \end{aligned}$$

Consequently, there exists a constant  $c > 0$  such that

$$(2.12) \quad \int_s^t (t-r)^{-\frac{1}{2}} dr \int_{\mathbb{R}^d} |\nabla_y p_{s,r}^{a,b}(x, y)| dy \leq c, \quad 0 \leq s < t \leq T.$$

*Proof.* Let  $u_t$  be in (2.6). By  $(A^{a,b})$ ,  $\sigma = \sqrt{2a}$ , (2.7) and (2.9), we find a constant  $c_1 > 0$  such that

$$|\bar{b}(t, x) - \bar{b}(t, y)| \leq c_1 |x - y|, \quad \|\bar{\sigma}(t, x) - \bar{\sigma}(t, y)\| \leq c_1 |x - y|^{\alpha \wedge \beta}, \quad t \in [0, T], x, y \in \mathbb{R}^d.$$

Let  $\bar{p}_{s,t}(x, y)$  be the density function of  $\mathcal{L}_{Y_{s,t}}$ . According to [10, Theorem 1.2], there exists a constant  $\kappa \geq 1$  and some  $\theta_{s,t} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that

$$(2.13) \quad |\nabla_y^i \bar{p}_{s,t}(x, y)| \leq \kappa (t-s)^{-\frac{i}{2}} g_\kappa(t-s, \theta_{s,t}(x) - y), \quad 0 \leq s < t \leq T, x, y \in \mathbb{R}^d, i = 0, 1,$$

where  $\nabla^0 f := f$ . Noting that  $X_{s,t}^x = \Theta_t^{-1}(Y_{s,t})$ , we have

$$(2.14) \quad p_{s,t}^{a,b}(x, y) = \bar{p}_{s,t}(\Theta_s(x), \Theta_t(y)) |\det(\nabla \Theta_t(y))|.$$

Combining this with (2.7), (2.10) and (2.13), we find a constant  $c_2 > 0$  such that

$$(2.15) \quad \begin{aligned} |\nabla_y p_{s,t}^{a,b}(x, y)| & \leq c_2 \kappa (t-s)^{-\frac{1}{2}} g_\kappa(t-s, \theta_{s,t}(\Theta_s(x)) - \Theta_t(y)) |\det(\nabla \Theta_t(y))| \\ & \quad + c_2 \|\nabla^2 u_t(y)\| p_{s,t}^{a,b}(x, y), \quad 0 \leq s < t, x, y \in \mathbb{R}^d. \end{aligned}$$

Since  $(p_0, q_0) \in \mathcal{K}$ , for any  $m > m_0$ , we have

$$(2.16) \quad \tilde{p} := \frac{(m-1)p_0}{m} > 1, \quad \tilde{q} := \frac{(m-1)q_0}{m} > 1, \quad \frac{d}{\tilde{p}} + \frac{2}{\tilde{q}} < 2.$$

By Krylov's estimate, see [19, Theorem 3.1], we find a constant  $c > 0$  such that

$$\begin{aligned}
(2.17) \quad & \int_s^t dr \int_{\mathbb{R}^d} \|\nabla^2 u_r(y)\|^{\frac{m}{m-1}} p_{s,r}^{a,b}(x,y) dy \\
&= \mathbb{E} \int_s^t \|\nabla^2 u_r\|^{\frac{m}{m-1}} (X_{s,r}^x) dr \leq c \|\nabla^2 u\|^{\frac{m}{m-1}}_{\tilde{L}_q^{\tilde{p}}(s,t)} = c (\|\nabla^2 u\|_{\tilde{L}_{q_0}^{p_0}(s,t)})^{\frac{m}{m-1}}.
\end{aligned}$$

This together with (2.8), (2.14) and (2.15) implies that for any  $m \in (m_0, 2)$ , there exists a constant  $c(m) > 0$  such that

$$\begin{aligned}
& \int_s^t \frac{g_{r,t}}{\sqrt{t-r}} dr \int_{\mathbb{R}^d} |\nabla_y p_{s,r}^{a,b}(x,y)| dy \leq c_2 \kappa \int_s^t g_{r,t} (t-r)^{-\frac{1}{2}} (r-s)^{-\frac{1}{2}} dr \\
& + c_2 \left( \int_s^t \left( \frac{g_{r,t}}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}} \left( \int_s^t dr \int_{\mathbb{R}^d} \|\nabla^2 u_r(y)\|^{\frac{m}{m-1}} p_{s,r}^{a,b}(x,y) dy \right)^{\frac{m-1}{m}} \\
& \leq c(m) \int_s^t \frac{g_{r,t}}{\sqrt{(t-r)(r-s)}} dr + c(m) \left( \int_s^t \left( \frac{g_{r,t}}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}}.
\end{aligned}$$

So, (2.11) holds. Letting  $g_{r,t} \equiv 1$  and  $m = \frac{m_0+2}{2}$ , we find a constant  $c > 0$  such that (2.11) implies (2.12). □

*Proof of Theorem 1.1.* By (1.1), it suffices to prove for  $\gamma = \delta_x, \tilde{\gamma} = \delta_y, x, y \in \mathbb{R}^d$ .

(a) We first consider  $x = y$ . Let  $f \in C_b^2(\mathbb{R}^d)$  with  $[f]_\psi \leq 1$ . By Itô's formula we have

$$P_{s,t}^{a_2,b_2} f(x) = f(x) + \int_s^t P_{s,r}^{a_2,b_2} (L_r^{a_2,b_2} f)(x) dr, \quad 0 \leq s \leq t \leq T.$$

This implies the Kolmogorov forward equation

$$(2.18) \quad \partial_t P_{s,t}^{a_2,b_2} f = P_{s,t}^{a_2,b_2} (L_t f), \quad \text{a.e. } t \in [s, T].$$

On the other hand, for  $(p, q) \in \mathcal{K}$  and  $t \in (0, T]$ , let  $\tilde{W}_{1,q,b_2^{(1)}}^{2,p}(0, t)$  be the set of all maps  $u : [0, t] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  satisfying

$$\|u\|_{0 \rightarrow t, \infty} + \|\nabla u\|_{0 \rightarrow t, \infty} + \|\nabla^2 u\|_{\tilde{L}_q^p(t)} + \|(\partial_s + b_2^{(1)} \cdot \nabla) u\|_{\tilde{L}_q^p(t)} < \infty.$$

By [19, Theorem 2.1], the PDE

$$(2.19) \quad (\partial_s + L_s^{a_2,b_2}) u_s = -L_s^{a_2,b_2} f, \quad s \in [0, t], u_t = 0$$

has a unique solution in the class  $\tilde{W}_{1,q,b_2^{(1)}}^{2,p}(0, t)$ . So, by Itô's formula [19, Lemma 3.3],

$$du_r(X_{s,r}^{2,x}) = -L_r^{a_2,b_2} f(X_{s,r}^{2,x}) + dM_r, \quad r \in [s, t]$$

holds for some martingale  $M_r$ . This and (2.18) yield

$$\begin{aligned} 0 &= \mathbb{E}u_t(X_{s,t}^{2,x}) = u_s(x) - \int_s^t (P_{s,r}^{a_2,b_2} L_r^{a_2,b_2} f) dr \\ &= u_s(x) - \int_s^t \frac{d}{dr} (P_{s,r}^{a_2,b_2} f) dr = u_s(x) - P_{s,t}^{a_2,b_2} f + f, \quad 0 \leq s \leq t \leq T. \end{aligned}$$

Combining this with (2.19), we derive  $P_{\cdot,t}^{a_2,b_2} f \in \tilde{W}_{1,q,b_2^{(1)}}^{2,p}(0,t)$  for  $t \in (0,T]$  and the Kolmogorov backward equation

$$(2.20) \quad \partial_s P_{s,t}^{a_2,b_2} f = \partial_s u_s = -L_s^{a_2,b_2}(u_s + f) = -L_s^{a_2,b_2} P_{s,t}^{a_2,b_2} f, \quad 0 \leq s \leq t \leq T.$$

By Itô's formula to  $P_{r,t}^{a_2,b_2} f(X_{s,r}^{1,x})$  for  $r \in [s,t]$ , see [19, Lemma 3.3], we derive

$$\begin{aligned} P_{s,t}^{a_1,b_1} f(x) - P_{s,t}^{a_2,b_2} f(x) &= \mathbb{E} \int_s^t (\partial_r + L_r^{a_1,b_1}) P_{r,t}^{a_2,b_2} f(X_{s,r}^{1,x}) dr \\ &= \int_s^t dr \int_{\mathbb{R}^d} p_{s,r}^{a_1,b_1}(x,y) (L_r^{a_1,b_1} - L_r^{a_2,b_2}) P_{r,t}^{a_2,b_2} f(y) dy. \end{aligned}$$

By the integration by parts formula, we obtain

$$\begin{aligned} &\left| \int_{\mathbb{R}^d} p_{s,r}^{a_1,b_1}(x,y) [\text{tr}\{(a_1 - a_2)(r,y) \nabla^2 P_{r,t}^{a_2,b_2} f(y)\}] dy \right| \\ &= \left| \int_{\mathbb{R}^d} \left\langle (a_1 - a_2)(r,y) \nabla_y p_{s,r}^{a_1,b_1}(x,y) + p_{s,r}^{a_1,b_1}(x,y) \text{div}(a_1 - a_2)(r,y), \nabla P_{r,t}^{a_2,b_2} f(y) \right\rangle dy \right|. \end{aligned}$$

Combining these with Lemma 2.2 and Lemma 2.3, for any  $m \in (m_0, 2)$ , we find constants  $c_1, c_2 > 0$  such that

$$\begin{aligned} |P_{s,t}^{a_1,b_1} f(x) - P_{s,t}^{a_2,b_2} f(x)| &\leq c_1 \int_s^t \frac{\psi((t-r)^{\frac{1}{2}}) \|a_1 - a_2\|_{r,\infty}}{\sqrt{t-r}} dr \int_{\mathbb{R}^d} |\nabla_y p_{s,r}^{a_1,b_1}(x,y)| dy \\ &\quad + c_1 \int_s^t \frac{\psi((t-r)^{\frac{1}{2}})}{(t-r)^{\frac{1}{2}}} (\|b_1 - b_2\|_{r,\infty} + \|\text{div}(a_1 - a_2)\|_{r,\infty}) dr \\ &\leq c_2 \int_s^t \frac{\psi((t-r)^{\frac{1}{2}})}{\sqrt{t-r}} \left( \frac{\|a_1 - a_2\|_{r,\infty}}{\sqrt{r-s}} + \|b_1 - b_2\|_{r,\infty} + \|\text{div}(a_1 - a_2)\|_{r,\infty} \right) dr \\ &\quad + c_2 \left( \int_s^t \left( \frac{\psi((t-r)^{\frac{1}{2}}) \|a_1 - a_2\|_{r,\infty}}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}} =: I_{s,t}. \end{aligned}$$

Therefore,

$$(2.21) \quad \mathbb{W}_\psi(P_{s,t}^{1,x}, P_{s,t}^{2,x}) \leq I_{s,t}, \quad 0 \leq s < t \leq T, \quad x \in \mathbb{R}^d.$$

(b) Let  $x, y \in \mathbb{R}^d$  and  $0 \leq s < t \leq T$ . By the triangle inequality for  $\mathbb{W}_\psi$ , (2.21) and Lemma 2.1, we obtain

$$(2.22) \quad \begin{aligned} \mathbb{W}_\psi(P_{s,t}^{1,x}, P_{s,t}^{2,y}) &\leq \mathbb{W}_\psi(P_{s,t}^{1,x}, P_{s,t}^{2,x}) + \mathbb{W}_\psi(P_{s,t}^{2,x}, P_{s,t}^{2,y}) \\ &\leq I_{s,t} + \psi((t-s)^{\frac{1}{2}}) \|P_{s,t}^{2,x} - P_{s,t}^{2,y}\|_{var} + \frac{\psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}} \mathbb{W}_1(P_{s,t}^{2,x}, P_{s,t}^{2,y}). \end{aligned}$$

By [15, Theorem 2.1] or [17, Theorem 1.1],  $(A^{a,b})$  for  $(a, b) = (a_2, b_2)$  implies that for some constant  $c_3 > 0$ ,

$$\mathbb{W}_1(P_{s,t}^{2,x}, P_{s,t}^{2,y}) \leq c_3 |x - y|, \quad \|P_{s,t}^{2,x} - P_{s,t}^{2,y}\|_{var} \leq \frac{c_3}{\sqrt{t-s}} |x - y|$$

holds for any  $0 \leq s < t \leq T$  and  $x, y \in \mathbb{R}^d$ . Combining this with (2.22), we derive (1.5) for  $\gamma = \delta_x$  and  $\tilde{\gamma} = \delta_y$ .

(c) It remains to prove (1.6). Let  $u$  be in (2.6) for  $(a, b) = (a_1, b_1)$ . Let  $\Theta_t(y) := y + u_t(y)$ , and

$$Y_{s,t}^{1,x} = \Theta_t(X_{s,t}^{1,x}), \quad Y_{s,t}^{2,y} = \Theta_t(X_{s,t}^{2,y}), \quad t \in [s, T].$$

By Itô's formula [19, Lemma 3.3], we obtain

$$\begin{aligned} dY_{s,t}^{1,x} &= \{b_1^{(1)}(t, \cdot) + \lambda u_t\}(X_{s,t}^{1,x})dt + \{(\nabla \Theta_t)\sigma_1(t, \cdot)\}(X_{s,t}^{1,x})dW_t, \\ dY_{s,t}^{2,y} &= \{b_1^{(1)}(t, \cdot) + \lambda u_t\}(X_{s,t}^{2,y})dt + \{(\nabla \Theta_t)(b_2 - b_1) + \text{tr}[(a_2 - a_1)(t, \cdot)\nabla^2 u_t]\}(X_{s,t}^{2,y})dt \\ &\quad + \{(\nabla \Theta_t)\sigma_2(t, \cdot)\}(X_{s,t}^{2,y})dW_t, \quad t \in [s, T], \quad Y_{s,s}^{1,x} = \Theta_s(x), \quad Y_{s,s}^{2,y} = \Theta_s(y). \end{aligned}$$

For any non-negative function  $f$  on  $\mathbb{R}^d$ , let

$$\mathcal{M}f(x) := \sup_{r \in (0,1]} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)dy, \quad x \in \mathbb{R}^d, B(x,r) := \{y \in \mathbb{R}^d : |y - x| < r\}.$$

By  $(A^{a,b})$  for  $a = a_i$ ,  $\sigma_i = \sqrt{2a_i}$ , (2.7), the maximal inequality in [17, Lemma 2.1], and Itô's formula, for any  $k \geq 1$  we find a constant  $c_1 > 1$  such that

$$(2.23) \quad c_1^{-1} |X_{s,t}^{1,x} - X_{s,t}^{2,y}|^{2k} \leq \xi_t := |Y_{s,t}^{1,x} - Y_{s,t}^{2,y}|^{2k} \leq c_1 |X_{s,t}^{1,x} - X_{s,t}^{2,y}|^{2k},$$

$$(2.24) \quad d\xi_t \leq c_1 \xi_t (1 + \eta_t) dt + c_1 \xi_t^{\frac{2k-1}{2k}} \gamma_t dt + c_1 \xi_t^{\frac{k-1}{k}} \|a_1 - a_2\|_{t,\infty}^2 dt + dM_t,$$

where  $M_t$  is a martingale and

$$\begin{aligned} \gamma_t &:= \|b_1 - b_2\|_{t,\infty} + \|a_1 - a_2\|_{t,\infty} \|\nabla^2 u_t\|(X_{s,t}^{2,y}), \\ \eta_t &:= \mathcal{M}(\|\nabla \sigma_1\|_{t,\infty}^2 + \|\nabla^2 u\|^2)(X_{s,t}^{1,x}) + \mathcal{M}(\|\nabla \sigma_1\|_{t,\infty}^2 + \|\nabla^2 u\|^2)(X_{s,t}^{2,y}). \end{aligned}$$

Note that for  $q \in (\frac{2k-1}{2k}, 1)$ ,

$$\mathbb{E} \left\{ \left( \sup_{r \in [s,t]} \xi_r^q \right)^{\frac{2k-1}{2kq}} \int_s^t \|a_1 - a_2\|_{r,\infty} \|\nabla^2 u_r\|(X_{s,r}^{2,y}) dr \right\}$$

$$\begin{aligned}
&\leq \left( \mathbb{E} \sup_{r \in [s, t]} \xi_r^q \right)^{\frac{2k-1}{2kq}} \left( \mathbb{E} \left( \int_s^t \|a_1 - a_2\|_{r, \infty} \|\nabla^2 u_r\| (X_{s,r}^{2,y}) dr \right)^{\frac{2kq}{2kq-2k+1}} \right)^{\frac{2kq-2k+1}{2kq}} \\
&\leq \left( \mathbb{E} \sup_{r \in [s, t]} \xi_r^q \right)^{\frac{2k-1}{2kq}} \left( \int_s^t \|a_1 - a_2\|_{r, \infty}^m dr \right)^{\frac{1}{m}} \\
&\quad \times \left( \mathbb{E} \left( \int_s^t \|\nabla^2 u_r\|^{\frac{m}{m-1}} (X_{s,r}^{2,y}) dr \right)^{\frac{2(m-1)kq}{m(2kq-2k+1)}} \right)^{\frac{2kq-2k+1}{2kq}}, \quad m > 1.
\end{aligned}$$

So, by the stochastic Grownwall inequality [18, Lemma 2.8] for  $q \in (\frac{2k-1}{2k}, 1)$ , [17, Lemma 2.1], and the Krylov estimate in [19, Theorem 3.1] which implies the Khasminskii inequality in [18, Lemma 3.5], we find constants  $c_2, c_3 > 0$  such that

$$\begin{aligned}
&\left[ \mathbb{E} \sup_{r \in [s, t]} \xi_r^q \right]^{\frac{1}{q}} \leq c_2 |x - y|^{2k} + c_2 \mathbb{E} \int_s^t \left\{ \xi_r^{\frac{2k-1}{2k}} \gamma_r dr + \xi_r^{\frac{k-1}{k}} \|a_1 - a_2\|_{r, \infty}^2 \right\} dr \\
&\leq c_2 |x - y|^{2k} + c_2 \mathbb{E} \left[ \left( \sup_{r \in [s, t]} \xi_r^q \right)^{\frac{2k-1}{2kq}} \int_s^t \gamma_r dr + \left( \sup_{r \in [s, t]} \xi_r^q \right)^{\frac{k-1}{kq}} \int_s^t \|a_1 - a_2\|_{r, \infty}^2 dr \right] \\
&\leq c_2 |x - y|^{2k} + \frac{1}{2} \left[ \mathbb{E} \sup_{r \in [s, t]} \xi_r^q \right]^{\frac{1}{q}} + c_3 \left( \int_s^t \|a_1 - a_2\|_{r, \infty}^2 dr \right)^k + c_3 \left( \int_s^t \|b_1 - b_2\|_{r, \infty} dr \right)^{2k} \\
&+ c_3 \left( \int_s^t \|a_1 - a_2\|_{r, \infty}^m dr \right)^{\frac{2k}{m}} \left( \mathbb{E} \left( \int_s^t \|\nabla^2 u_r\|^{\frac{m}{m-1}} (X_{s,r}^{2,y}) dr \right)^{\frac{2(m-1)kq}{m(2kq-2k+1)}} \right)^{\frac{2kq-2k+1}{q}}, \quad m > 1.
\end{aligned}$$

Noting that [11, Theorem 2.1(3)] implies

$$\left[ \mathbb{E} \sup_{r \in [s, t]} \xi_r^q \right] < \infty,$$

we obtain

$$\begin{aligned}
(2.25) \quad &\left[ \mathbb{E} \sup_{r \in [s, t]} \xi_r^q \right]^{\frac{1}{q}} \leq 2c_2 |x - y|^{2k} + 2c_3 \left( \int_s^t \|a_1 - a_2\|_{r, \infty}^2 dr \right)^k \\
&+ 2c_3 \left( \int_s^t \|b_1 - b_2\|_{r, \infty} dr \right)^{2k} \\
&+ 2c_3 \left( \int_s^t \|a_1 - a_2\|_{r, \infty}^m dr \right)^{\frac{2k}{m}} \left( \mathbb{E} \left( \int_s^t \|\nabla^2 u_r\|^{\frac{m}{m-1}} (X_{s,r}^{2,y}) dr \right)^{\frac{2(m-1)kq}{m(2kq-2k+1)}} \right)^{\frac{2kq-2k+1}{q}}.
\end{aligned}$$

Recall that  $(\tilde{p}, \tilde{q})$  is defined in (2.16). By (2.8), [19, Theorem 3.1] and [18, Lemma 3.5], we find a constant  $c_4 > 0$  such that

$$\mathbb{E} \left( \int_s^t \|\nabla^2 u_r\|^{\frac{m}{m-1}} (X_{s,r}^{2,y}) dr \right)^{\frac{2(m-1)kq}{m(2kq-2k+1)}}$$

$$\leq c_4(\|\nabla^2 u\|^{\frac{m}{m-1}}_{\tilde{L}^{\tilde{p}}_q(s,t)})^{\frac{2(m-1)kq}{m(2kq-2k+1)}} = c_4(\|\nabla^2 u\|_{\tilde{L}^{p_0}_{q_0}(0,T)})^{\frac{2kq}{2kq-2k+1}} < \infty.$$

Combining this with (2.25), we find a constant  $c_5 > 0$  such that

$$\begin{aligned} (\mathbb{E}|Y_{s,t}^{1,x} - Y_{s,t}^{2,y}|^k)^2 &\leq \left[ \mathbb{E} \sup_{r \in [s,t]} \xi_r^q \right]^{\frac{1}{q}} \leq c_5 |x - y|^{2k} + c_5 \left( \int_s^t \|b_1 - b_2\|_{r,\infty} dr \right)^{2k} \\ &\quad + c_5 \left( \int_s^t \|a_1 - a_2\|_{r,\infty}^m dr \right)^{\frac{2k}{m}} + c_5 \left( \int_s^t \|a_1 - a_2\|_{r,\infty}^2 dr \right)^k. \end{aligned}$$

Noting that (2.23) implies

$$\mathbb{W}_k(P_{s,t}^{1,x}, P_{s,t}^{2,y})^k \leq \sqrt{c_1} \mathbb{E}|Y_{s,t}^{1,x} - Y_{s,t}^{2,y}|^k,$$

by Jensen's inequality we derive (1.6) for some constant  $C > 0$  and  $\gamma = \delta_x, \tilde{\gamma} = \delta_y$ . □

### 3 Proof of Theorem 1.3

Once the well-posedness of (1.7) is proved, the proof of [7, (1.5)] implies (1.9) under  $(B^{a,b})$ . We skip the details to save space. So, in the following we only prove the well-posedness and estimate (1.11).

(a) Let  $X_0$  be  $\mathcal{F}_0$ -measurable with  $\gamma := \mathcal{L}_{X_0} \in \mathcal{P}_k$ . Let

$$\mathcal{C}_T^\gamma := \{\mu \in C([0, T]; \mathcal{P}_k) : \mu_0 = \gamma\}.$$

For any  $\lambda \geq 0$ ,  $C_T^\gamma$  is a complete space under the metric

$$\rho_\lambda(\mu, \tilde{\mu}) := \sup_{t \in [0, T]} e^{-\lambda t} \{\mathbb{W}_\psi(\mu_t, \tilde{\mu}_t) + \mathbb{W}_k(\mu_t, \tilde{\mu}_t)\}.$$

For any  $\mu \in C([0, T]; \mathcal{P}_k)$ , let

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

According to [11, Theorem 2.1],  $(B^{a,b})$  implies that the SDE

$$dX_t^\mu = b_t^\mu(X_t^\mu)dt + \sigma_t^\mu(X_t^\mu)dW_t, \quad t \in [0, T], X_0^\mu = X_0$$

is well-posed, and

$$\mathbb{E} \left[ \sup_{s \in [0, T]} |X_s^\mu|^k \right] < \infty.$$

So, we define a map

$$\Phi^\gamma : \mathcal{C}_T^\gamma \rightarrow \mathcal{C}_T^\gamma; \quad \mu \mapsto \{(\Phi^\gamma \mu)_t := \mathcal{L}_{X_t^\mu}\}_{t \in [0, T]}.$$



According to [9, Theorem 3.1], if  $\Phi^\gamma$  has a unique fixed point in  $\mathcal{C}_T^\gamma$ , then (1.7) is well-posed for distributions in  $\mathcal{P}_k$ .

(b) Let  $\tilde{\gamma} \in \mathcal{P}_k$  which may be different from  $\gamma$ , and let  $\tilde{\mu} \in \mathcal{C}_T^{\tilde{\gamma}}$ . We estimate the  $\rho_\lambda$ -distance between  $\Phi^\gamma \mu$  and  $\Phi^{\tilde{\gamma}} \tilde{\mu}$ . By Theorem 1.1 and  $(B^{a,b})$ , for any  $m \in (m_0, 2)$ , there exist constants  $c_1, c_2 > 0$  such that

$$\begin{aligned}
& \mathbb{W}_\psi((\Phi^\gamma \mu)_t, (\Phi^{\tilde{\gamma}} \tilde{\mu})_t) + \mathbb{W}_k((\Phi^\gamma \mu)_t, (\Phi^{\tilde{\gamma}} \tilde{\mu})_t) \\
& \leq \frac{c_1 \psi(t^{\frac{1}{2}})}{\sqrt{t}} \mathbb{W}_k(\gamma, \tilde{\gamma}) + c_1 \left( \int_0^t \|a^\mu - a^{\tilde{\mu}}\|_{r,\infty}^2 dr \right)^{\frac{1}{2}} \\
& \quad + c_1 \left( \int_0^t \left( \frac{\psi((t-r)^{\frac{1}{2}}) \|a^\mu - a^{\tilde{\mu}}\|_{r,\infty}}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}} \\
& \quad + c_1 \int_0^t \frac{c_1 \psi((t-r)^{\frac{1}{2}})}{\sqrt{t-r}} \left( \frac{\|a^\mu - a^{\tilde{\mu}}\|_{r,\infty}}{\sqrt{r}} + \|b^\mu - b^{\tilde{\mu}}\|_{r,\infty} + \|\operatorname{div}(a^\mu - a^{\tilde{\mu}})\|_{r,\infty} \right) dr \\
& \leq \frac{c_1 \psi(t^{\frac{1}{2}})}{\sqrt{t}} \mathbb{W}_k(\gamma, \tilde{\gamma}) + c_2 \left( \int_0^t \left( \mathbb{W}_\psi(\mu_r, \tilde{\mu}_r) + \mathbb{W}_k(\mu_r, \tilde{\mu}_r) \right)^2 dr \right)^{\frac{1}{2}} \\
& \quad + c_2 \left( \int_0^t \left( \frac{\psi((t-r)^{\frac{1}{2}}) (\mathbb{W}_\psi(\mu_r, \tilde{\mu}_r) + \mathbb{W}_k(\mu_r, \tilde{\mu}_r))}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}} \\
& \quad + c_2 \int_0^t \frac{\psi((t-r)^{\frac{1}{2}})}{\sqrt{r(t-r)}} (1 + \sqrt{r} \rho_r) \left( \mathbb{W}_\psi(\mu_r, \tilde{\mu}_r) + \mathbb{W}_k(\mu_r, \tilde{\mu}_r) \right) dr
\end{aligned}$$

Let  $\gamma = \tilde{\gamma}$ . We obtain

$$\rho_\lambda(\Phi^\gamma \mu, \Phi^\gamma \tilde{\mu}) \leq \delta(\lambda) \rho_\lambda(\mu, \tilde{\mu}),$$

where by  $(B^{a,b})$  and  $m \in (m_0, 2)$ , as  $\lambda \rightarrow \infty$  we have

$$\begin{aligned}
\delta(\lambda) &:= c_2 \sup_{t \in [0, T]} \left[ \int_0^t \frac{\psi((t-r)^{\frac{1}{2}}) e^{-\lambda(t-r)}}{\sqrt{t-r}} \left( \frac{1}{\sqrt{r}} + \rho_r \right) dr + \left( \int_0^t e^{-2\lambda(t-r)} dr \right)^{\frac{1}{2}} \right] \\
& \quad + c_2 \left( \int_0^t \left( \frac{\psi((t-r)^{\frac{1}{2}}) e^{-\lambda(t-r)}}{\sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}} \rightarrow 0.
\end{aligned}$$

So,  $\Phi^\gamma$  is  $\rho_\lambda$ -contractive on  $\mathcal{C}_T^\gamma$  for large  $\lambda > 0$ , and hence has a unique fixed point. This implies the well-posedness of (1.7) for distributions in  $\mathcal{P}_k$ .

(c) For  $s \in [0, T]$ , let  $P_{s,t}^* \gamma = \mathcal{L}_{X_{s,t}^\gamma}$ , where  $X_{s,t}^\gamma$  solves (1.7) for  $t \in [s, T]$  and  $\mathcal{L}_{X_{s,s}^\gamma} = \gamma$ . By (1.9) for  $s$  replacing 0, we have

$$\sup_{t \in [s, T]} (P_{s,t}^* \gamma)(|\cdot|^k) < \infty, \quad \gamma \in \mathcal{P}_k.$$

Since  $\psi$  has growth slower than linear, and (2.1) implies the boundedness of  $\frac{r}{\psi(r)}$  for  $r \in [0, T]$ , this implies that for any  $\gamma, \tilde{\gamma} \in \mathcal{P}_k$  and  $s \in [0, T]$ ,

$$(3.1) \quad \sup_{r \in [s, t]} (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) < \infty, \quad t \in [s, T],$$

$$(3.2) \quad \Gamma_{s,t} := \sup_{r \in [s, t]} \frac{\sqrt{r-s}}{\psi((r-s)^{\frac{1}{2}})} (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) < \infty, \quad t \in [s, T].$$

Let

$$\begin{aligned} a_1(t, x) &:= a_t(x, P_{s,t}^* \gamma), \quad b_1(t, x) := b_t(x, P_{s,t}^* \gamma), \\ a_2(t, x) &:= a_t(x, P_{s,t}^* \tilde{\gamma}), \quad b_1(t, x) := b_t(x, P_{s,t}^* \tilde{\gamma}), \quad (t, x) \in [s, T] \times \mathbb{R}^d. \end{aligned}$$

Then  $P_{s,t}^* \gamma = P_{s,t}^{1,\gamma}$ ,  $P_{s,t}^* \tilde{\gamma} = P_{s,t}^{2,\tilde{\gamma}}$ , and (1.1) implies

$$(3.3) \quad P_{s,t}^* \gamma = \int_{\mathbb{R}^d} P_{s,t}^{1,x} \gamma(\mathrm{d}x), \quad P_{s,t}^* \tilde{\gamma} = \int_{\mathbb{R}^d} P_{s,t}^{2,x} \tilde{\gamma}(\mathrm{d}x).$$

Thus, by Theorem 1.1 and  $(B^{a,b})$ , for any  $m \in (m_0, 2)$ , we find a constant  $k_0 > 0$  such that

$$\begin{aligned} (3.4) \quad \mathbb{W}_\psi(P_{s,t}^* \gamma, P_{s,t}^* \tilde{\gamma}) &= \mathbb{W}_\psi(P_{s,t}^{1,\gamma}, P_{s,t}^{2,\tilde{\gamma}}) \leq \frac{k_0 \psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}} \mathbb{W}_1(\gamma, \tilde{\gamma}) \\ &+ k_0 \int_s^t \frac{\psi((t-r)^{\frac{1}{2}})}{\sqrt{(r-s)(t-r)}} \left(1 + \rho_r \sqrt{r-s}\right) (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) \mathrm{d}r \\ &+ k_0 \left( \int_s^t \left( \frac{\psi((t-r)^{\frac{1}{2}}) (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma})}{\sqrt{t-r}} \right)^m \mathrm{d}r \right)^{\frac{1}{m}}, \end{aligned}$$

$$\begin{aligned} (3.5) \quad \mathbb{W}_k(P_{s,t}^* \gamma, P_{s,t}^* \tilde{\gamma}) &= \mathbb{W}_k(P_{s,t}^{1,\gamma}, P_{s,t}^{2,\tilde{\gamma}}) \leq k_0 \mathbb{W}_k(\gamma, \tilde{\gamma}) \\ &+ k_0 \int_s^t \rho_r (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) \mathrm{d}r + k_0 \left( \int_s^t (\mathbb{W}_\psi + \mathbb{W}_k)^2(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) \mathrm{d}r \right)^{\frac{1}{2}}. \end{aligned}$$

By combining these with the definition of  $\Gamma_{s,t}$  in (3.2), we find a constant  $k_1 > 0$  such that

$$\begin{aligned} (3.6) \quad \Gamma_{s,t} &\leq k_1 \mathbb{W}_k(\gamma, \tilde{\gamma}) + k_1 \Gamma_{s,t} h(t-s), \quad 0 \leq s < t \leq T, \\ h(t) &:= \sup_{(s,\theta) \in (0,t] \times [0,T-t]} \frac{\sqrt{s}}{\psi(s^{\frac{1}{2}})} \int_0^s \frac{\psi(r^{\frac{1}{2}}) \psi((s-r)^{\frac{1}{2}})}{\sqrt{r(s-r)}} \left( \frac{1}{\sqrt{r}} + \rho_{\theta+r} \right) \mathrm{d}r \\ &+ \sup_{s \in (0,t]} \frac{\sqrt{s}}{\psi(s^{\frac{1}{2}})} \left( \int_0^s \left( \frac{\psi((s-r)^{\frac{1}{2}}) \psi(r^{\frac{1}{2}})}{\sqrt{r} \sqrt{s-r}} \right)^m \mathrm{d}r \right)^{\frac{1}{m}} \\ &+ \left( \int_0^t \left( \frac{\psi(r^{\frac{1}{2}})}{\sqrt{r}} \right)^2 \mathrm{d}r \right)^{\frac{1}{2}}, \quad t \in (0, T]. \end{aligned}$$

Note that

$$\begin{aligned}
(3.7) \quad & \frac{\sqrt{s}}{\psi(s^{\frac{1}{2}})} \int_0^s \frac{\psi(r^{\frac{1}{2}})\psi((s-r)^{\frac{1}{2}})}{r\sqrt{s-r}} dr \\
& \leq \frac{\sqrt{s}}{\psi(s^{\frac{1}{2}})} \left( \int_0^{\frac{s}{2}} \frac{\psi(s^{\frac{1}{2}})}{\sqrt{s/2}} \cdot \frac{\psi(r^{\frac{1}{2}})}{r} dr + \int_{\frac{s}{2}}^s \frac{\psi((s-r)^{\frac{1}{2}})}{s-r} \cdot \frac{\sqrt{s}\psi(s^{\frac{1}{2}})}{s/2} dr \right) \\
& \leq (2 + \sqrt{2}) \int_0^{\frac{s}{2}} \frac{\psi(r^{\frac{1}{2}})}{r} dr = 2(2 + \sqrt{2}) \int_0^{\sqrt{s/2}} \frac{\psi(r)}{r} dr.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
(3.8) \quad & \frac{\sqrt{s}}{\psi(s^{\frac{1}{2}})} \left( \int_0^s \left( \frac{\psi((s-r)^{\frac{1}{2}})\psi(r^{\frac{1}{2}})}{\sqrt{r}\sqrt{s-r}} \right)^m dr \right)^{\frac{1}{m}} \\
& \leq \sqrt{2} \left( \left( \int_0^{\frac{s}{2}} \left( \frac{\psi(r^{\frac{1}{2}})}{\sqrt{r}} \right)^m dr \right)^{\frac{1}{m}} + \left( \int_{\frac{s}{2}}^s \left( \frac{\psi((s-r)^{\frac{1}{2}})}{s-r} \right)^m dr \right)^{\frac{1}{m}} \right) \\
& \leq 2\sqrt{2} \left( \int_0^{\frac{s}{2}} \left( \frac{\psi(r^{\frac{1}{2}})}{\sqrt{r}} \right)^m dr \right)^{\frac{1}{m}},
\end{aligned}$$

$$\begin{aligned}
(3.9) \quad & \frac{\sqrt{s}}{\psi(s^{\frac{1}{2}})} \int_0^s \frac{\psi(r^{\frac{1}{2}})\psi((s-r)^{\frac{1}{2}})}{\sqrt{r}(s-r)} \rho_{\theta+r} dr \\
& = \frac{\sqrt{s}}{\psi(s^{\frac{1}{2}})} \left( \int_0^{\frac{s}{2}} \frac{\psi(s^{\frac{1}{2}})}{\sqrt{s/2}} \cdot \frac{\psi(r^{\frac{1}{2}})}{\sqrt{r}} \rho_{\theta+r} dr + \int_{\frac{s}{2}}^s \frac{\psi((s-r)^{\frac{1}{2}})}{\sqrt{s-r}} \cdot \frac{\sqrt{s}\psi(s^{\frac{1}{2}})}{s/\sqrt{2}} \rho_{\theta+r} dr \right) \\
& \leq 2\sqrt{2} \int_0^s \left( \frac{\psi(r^{\frac{1}{2}})}{\sqrt{r}} + \frac{\psi((s-r)^{\frac{1}{2}})}{\sqrt{s-r}} \right) \rho_{\theta+r} dr \leq 4\sqrt{2} \left( \int_0^s \frac{\psi(r^{\frac{1}{2}})^2}{r} dr \right)^{\frac{1}{2}} \left( \int_0^T \rho_r^2 dr \right)^{\frac{1}{2}}.
\end{aligned}$$

Combining these with (1.10), we conclude that  $h(t)$  defined in (3.6) satisfies  $h(t) \rightarrow 0$  as  $t \rightarrow 0$ . Letting  $r_0 > 0$  such that  $k_1 h(t) \leq \frac{1}{2}$  for  $t \in [0, r_0]$ , we deduce from (3.2) and (3.6) that

$$\frac{\sqrt{t-s}}{\psi((t-s)^{\frac{1}{2}})} (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,t}^* \gamma, P_{s,t}^* \tilde{\gamma}) \leq \Gamma_{s,t} \leq 2k_1 \mathbb{W}_k(\tilde{\gamma}, \gamma)$$

holds for all  $s \in [0, T)$  and  $t \in (s, (s + r_0) \wedge T]$ . Consequently,

$$\begin{aligned}
& (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,t}^* \gamma, P_{s,t}^* \tilde{\gamma}) \leq \frac{2k_1 \psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}} \mathbb{W}_k(\gamma, \tilde{\gamma}), \\
& s \in [0, T), t \in (s, (s + r_0) \wedge T], \gamma, \tilde{\gamma} \in \mathcal{P}_k.
\end{aligned}$$

Combining this with the flow property

$$P_{s,t}^* = P_{r,t}^* P_{s,r}^*, \quad 0 \leq s \leq r \leq t \leq T,$$

we find a constant  $k_2 > 0$  such that

$$(3.10) \quad (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,t}^* \gamma, P_{s,t}^* \tilde{\gamma}) \leq \frac{k_2 \psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}} \mathbb{W}_k(\gamma, \tilde{\gamma}), \quad t \in (s, T], \gamma, \tilde{\gamma} \in \mathcal{P}_k.$$

By the conditions on  $\psi$  in  $(B^{a,b})(3)$  and (1.10), we have

$$\sup_{t \in (0, T]} \left\{ \int_0^t \frac{\psi(r^{\frac{1}{2}}) \psi((t-r)^{\frac{1}{2}})}{r \sqrt{t-r}} (1 + \rho_r \sqrt{r}) dr + \left( \int_0^t \left( \frac{\psi(r^{\frac{1}{2}})}{\sqrt{r}} \right)^2 dr \right)^{\frac{1}{2}} + \left( \int_0^t \left( \frac{\psi((t-r)^{\frac{1}{2}}) \psi(r^{\frac{1}{2}})}{\sqrt{r} \sqrt{t-r}} \right)^m dr \right)^{\frac{1}{m}} \right\} < \infty.$$

Therefore, substituting (3.10) into (3.4) and (3.5), we derive (1.11) for some constant  $c > 0$ .

## 4 Proof of Theorem 1.5

(a) We use the notations in step (c) in the proof of Theorem 1.3. By Pinsker's inequality, [13, (1.3)] and  $(B^{a,b})$  with  $\|\rho\|_\infty < \infty$ , we find constants  $\varepsilon \in (0, \frac{1}{2}]$ ,  $c_1 > 0$  such that

$$\begin{aligned} \|P_{s,t}^{1,x} - P_{s,t}^{2,y}\|_{var} &\leq \sqrt{2 \text{Ent}(P_{s,t}^{1,x} | P_{s,t}^{2,y})} \\ &\leq \frac{c_1 |x - y|}{\sqrt{t-s}} + \frac{c_1}{\sqrt{t-s}} \left( \int_s^t (\mathbb{W}_\psi + \mathbb{W}_k)^2(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) dr \right)^{\frac{1}{2}} \\ &\quad + c_1 \sqrt{\log(1 + (t-s)^{-1})} \sup_{r \in [s+\varepsilon(t-s), t]} (\mathbb{W}_\psi + \mathbb{W}_k)^2(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) dr, \quad t \in [s, T]. \end{aligned}$$

Combining this with (3.3) and Lemma 2.1, we obtain

$$\begin{aligned} (4.1) \quad &\mathbb{W}_\psi(P_{s,t}^* \gamma, P_{s,t}^* \tilde{\gamma}) - \frac{\psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}} \mathbb{W}_1(P_{s,t}^{1,\gamma}, P_{s,t}^{2,\tilde{\gamma}}) \leq \psi((t-s)^{\frac{1}{2}}) \|P_{s,t}^{1,\gamma} - P_{s,t}^{2,\tilde{\gamma}}\|_{var} \\ &\leq \frac{\psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}} \left( \int_s^t (\mathbb{W}_\psi + \mathbb{W}_k)^2(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) dr \right)^{\frac{1}{2}} \\ &\quad + c_1 \psi((t-s)^{\frac{1}{2}}) \sqrt{\log(1 + (t-s)^{-1})} \sup_{r \in [s+\varepsilon(t-s), t]} (\mathbb{W}_\psi + \mathbb{W}_k)(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) \end{aligned}$$

for  $t \in [s, T]$ . On the other hand, since  $b^{(0)}$  is bounded,  $\|b^{(0)}\|_{\tilde{L}_{q_0}^{p_0}(T)} < \infty$  holds for any  $p_0, q_0 > 2$ , so that (1.6) holds for  $m = 2$ . Then there exists a constant  $c_2 > 0$  such that

$$\begin{aligned} (4.2) \quad &\mathbb{W}_1(P_{s,t}^{1,\gamma}, P_{s,t}^{2,\tilde{\gamma}}) \leq \mathbb{W}_k(P_{s,t}^{1,\gamma}, P_{s,t}^{2,\tilde{\gamma}}) \\ &\leq c_2 \mathbb{W}_k(\gamma, \tilde{\gamma}) + c_2 \left( \int_s^t (\mathbb{W}_\psi + \mathbb{W}_k)^2(P_{s,r}^* \gamma, P_{s,r}^* \tilde{\gamma}) dr \right)^{\frac{1}{2}}. \end{aligned}$$

Combining this with (4.1), we find a constant  $c_3 > 0$  such that instead of (3.6) we have

$$(4.3) \quad \begin{aligned} \Gamma_{s,t} &\leq c_3 \mathbb{W}_k(\gamma, \tilde{\gamma}) + c_2 h(t-s) \Gamma_{s,t}, \quad 0 \leq s \leq t \leq T, \\ h(t) &:= \left( \int_0^t \frac{\psi(s^{\frac{1}{2}})^2}{s} ds \right)^{\frac{1}{2}} + \sup_{r \in (0,t]} \psi(r^{\frac{1}{2}}) \sqrt{\log(1+r^{-1})}, \quad t > 0. \end{aligned}$$

Since  $\int_0^1 \frac{\psi(r)^2}{r} dr < \infty$ , we have  $h(t) \rightarrow 0$  as  $t \rightarrow 0$  if  $\lim_{r \rightarrow 0} \psi(r)^2 \log(1+r^{-1}) = 0$ , so that (1.11) follows as explained in step (c) in the proof of Theorem 1.3.

(b) Next, by (3.3), [13, (1.3)] and  $(B^{a,b})$  with  $\|\rho\|_\infty < \infty$ , we find constants  $\varepsilon \in (0, \frac{1}{2}]$ ,  $c_1 > 0$  such that for any  $\gamma, \tilde{\gamma} \in \mathcal{P}_k$ ,

$$\begin{aligned} \text{Ent}(P_t^* \gamma | P_t^* \tilde{\gamma}) &\leq \frac{\mathbb{W}_2(\gamma, \tilde{\gamma})^2}{t} + \frac{c_1}{t} \int_0^t (\mathbb{W}_\psi + \mathbb{W}_k)^2(P_r^* \gamma, P_r^* \tilde{\gamma}) dr \\ &\quad + c_1 \log(1+t^{-1}) \sup_{r \in [\varepsilon t, t]} (\mathbb{W}_\psi + \mathbb{W}_k)^2(P_r^* \gamma, P_r^* \tilde{\gamma}), \quad t \in (0, T]. \end{aligned}$$

Combining this with (1.11), we find a constant  $c > 0$  such that (1.12) holds.

(c) If either  $\|b\|_\infty < \infty$  or (1.13) holds, then we may apply [13, (1.4)] to delete the term  $\log(1+(t-s)^{-1})$  from the above calculations, so that  $h(t)$  in (4.3) becomes  $(\int_0^t \frac{\psi(s^{\frac{1}{2}})^2}{s} ds)^{\frac{1}{2}}$  which goes to 0 as  $t \rightarrow 0$ . Therefore, (1.11) and (1.14) hold for some constant  $c > 0$  as shown above.

## References

- [1] V. I. Bogachev, M. Röckner, S. V. Shaposhnikov, *Distances between transition probabilities of diffusions and applications to nonlinear Fokker-Planck-Kolmogorov equations*, J. Funct. Anal. 271(2016), 1262-1300.
- [2] P.-E. Chaudru de Raynal, N. Frikha, *Well-posedness for some non-linear SDEs and related PDE on the Wasserstein space*, J. Math. Pures Appl. 159(2022), 1-167.
- [3] X. Huang, *McKean-Vlasov SDEs with singularity in distribution variable and distribution dependent diffusion*, arXiv:2302.05845.
- [4] X. Huang, M. Röckner, F.-Y. Wang, *Non-linear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs*, Discrete Contin. Dyn. Syst. 39(2019), 3017-3035.
- [5] X. Huang, Y. Song, *Well-posedness and regularity for distribution dependent SPDEs with singular drifts*, Nonlinear Anal. 203(2021), 112167.
- [6] X. Huang, F.-Y. Wang, *Distribution dependent SDEs with singular coefficients*, Stochastic Process. Appl. 129(2019), 4747-4770.

- [7] X. Huang, F.-Y. Wang, *Singular McKean-Vlasov (reflecting) SDEs with distribution dependent noise*, J. Math. Anal. Appl. 514(2022), 126301 21pp.
- [8] X. Huang, F.-Y. Wang, *Log-Harnack inequality and Bismut formula for singular McKean-Vlasov SDEs*, arXiv:2207.11536.
- [9] X. Huang, P. Ren, F.-Y. Wang, *Distribution dependent stochastic differential equations*, Front. Math. China 16(2021), 257–301.
- [10] S. Menozzi, A. Pesce, X. Zhang, *Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift*, J. Diff. Equat. 272(2021), 330–369.
- [11] P. Ren, *Singular McKean-Vlasov SDEs: well-posedness, regularities and Wang’s Harnack inequality*, Stoch. Proc. Appl. 156(2023), 291–311.
- [12] P. Ren, F.-Y. Wang, *Exponential convergence in entropy and Wasserstein for McKean-Vlasov SDEs*, Nonlinear Anal. 206(2021), 112259.
- [13] P. Ren, F.-Y. Wang, *Entropy estimate between diffusion processes and application to McKean-Vlasov SDEs*, arXiv:2302.13500.
- [14] F.-Y. Wang, *Distribution-dependent SDEs for Landau type equations*, Stoch. Proc. Appl. 128(2018), 595-621.
- [15] F.-Y. Wang, *Derivative formula for singular McKean-Vlasov SDEs*, Comm. Pure Appl. Anal. 22(2023), 1866–1898.
- [16] F.-Y. Wang, *Distribution dependent reflecting stochastic differential equations*, to appear in Sci. China Math. arXiv:2106.12737.
- [17] P. Xia, L. Xie, X. Zhang, G. Zhao,  *$L^q(L^p)$ -theory of stochastic differential equations*, Stochastic Process. Appl. 130(2020), 5188-5211.
- [18] L. Xie, X. Zhang, *Ergodicity of stochastic differential equations with jumps and singular coefficients*, Ann. Inst. Henri Poincaré Probab. Stat. 56(2020), 175-229.
- [19] S.-Q. Zhang, C. Yuan, *A study on Zvonkin’s transformation for stochastic differential equations with singular drift and related applications*, J. Diff. Equat. 297(2021), 277–319.