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

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

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, panparen@cityu.edu.hk, wangfy@tju.edu.cn

March 7, 2024

Abstract

The L^k-Wasserstein distance $\mathbb{W}_k (k \geq 1)$ and the probability distance $\mathbb{W}_\psi$ induced by a concave function ψ, 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 Ξ be the space of (a, b), where

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

*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 $\mathbb{R}^d$:

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

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

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

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 $\gamma\in\mathscr{P}$, where $\mathscr{P}$ 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

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

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

$$\mathrm{Ent}(P_{s,t}^{1,\gamma}|P_{s,t}^{2,\tilde\gamma}):=\int_{\mathbb{R}^d}\Big(\log\frac{\mathrm{d}P_{s,t}^{1,\gamma}}{\mathrm{d}P_{s,t}^{2,\tilde\gamma}}\Big)\mathrm{d}P_{s,t}^{1,\gamma},\quad 0\le s<t\le T,\gamma,\tilde\gamma\in\mathscr{P}$$

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

$$\mathscr{A}:=\big\{\psi:[0,\infty)\to[0,\infty)\text{ is increasing and concave, }\psi(r)>0\text{ for }r>0\big\}.$$

For any $\psi\in\mathscr{A}$, the ψ-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

$$\mathscr{P}_\psi:=\Big\{\mu\in\mathscr{P}:\ \|\mu\|_\psi:=\int_{\mathbb{R}^d}\psi(|x|)\mu(\mathrm{d}x)<\infty\Big\}$$

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

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

where $\mu(f) := \int_{\mathbb{R}^d} f \mathrm{d}\mu$ 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

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

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

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

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

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

is a Polish space under $\mathbb{W}_k$. Since ψ has at most linear growth, we have $\mathscr{P}_k \subset \mathscr{P}_\psi$, and $\mathscr{P}_k$ 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 $p \ge 1$, $L^p(\mathbb{R}^d)$ is the class of measurable functions f on $\mathbb{R}^d$ such that

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

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

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

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

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

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

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

We make the following assumptions for the coefficients $(a,b) \in \Xi$, where ∇ 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 $\{(p_i,q_i)\}_{0\le i\le l} \subset \mathscr{K}$ such that the following conditions hold.

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

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

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

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

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

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

Let $\sigma(t,x) := \sqrt{2a(t,x)}$, and let W_t be a d-dimensional Brownian motion on a probability basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\in[0,T]}, \mathbb{P})$. 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

$$\mathrm{d}X_{s,t}^x = b(t,X_{s,t}^x)\mathrm{d}t + \sigma(t,X_{s,t}^x)\mathrm{d}W_t, \quad t\in[s,T] \tag{1.3}$$

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

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

The associated Markov semigroup is given by

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

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

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

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

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

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* $(p_i,q_i)_{0\le i\le l}$, *such that for any* $\psi\in\mathscr{A}$ *and* $\gamma,\tilde\gamma\in\mathscr{P}$,

$$\begin{aligned}\mathbb{W}_\psi(P_{s,t}^{1,\gamma},P_{s,t}^{2,\tilde\gamma}) &\le \frac{c\psi((t-s)^{\frac12})}{\sqrt{t-s}}\mathbb{W}_1(\gamma,\tilde\gamma)+c\int_s^t\frac{\psi((t-r)^{\frac12})\|a_1-a_2\|_{r,\infty}}{\sqrt{(r-s)(t-r)}}\mathrm{d}r\\ &+c\bigg(\int_s^t\bigg(\frac{\psi((t-r)^{\frac12})\|a_1-a_2\|_{r,\infty}}{\sqrt{t-r}}\bigg)^m\mathrm{d}r\bigg)^{\frac1m}\\ &+c\int_s^t\frac{\psi((t-r)^{\frac12})}{\sqrt{t-r}}\big\{\|b_1-b_2\|_{r,\infty}+\|\mathrm{div}(a_1-a_2)\|_{r,\infty}\big\}\mathrm{d}r,\quad 0\le s<t\le T,\ \gamma,\tilde\gamma\in\mathscr{P}.\end{aligned}\tag{1.5}$$

Moreover, for any $k\ge1$, *there exists a constant* $C>0$ *depending only on* k,K,d,T *and* $(p_i,q_i)_{0\le i\le l}$, *such that for any* $\gamma,\tilde\gamma\in\mathscr{P}$ *and* $0\le s\le t\le T$,

$$\mathbb{W}_k(P_{s,t}^{1,\gamma},P_{s,t}^{2,\tilde\gamma})\le C\bigg[\mathbb{W}_k(\gamma,\tilde\gamma)+\int_s^t\|b_1-b_2\|_{r,\infty}\mathrm{d}r+\bigg(\int_s^t\|a_1-a_2\|_{r,\infty}^2\mathrm{d}r\bigg)^{\frac12}\bigg].\tag{1.6}$$

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

$$\mathrm{d}X_t=b_t(X_t,\mathscr{L}_{X_t})\mathrm{d}t+\sigma_t(X_t,\mathscr{L}_{X_t})\mathrm{d}W_t,\quad t\in[0,T],\tag{1.7}$$

where $\mathscr{L}_{X_t}$ is the distribution of X_t, and for some $k\ge1$,

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

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

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

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

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

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

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

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

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 μ, 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 μ, and the drift is singular in the time-spatial variables. For any $\mu \in C_b^w([0,T];\mathscr{P}_k)$, 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 $\psi \in \mathscr{A}$ with $\lim_{t\to 0}\psi(t) = 0$.

(1) $(A^{a,b})$ holds for $(a,b) = (a^\mu, b^\mu)$ uniformly in $\mu \in C_b^w([0,T];\mathscr{P}_k)$, with drift decomposition $b^\mu = (b^\mu)^{(0)} + (b^\mu)^{(1)}$.

(2) There exists a constant $K > 0$ such that

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

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

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

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 σ in distribution is given by singular integral kernels. Let $\psi \in \mathscr{A}$ with $\lim_{t\to 0}\psi(t) = 0$ and let*

$$\begin{aligned} b_t(\cdot,\mu) &= b_t^{(0)} + \int_{\mathbb{R}^d} \tilde b_t(\cdot,y)\mu(\mathrm{d}y),\\ \sigma_t(\cdot,\mu) &= \sqrt{\lambda I + \int_{\mathbb{R}^d} (\tilde\sigma_t\tilde\sigma_t^*)(\cdot,y)\mu(\mathrm{d}y)}, \quad (t,\mu) \in [0,T]\times\mathscr{P}_k, \end{aligned}$$

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

$$|\tilde b_t(x,y) - \tilde b_t(\tilde x,\tilde y)| \le K\big(|x-\tilde x| + \psi(|y-\tilde y|)\big), \quad x,\tilde x,y,\tilde y \in \mathbb{R}^d, t \in [0,T]$$

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

$$\begin{aligned} &\|\tilde\sigma_t(x,y) - \tilde\sigma_t(\tilde x,\tilde y)\| \le K\big(|x-\tilde x| + \psi(|y-\tilde y|)\big),\\ &|\nabla\tilde\sigma_t(\cdot,y)(x) - \nabla\tilde\sigma_t(\cdot,\tilde y)(x)| \le K\psi(|y-\tilde y|), \quad x,\tilde x,y,\tilde y \in \mathbb{R}^d, t \in [0,T]. \end{aligned}$$

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* $\mathscr{P}_k$. *Moreover, for any* $n \in \mathbb{N}$, *there exists a constant* $c > 0$ *such that any solution satisfies*

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X_t|^n \Big| \mathscr{F}_0\Big] \le c(1 + |X_0|^n). \tag{1.9}$$

(2) *If* ψ *is a Dini function, i.e.*

$$\int_0^1 \frac{\psi(s)}{s}\,\mathrm{d}s < \infty, \tag{1.10}$$

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

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

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* $k \ge 1$ *(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 ψ not satisfying (1.10), and to estimate the relative entropy $\mathrm{Ent}(P_t^*\gamma|P_t^*\tilde\gamma)$, we need the drift to be Dini continuous in the spatial variable.

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

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

Then the following assertions hold.

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

$$\begin{aligned} \mathrm{Ent}(P_t^*\gamma|P_t^*\tilde\gamma) \le{}& \frac{c\mathbb{W}_2(\gamma,\tilde\gamma)^2}{t}\\ &+ c\mathbb{W}_k(\gamma,\tilde\gamma)^2\Big(\frac{1}{t}\int_0^t \frac{\psi(r)^2}{r}\mathrm{d}r + \frac{\psi(t^{\frac{1}{2}})^2}{t}\log(1+t^{-1})\Big). \end{aligned} \tag{1.12}$$

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

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

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

$$\mathrm{Ent}(P_t^*\gamma|P_t^*\tilde\gamma) \le \frac{c\mathbb{W}_2(\gamma,\tilde\gamma)^2}{t} + \frac{c\mathbb{W}_k(\gamma,\tilde\gamma)^2}{t}\int_0^t \frac{\psi(r)^2}{r}\mathrm{d}r, \quad t\in(0,T], \gamma,\tilde\gamma\in\mathscr{P}_k. \tag{1.14}$$

Remark 1.6. *When* $k \le 2$, (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* $\psi \in \mathscr{A}$,

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

Proof. Since ψ is nonnegative and concave, we have

$$\psi(Rr) \le R\psi(r), \quad r\ge 0, R\ge 1. \tag{2.1}$$

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

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

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 $[f]_\psi \le 1$, Jensen's inequality and (2.1), we obtain

$$|f_t(x)-f(x)| \le \mathbb{E}[\psi(|B_t|)] \le \psi(\mathbb{E}|B_t|) \le \psi((dt)^{\frac12}) \le \sqrt{d}\psi(t^{\frac12}), \quad t\ge 0, x\in\mathbb{R}^d.$$

So,

$$\sup_{[f]_\psi\le 1}\big|\gamma(f_t-f)-\tilde\gamma(f_t-f)\big| \le \sqrt{d}\psi(t^{\frac12})\|\gamma-\tilde\gamma\|_{var}, \quad t\ge 0. \tag{2.2}$$

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

$$|\nabla f_t(x)| = \left|\nabla_x \int_{\mathbb{R}^d}(2\pi t)^{-\frac d2}\mathrm{e}^{-\frac{|x-y|^2}{2t}}(f(y)-f(z))\mathrm{d}y\right|_{z=x}$$

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

Combining this with (2.2) and noting that

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

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

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

Then the proof is finished. □

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

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

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

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

$$
\nabla P^{a,b}_{s,t} f(x) = \mathbb{E}\big[f(X^x_{s,t}) M^x_{s,t}\big] \tag{2.3}
$$

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

$$
\mathbb{E}[M^x_{s,t}] = 0, \quad \mathbb{E}|M^x_{s,t}|^2 \le c_1^2 (t-s)^{-1}. \tag{2.4}
$$

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

$$
|\nabla P^{a,b}_{s,t} f(x)| = \Big|\mathbb{E}\big[\{f(X^x_{s,t}) - f(z)\} M^x_{s,t}\big]\Big| \le \mathbb{E}\big[\psi(|X^x_{s,t} - z|) |M^x_{s,t}|\big].
$$

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

$$
\begin{aligned}
|\nabla P^{a,b}_{s,t} f(x)| &\le \mathbb{E}[|M^x_{s,t}|] \psi\bigg(\frac{\mathbb{E}[|X^x_{s,t} - z| \cdot |M^x_{s,t}|]}{\mathbb{E}[|M^x_{s,t}|]}\bigg) \\
&\le \mathbb{E}[|M^x_{s,t}|] \psi\bigg(\frac{(\mathbb{E}[|M^x_{s,t}|]^2)^{\frac{1}{2}}}{\mathbb{E}[|M^x_{s,t}|]} \big(\mathbb{E}|X^x_{s,t} - z|^2\big)^{\frac{1}{2}}\bigg).
\end{aligned}
$$

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

$$\sup_{[f]_\psi\le 1}|\nabla P^{a,b}_{s,t}f(x)|\le c_1(t-s)^{-\frac{1}{2}}\inf_{z\in\mathbb{R}^d}\psi\Big(\big\{\mathbb{E}|X^x_{s,t}-z|^2\big\}^{\frac{1}{2}}\Big),\quad 0\le s<t\le T, x\in\mathbb{R}^d. \tag{2.5}$$

(b) To estimate $\inf_{z\in\mathbb{R}^d}\mathbb{E}|X^x_{s,t}-z|^2$, 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

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

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

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

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

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

$$\mathrm{d}Y_{s,t}=\bar b(t,Y_{s,t})\mathrm{d}t+\bar\sigma(t,Y_{s,t})\mathrm{d}W_t,\quad t\in[s,T], Y_{s,s}=x+u_s(x),$$

where

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

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

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

Let

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

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

$$\begin{aligned}\mathrm{d}|Y_{s,t}-\theta_{s,t}|^2&=\Big\{2\big\langle Y_{s,t}-\theta_{s,t},\bar b(t,Y_{s,t})-\bar b(t,\theta_{s,t})\big\rangle+\|\bar\sigma(t,Y_{s,t})\|^2_{HS}\Big\}\mathrm{d}t+\mathrm{d}M_t\\&\le c_2\Big\{|Y_{s,t}-\theta_{s,t}|^2+1\Big\}\mathrm{d}t+\mathrm{d}M_t,\quad t\in[s,T],|Y_{s,s}-\theta_{s,s}|=0.\end{aligned}$$

Thus,

$$\mathbb{E}\big[|Y_{s,t}-\theta_{s,t}|^2\big]\le c_2\mathrm{e}^{c_2T}(t-s),\quad 0\le s\le t\le T.$$

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

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

Combining this with (2.5) and (2.1), we finish the proof. □

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

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

be the standard Gaussian heat kernel with parameter κ.

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 $0 \le g_{\cdot,t} \in \mathscr{B}([0,t])$,*

$$\begin{aligned}&\int_s^t \frac{g_{r,t}}{\sqrt{t-r}} \mathrm{d}r \int_{\mathbb{R}^d} |\nabla_y p^{a,b}_{s,r}(x,y)| \mathrm{d}y \\ &\le c(m) \int_s^t \frac{g_{r,t}}{\sqrt{(t-r)(r-s)}} \mathrm{d}r + c(m) \bigg(\int_s^t \Big(\frac{g_{r,t}}{\sqrt{t-r}} \Big)^m \mathrm{d}r \bigg)^{\frac{1}{m}}, \quad s \in [0,t].\end{aligned} \tag{2.11}$$

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

$$\int_s^t (t-r)^{-\frac{1}{2}} \mathrm{d}r \int_{\mathbb{R}^d} |\nabla_y p^{a,b}_{s,r}(x,y)| \mathrm{d}y \le c, \quad 0 \le s < t \le T. \tag{2.12}$$

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

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

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

$$|\nabla^i_y \bar p_{s,t}(x,y)| \le \kappa (t-s)^{-\frac{i}{2}} g_\kappa(t-s, \theta_{s,t}(x) - y), \quad 0 \le s < t \le T, x, y \in \mathbb{R}^d, i = 0, 1, \tag{2.13}$$

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

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

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

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

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

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

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

$$\begin{aligned}(2.17)\qquad &\int_s^t \mathrm{d}r\int_{\mathbb{R}^d}\|\nabla^2 u_r(y)\|^{\frac{m}{m-1}} p^{a,b}_{s,r}(x,y)\mathrm{d}y\\ &=\mathbb{E}\int_s^t \|\nabla^2 u_r\|^{\frac{m}{m-1}}(X^x_{s,r})\mathrm{d}r\le c\|\|\nabla^2 u\|^{\frac{m}{m-1}}\|_{\tilde L^{\tilde p}_{\tilde q}(s,t)} = c(\|\nabla^2 u\|_{\tilde L^{p_0}_{q_0}(s,t)})^{\frac{m}{m-1}}.\end{aligned}$$

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

$$\begin{aligned}&\int_s^t \frac{g_{r,t}}{\sqrt{t-r}}\mathrm{d}r\int_{\mathbb{R}^d}|\nabla_y p^{a,b}_{s,r}(x,y)|\mathrm{d}y\le c_2\kappa\int_s^t g_{r,t}(t-r)^{-\frac12}(r-s)^{-\frac12}\mathrm{d}r\\ &\quad + c_2\bigg(\int_s^t\Big(\frac{g_{r,t}}{\sqrt{t-r}}\Big)^m\mathrm{d}r\bigg)^{\frac1m}\bigg(\int_s^t \mathrm{d}r\int_{\mathbb{R}^d}\|\nabla^2 u_r(y)\|^{\frac{m}{m-1}}p^{a,b}_{s,r}(x,y)\mathrm{d}y\bigg)^{\frac{m-1}{m}}\\ &\le c(m)\int_s^t\frac{g_{r,t}}{\sqrt{(t-r)(r-s)}}\mathrm{d}r + c(m)\bigg(\int_s^t\Big(\frac{g_{r,t}}{\sqrt{t-r}}\Big)^m\mathrm{d}r\bigg)^{\frac1m}.\end{aligned}$$

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 $[f]_\psi\le 1$. By Itô's formula we have

$$P^{a_2,b_2}_{s,t}f(x)=f(x)+\int_s^t P^{a_2,b_2}_{s,r}(L^{a_2,b_2}_r f)(x)\mathrm{d}r,\quad 0\le s\le t\le T.$$

This implies the Kolmogorov forward equation

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

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

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

By [19, Theorem 2.1], the PDE

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

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

$$\mathrm{d}u_r(X^{2,x}_{s,r})=-L^{a_2,b_2}_r f(X^{2,x}_{s,r})+\mathrm{d}M_r,\quad r\in[s,t]$$

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

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

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

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

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

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

By the integration by parts formula, we obtain

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

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

$$|P_{s,t}^{a_1,b_1} f(x) - P_{s,t}^{a_2,b_2} f(x)| \le c_1 \int_s^t \frac{\psi((t-r)^{\frac{1}{2}})\|a_1-a_2\|_{r,\infty}}{\sqrt{t-r}}\mathrm{d}r \int_{\mathbb{R}^d} |\nabla_y p_{s,r}^{a_1,b_1}(x,y)|\mathrm{d}y$$
$$+ c_1 \int_s^t \frac{\psi((t-r)^{\frac{1}{2}})}{(t-r)^{\frac{1}{2}}}\big(\|b_1-b_2\|_{r,\infty} + \|\mathrm{div}(a_1-a_2)\|_{r,\infty}\big)\mathrm{d}r$$
$$\le c_2 \int_s^t \frac{\psi((t-r)^{\frac{1}{2}})}{\sqrt{t-r}}\left(\frac{\|a_1-a_2\|_{r,\infty}}{\sqrt{r-s}} + \|b_1-b_2\|_{r,\infty} + \|\mathrm{div}(a_1-a_2)\|_{r,\infty}\right)\mathrm{d}r$$
$$+ c_2\left(\int_s^t \left(\frac{\psi((t-r)^{\frac{1}{2}})\|a_1-a_2\|_{r,\infty}}{\sqrt{t-r}}\right)^m \mathrm{d}r\right)^{\frac{1}{m}} =: I_{s,t}.$$

Therefore,

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

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

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

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 $c_3 > 0$,

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

holds for any $0 \le s < t \le T$ 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

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

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

$$
\begin{aligned}
\mathrm{d}Y^{1,x}_{s,t} &= \big\{b^{(1)}_1(t,\cdot) + \lambda u_t\big\}(X^{1,x}_{s,t})\mathrm{d}t + \big\{(\nabla\Theta_t)\sigma_1(t,\cdot)\big\}(X^{1,x}_{s,t})\mathrm{d}W_t, \\
\mathrm{d}Y^{2,y}_{s,t} &= \big\{b^{(1)}_1(t,\cdot) + \lambda u_t\big\}(X^{2,y}_{s,t})\mathrm{d}t + \big\{(\nabla\Theta_t)(b_2 - b_1) + \mathrm{tr}[(a_2 - a_1)(t,\cdot)\nabla^2 u_t]\big\}(X^{2,y}_{s,t})\mathrm{d}t \\
&\quad + \big\{(\nabla\Theta_t)\sigma_2(t,\cdot)\big\}(X^{2,y}_{s,t})\mathrm{d}W_t, \quad t \in [s,T], \ Y^{1,x}_{s,s} = \Theta_s(x), \ Y^{2,y}_{s,s} = \Theta_s(y).
\end{aligned}
$$

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

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

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 $k \ge 1$ we find a constant $c_1 > 1$ such that

$$
c_1^{-1}|X^{1,x}_{s,t} - X^{2,y}_{s,t}|^{2k} \le \xi_t := |Y^{1,x}_{s,t} - Y^{2,y}_{s,t}|^{2k} \le c_1|X^{1,x}_{s,t} - X^{2,y}_{s,t}|^{2k}, \tag{2.23}
$$

$$
\mathrm{d}\xi_t \le c_1\xi_t(1+\eta_t)\mathrm{d}t + c_1\xi_t^{\frac{2k-1}{2k}}\gamma_t\mathrm{d}t + c_1\xi_t^{\frac{k-1}{k}}\|a_1 - a_2\|^2_{t,\infty}\mathrm{d}t + \mathrm{d}M_t, \tag{2.24}
$$

where M_t is a martingale and

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

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

$$
\mathbb{E}\bigg\{\Big(\sup_{r\in[s,t]}\xi_r^q\Big)^{\frac{2k-1}{2kq}}\int_s^t \|a_1 - a_2\|_{r,\infty}\|\nabla^2 u_r\|(X^{2,y}_{s,r})\mathrm{d}r\bigg\}
$$

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

$$
\le \left(\mathbb{E}\sup_{r\in[s,t]}\xi_r^q\right)^{\frac{2k-1}{2kq}} \left(\int_s^t \|a_1-a_2\|_{r,\infty}^m \mathrm{d}r\right)^{\frac{1}{m}}
$$

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

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

$$
\left[\mathbb{E}\sup_{r\in[s,t]}\xi_r^q\right]^{\frac{1}{q}} \le c_2|x-y|^{2k} + c_2\mathbb{E}\int_s^t \left\{\xi_r^{\frac{2k-1}{2k}}\gamma_r\mathrm{d}r + \xi_r^{\frac{k-1}{k}}\|a_1-a_2\|_{r,\infty}^2\right\}\mathrm{d}r
$$

$$
\le c_2|x-y|^{2k} + c_2\mathbb{E}\left[\Big(\sup_{r\in[s,t]}\xi_r^q\Big)^{\frac{2k-1}{2kq}}\int_s^t\gamma_r\mathrm{d}r + \Big(\sup_{r\in[s,t]}\xi_r^q\Big)^{\frac{k-1}{kq}}\int_s^t\|a_1-a_2\|_{r,\infty}^2\mathrm{d}r\right]
$$

$$
\le c_2|x-y|^{2k} + \frac{1}{2}\left[\mathbb{E}\sup_{r\in[s,t]}\xi_r^q\right]^{\frac{1}{q}} + c_3\Big(\int_s^t\|a_1-a_2\|_{r,\infty}^2\mathrm{d}r\Big)^k + c_3\Big(\int_s^t\|b_1-b_2\|_{r,\infty}\mathrm{d}r\Big)^{2k}
$$

$$
+ c_3\Big(\int_s^t\|a_1-a_2\|_{r,\infty}^m\mathrm{d}r\Big)^{\frac{2k}{m}} \left(\mathbb{E}\left(\int_s^t \|\nabla^2 u_r\|^{\frac{m}{m-1}}(X_{s,r}^{2,y})\mathrm{d}r\right)^{\frac{2(m-1)kq}{m(2kq-2k+1)}}\right)^{\frac{2kq-2k+1}{q}}, \quad m>1.
$$

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

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

we obtain

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

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

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

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

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

$$\begin{aligned}\big(\mathbb{E}|Y^{1,x}_{s,t} - Y^{2,y}_{s,t}|^k\big)^2 &\le \Big[\mathbb{E}\sup_{r\in[s,t]} \xi_r^q\Big]^{\frac{1}{q}} \le c_5|x-y|^{2k} + c_5\bigg(\int_s^t \|b_1 - b_2\|_{r,\infty}\mathrm{d}r\bigg)^{2k} \\ &+ c_5\bigg(\int_s^t \|a_1 - a_2\|^m_{r,\infty}\mathrm{d}r\bigg)^{\frac{2k}{m}} + c_5\bigg(\int_s^t \|a_1 - a_2\|^2_{r,\infty}\mathrm{d}r\bigg)^k.\end{aligned}$$

Noting that (2.23) implies

$$\mathbb{W}_k(P^{1,x}_{s,t}, P^{2,y}_{s,t})^k \le \sqrt{c_1}\mathbb{E}|Y^{1,x}_{s,t} - Y^{2,y}_{s,t}|^k,$$

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 $\mathscr{F}_0$-measurable with $\gamma := \mathscr{L}_{X_0} \in \mathscr{P}_k$. Let

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

For any $\lambda \ge 0$, C^γ_T is a complete space under the metric

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

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

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

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

$$\mathrm{d}X^\mu_t = b^\mu_t(X^\mu_t)\mathrm{d}t + \sigma^\mu_t(X^\mu_t)\mathrm{d}W_t, \quad t \in [0,T], X^\mu_0 = X_0$$

is well-posed, and

$$\mathbb{E}\Big[\sup_{s\in[0,T]} |X^\mu_t|^k\Big] < \infty.$$

So, we define a map

$$\Phi^\gamma : \mathscr{C}^\gamma_T \to \mathscr{C}^\gamma_T; \quad \mu \mapsto \big\{(\Phi^\gamma\mu)_t := \mathscr{L}_{X^\mu_t}\big\}_{t\in[0,T]}.$$

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

(b) Let $\tilde\gamma \in \mathscr{P}_k$ which may be different from γ, and let $\tilde\mu \in \mathscr{C}_T^{\tilde\gamma}$. We estimate the ρ_λ-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

$$\begin{aligned}
&\mathbb{W}_\psi\big((\Phi^\gamma\mu)_t, (\Phi^{\tilde\gamma}\tilde\mu)_t\big) + \mathbb{W}_k\big((\Phi^\gamma\mu)_t, (\Phi^{\tilde\gamma}\tilde\mu)_t\big)\\
&\le \frac{c_1\psi(t^{\frac12})}{\sqrt t}\mathbb{W}_k(\gamma,\tilde\gamma) + c_1\bigg(\int_0^t \|a^\mu - a^{\tilde\mu}\|_{r,\infty}^2 \mathrm{d}r\bigg)^{\frac12}\\
&\quad + c_1\left(\int_0^t \left(\frac{\psi((t-r)^{\frac12})\|a^\mu - a^{\tilde\mu}\|_{r,\infty}}{\sqrt{t-r}}\right)^m \mathrm{d}r\right)^{\frac1m}\\
&\quad + c_1\int_0^t \frac{c_1\psi((t-r)^{\frac12})}{\sqrt{t-r}}\Big(\frac{\|a^\mu - a^{\tilde\mu}\|_{r,\infty}}{\sqrt r} + \|b^\mu - b^{\tilde\mu}\|_{r,\infty} + \|\mathrm{div}(a^\mu - a^{\tilde\mu})\|_{r,\infty}\Big)\mathrm{d}r\\
&\le \frac{c_1\psi(t^{\frac12})}{\sqrt t}\mathbb{W}_k(\gamma,\tilde\gamma) + c_2\bigg(\int_0^t \Big(\mathbb{W}_\psi(\mu_r,\tilde\mu_r) + \mathbb{W}_k(\mu_r,\tilde\mu_r)\Big)^2\mathrm{d}r\bigg)^{\frac12}\\
&\quad + c_2\left(\int_0^t \left(\frac{\psi((t-r)^{\frac12})(\mathbb{W}_\psi(\mu_r,\tilde\mu_r) + \mathbb{W}_k(\mu_r,\tilde\mu_r))}{\sqrt{t-r}}\right)^m \mathrm{d}r\right)^{\frac1m}\\
&\quad + c_2\int_0^t \frac{\psi((t-r)^{\frac12})}{\sqrt{r(t-r)}}(1+\sqrt r\rho_r)\Big(\mathbb{W}_\psi(\mu_r,\tilde\mu_r) + \mathbb{W}_k(\mu_r,\tilde\mu_r)\Big)\mathrm{d}r
\end{aligned}$$

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

$$\rho_\lambda(\Phi^\gamma\mu, \Phi^\gamma\tilde\mu) \le \delta(\lambda)\rho_\lambda(\mu,\tilde\mu),$$

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

$$\begin{aligned}
\delta(\lambda) := c_2 \sup_{t\in[0,T]}\bigg[\int_0^t \frac{\psi((t-r)^{\frac12})\mathrm{e}^{-\lambda(t-r)}}{\sqrt{t-r}}\Big(\frac{1}{\sqrt r} + \rho_r\Big)\mathrm{d}r + \bigg(\int_0^t \mathrm{e}^{-2\lambda(t-r)}\mathrm{d}r\bigg)^{\frac12}\bigg]\\
+ c_2\left(\int_0^t \left(\frac{\psi((t-r)^{\frac12})\mathrm{e}^{-\lambda(t-r)}}{\sqrt{t-r}}\right)^m \mathrm{d}r\right)^{\frac1m} \to 0.
\end{aligned}$$

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

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

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

Since ψ 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 $\gamma, \tilde{\gamma} \in \mathscr{P}_k$ and $s \in [0,T)$,

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

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

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

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

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

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

$$\begin{aligned} &\mathbb{W}_k(P^*_{s,t}\gamma, P^*_{s,t}\tilde{\gamma}) = \mathbb{W}_k(P^{1,\gamma}_{s,t}, P^{2,\tilde{\gamma}}_{s,t}) \le k_0\mathbb{W}_k(\gamma,\tilde{\gamma})\\ &+ k_0\int_s^t \rho_r\big(\mathbb{W}_\psi + \mathbb{W}_k\big)(P^*_{s,r}\gamma, P^*_{s,r}\tilde{\gamma})\mathrm{d}r + k_0\bigg(\int_s^t \big(\mathbb{W}_\psi + \mathbb{W}_k\big)^2(P^*_{s,r}\gamma, P^*_{s,r}\tilde{\gamma})\mathrm{d}r\bigg)^{\frac{1}{2}}. \end{aligned} \tag{3.5}$$

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

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

Note that

$$
\begin{aligned}
&\frac{\sqrt{s}}{\psi(s^{\frac12})}\int_0^s \frac{\psi(r^{\frac12})\psi((s-r)^{\frac12})}{r\sqrt{s-r}}\mathrm{d}r\\
&\le \frac{\sqrt{s}}{\psi(s^{\frac12})}\bigg(\int_0^{\frac{s}{2}}\frac{\psi(s^{\frac12})}{\sqrt{s/2}}\cdot\frac{\psi(r^{\frac12})}{r}\mathrm{d}r+\int_{\frac{s}{2}}^s\frac{\psi((s-r)^{\frac12})}{s-r}\cdot\frac{\sqrt{s}\psi(s^{\frac12})}{s/2}\mathrm{d}r\bigg)\\
&\le \big(2+\sqrt{2}\big)\int_0^{\frac{s}{2}}\frac{\psi(r^{\frac12})}{r}\mathrm{d}r = 2\big(2+\sqrt{2}\big)\int_0^{\sqrt{s/2}}\frac{\psi(r)}{r}\mathrm{d}r.
\end{aligned}\tag{3.7}
$$

Similarly, we have

$$
\begin{aligned}
&\frac{\sqrt{s}}{\psi(s^{\frac12})}\left(\int_0^s\left(\frac{\psi((s-r)^{\frac12})\psi(r^{\frac12})}{\sqrt{r}\sqrt{s-r}}\right)^m\mathrm{d}r\right)^{\frac1m}\\
&\le\sqrt{2}\left(\left(\int_0^{\frac{s}{2}}\left(\frac{\psi(r^{\frac12})}{\sqrt{r}}\right)^m\mathrm{d}r\right)^{\frac1m}+\left(\int_{\frac{s}{2}}^s\left(\frac{\psi((s-r)^{\frac12})}{s-r}\right)^m\mathrm{d}r\right)^{\frac1m}\right)\\
&\le 2\sqrt{2}\left(\int_0^{\frac{s}{2}}\left(\frac{\psi(r^{\frac12})}{\sqrt{r}}\right)^m\mathrm{d}r\right)^{\frac1m},
\end{aligned}\tag{3.8}
$$

$$
\begin{aligned}
&\frac{\sqrt{s}}{\psi(s^{\frac12})}\int_0^s\frac{\psi(r^{\frac12})\psi((s-r)^{\frac12})}{\sqrt{r(s-r)}}\rho_{\theta+r}\mathrm{d}r\\
&= \frac{\sqrt{s}}{\psi(s^{\frac12})}\bigg(\int_0^{\frac{s}{2}}\frac{\psi(s^{\frac12})}{\sqrt{s/2}}\cdot\frac{\psi(r^{\frac12})}{\sqrt{r}}\rho_{\theta+r}\mathrm{d}r+\int_{\frac{s}{2}}^s\frac{\psi((s-r)^{\frac12})}{\sqrt{s-r}}\cdot\frac{\sqrt{s}\psi(s^{\frac12})}{s/\sqrt{2}}\rho_{\theta+r}\mathrm{d}r\bigg)\\
&\le 2\sqrt{2}\int_0^s\Big(\frac{\psi(r^{\frac12})}{\sqrt{r}}+\frac{\psi((s-r)^{\frac12})}{\sqrt{s-r}}\Big)\rho_{\theta+r}\mathrm{d}r\le 4\sqrt{2}\bigg(\int_0^s\frac{\psi(r^{\frac12})^2}{r}\mathrm{d}r\bigg)^{\frac12}\bigg(\int_0^T\rho_r^2\mathrm{d}r\bigg)^{\frac12}.
\end{aligned}\tag{3.9}
$$

Combining these with (1.10), we conclude that $h(t)$ defined in (3.6) satisfies $h(t)\to 0$ as $t\to 0$. Letting $r_0>0$ such that $k_1h(t)\le\frac12$ for $t\in[0,r_0]$, we deduce form (3.2) and (3.6) that

$$\frac{\sqrt{t-s}}{\psi((t-s)^{\frac12})}(\mathbb{W}_\psi+\mathbb{W}_k)(P_{s,t}^*\gamma,P_{s,t}^*\tilde\gamma)\le\Gamma_{s,t}\le 2k_1\mathbb{W}_k(\tilde\gamma,\gamma)$$

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

$$
\begin{gathered}
(\mathbb{W}_\psi+\mathbb{W}_k)(P_{s,t}^*\gamma,P_{s,t}^*\tilde\gamma)\le\frac{2k_1\psi((t-s)^{\frac12})}{\sqrt{t-s}}\mathbb{W}_k(\gamma,\tilde\gamma),\\
s\in[0,T), t\in(s,(s+r_0)\wedge T],\ \gamma,\tilde\gamma\in\mathscr{P}_k.
\end{gathered}
$$

Combining this with the flow property

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

we find a constant $k_2 > 0$ such that

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

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

$$\sup_{t\in(0,T]}\Bigg\{\int_0^t \frac{\psi(r^{\frac{1}{2}})\psi((t-r)^{\frac{1}{2}})}{r\sqrt{t-r}}\Big(1+\rho_r\sqrt{r}\Big)\mathrm{d}r + \bigg(\int_0^t \Big(\frac{\psi(r^{\frac{1}{2}})}{\sqrt{r}}\Big)^2\mathrm{d}r\bigg)^{\frac{1}{2}} + \bigg(\int_0^t \bigg(\frac{\psi((t-r)^{\frac{1}{2}})\psi(r^{\frac{1}{2}})}{\sqrt{r}\sqrt{t-r}}\bigg)^m\mathrm{d}r\bigg)^{\frac{1}{m}}\Bigg\} < \infty.$$

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

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

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

$$\begin{aligned}
&\mathbb{W}_\psi(P^*_{s,t}\gamma, P^*_{s,t}\tilde{\gamma}) - \frac{\psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}}\mathbb{W}_1(P^{1,\gamma}_{s,t}, P^{2,\tilde{\gamma}}_{s,t}) \le \psi((t-s)^{\frac{1}{2}})\|P^{1,\gamma}_{s,t} - P^{2,\tilde{\gamma}}_{s,t}\|_{var} \\
&\le \frac{\psi((t-s)^{\frac{1}{2}})}{\sqrt{t-s}}\bigg(\int_s^t (\mathbb{W}_\psi+\mathbb{W}_k)^2(P^*_{s,r}\gamma, P^*_{s,r}\tilde{\gamma})\mathrm{d}r\bigg)^{\frac{1}{2}} \\
&\quad + c_1\psi((t-s)^{\frac{1}{2}})\sqrt{\log(1+(t-s)^{-1})}\sup_{r\in[s+\varepsilon(t-s),t]}(\mathbb{W}_\psi+\mathbb{W}_k)(P^*_{s,r}\gamma, P^*_{s,r}\tilde{\gamma})
\end{aligned} \tag{4.1}$$

for $t\in[s,T]$. On the other hand, since $b^{(0)}$ is bounded, $\|b^{(0)}\|_{\tilde{L}^{p_0}_{q_0}(T)} < \infty$ 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

$$\begin{aligned}
&\mathbb{W}_1(P^{1,\gamma}_{s,t}, P^{2,\tilde{\gamma}}_{s,t}) \le \mathbb{W}_k(P^{1,\gamma}_{s,t}, P^{2,\tilde{\gamma}}_{s,t}) \\
&\le c_2\mathbb{W}_k(\gamma,\tilde{\gamma}) + c_2\bigg(\int_s^t (\mathbb{W}_\psi+\mathbb{W}_k)^2(P^*_{s,r}\gamma, P^*_{s,r}\tilde{\gamma})\mathrm{d}r\bigg)^{\frac{1}{2}}.
\end{aligned} \tag{4.2}$$

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

$$(4.3)\qquad \begin{aligned} &\Gamma_{s,t} \le c_3 \mathbb{W}_k(\gamma,\tilde\gamma) + c_2 h(t-s)\Gamma_{s,t}, \quad 0\le s\le t\le T,\\ &h(t) := \bigg(\int_0^t \frac{\psi(s^{\frac12})^2}{s}\text{d}s\bigg)^{\frac12} + \sup_{r\in(0,t]} \psi(r^{\frac12})\sqrt{\log(1+r^{-1})}, \quad t>0.\end{aligned}$$

Since $\int_0^1 \frac{\psi(r)^2}{r}\text{d}r < \infty$, we have $h(t)\to 0$ as $t\to 0$ if $\lim_{r\to 0}\psi(r)^2\log(1+r^{-1}) = 0$, 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 $\varepsilon\in(0,\frac12], c_1 > 0$ such that for any $\gamma,\tilde\gamma\in\mathscr{P}_k$,

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

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 $\big(\int_0^t \frac{\psi(s^{\frac12})^2}{s}\text{d}s\big)^{\frac12}$ which goes to 0 as $t\to 0$. 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,* Stochatic 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.