

# McKean-Vlasov SDEs with Drifts Discontinuous under Wasserstein Distance\*

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

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

xinghuang@tju.edu.cn

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

wangfy@tju.edu.cn

March 9, 2020

## Abstract

Existence and uniqueness are proved for McKean-Vlasov type distribution dependent SDEs with singular drifts satisfying an integrability condition in space variable and the Lipschitz condition in distribution variable with respect to  $\mathbb{W}_0$  or  $\mathbb{W}_0 + \mathbb{W}_\theta$  for some  $\theta \geq 1$ , where  $\mathbb{W}_0$  is the total variation distance and  $\mathbb{W}_\theta$  is the  $L^\theta$ -Wasserstein distance. This improves some existing results where the drift is continuous in the distribution variable with respect to the Wasserstein distance.

AMS subject Classification: 60H1075, 60G44.

Keywords: Distribution dependent SDEs, total variation distance, Wasserstein distance, Krylov's estimate, Zvonkin's transform.

## 1 Introduction

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

$$(1.1) \quad dX_t = b_t(X_t, \mathcal{L}_{X_t})dt + \sigma_t(X_t, \mathcal{L}_{X_t})dW_t, \quad t \in [0, T],$$

where  $T > 0$  is a fixed time,  $(W_t)_{t \in [0, T]}$  is the  $m$ -dimensional Brownian motion on a complete filtration probability space  $(\Omega, \{\mathcal{F}_t\}_{t \in [0, T]}, \mathbb{P})$ ,  $\mathcal{L}_{X_t}$  is the law of  $X_t$ ,

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

---

\*Supported in part by NNSFC (11771326, 11831014, 11801406).

are measurable, and  $\mathcal{P}$  is the space of all probability measures on  $\mathbb{R}^d$  equipped with the weak topology.

This type SDEs are also called McKean-Vlasov SDEs and mean field SDEs, and have been intensively investigated due to its wide applications, see for instance [1, 2, 5, 8, 10, 11, 12, 20, 22] and references within.

An adapted continuous process on  $\mathbb{R}^d$  is called a (strong) solution of (1.1), if

$$(1.2) \quad \mathbb{E} \int_0^T \{ |b_t(X_t, \mathcal{L}_{X_t})| + \|\sigma_t(X_t, \mathcal{L}_{X_t})\|^2 \} dt < \infty,$$

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

$$(1.3) \quad X_t = X_0 + \int_0^t b_s(X_s, \mathcal{L}_{X_s}) ds + \int_0^t \sigma_s(X_s, \mathcal{L}_{X_s}) dW_s, \quad t \in [0, T].$$

We call (1.1) (strongly) well-posed for an  $\mathcal{F}_0$ -measurable initial value  $X_0$ , if (1.1) has a unique solution starting at  $X_0$ .

When a different probability measure  $\tilde{\mathbb{P}}$  is concerned, we use  $\mathcal{L}_\xi | \tilde{\mathbb{P}}$  to denote the law of a random variable  $\xi$  under the probability  $\tilde{\mathbb{P}}$ , and use  $\mathbb{E}_{\tilde{\mathbb{P}}}$  to stand for the expectation under  $\tilde{\mathbb{P}}$ . For any  $\mu_0 \in \mathcal{P}$ ,  $(\tilde{X}_t, \tilde{W}_t)_{t \in [0, T]}$  is called a weak solution to (1.1) starting at  $\mu_0$ , if  $(\tilde{W}_t)_{t \in [0, T]}$  is the  $m$ -dimensional Brownian motion under a complete filtration probability space  $(\tilde{\Omega}, \{\tilde{\mathcal{F}}_t\}_{t \in [0, T]}, \tilde{\mathbb{P}})$ ,  $(\tilde{X}_t)_{t \in [0, T]}$  is a continuous  $\tilde{\mathcal{F}}_t$ -adapted process on  $\mathbb{R}^d$  with  $\mathcal{L}_{\tilde{X}_0} | \tilde{\mathbb{P}} = \mu_0$ , and (1.2)-(1.3) hold for  $(\tilde{X}, \tilde{W}, \tilde{\mathbb{P}}, \mathbb{E}_{\tilde{\mathbb{P}}})$  replacing  $(X, W, \mathbb{P}, \mathbb{E})$ . We call (1.1) weakly well-posed for an initial distribution  $\mu_0$ , if it has a unique weak solution starting at  $\mu_0$ ; i.e. it has a weak solution  $(\tilde{X}_t, \tilde{W}_t)_{t \in [0, T]}$  with initial distribution  $\mu_0$  under some complete filtration probability space  $(\tilde{\Omega}, \{\tilde{\mathcal{F}}_t\}_{t \in [0, T]}, \tilde{\mathbb{P}})$ , and  $\mathcal{L}_{\tilde{X}_{[0, T]}} | \tilde{\mathbb{P}} = \mathcal{L}_{\tilde{X}_{[0, T]}} | \tilde{\mathbb{P}}$  holds for any other weak solution with the same initial distribution  $(\tilde{X}_t, \tilde{W}_t)_{t \in [0, T]}$  under some complete filtration probability space  $(\bar{\Omega}, \{\bar{\mathcal{F}}_t\}_{t \in [0, T]}, \bar{\mathbb{P}})$ .

Recently, the (weak and strong) well-posedness is studied in [3, 4, 6, 13, 16, 17, 19] for (1.1) with  $\sigma_t(x, \gamma) = \sigma_t(x)$  independent of the distribution variable  $\gamma$ , and with singular drift  $b_t(x, \gamma)$ . See also [12, 16] for the case with memory. We briefly recall some conditions on  $b$  which together with a regular and non-degenerate condition on  $\sigma$  implies the well-posedness of (1.1). To this end, we recall the  $L^\theta$ -Wasserstein distance  $\mathbb{W}_\theta$  for  $\theta > 0$ :

$$\mathbb{W}_\theta(\gamma, \tilde{\gamma}) := \inf_{\pi \in \mathcal{C}(\gamma, \tilde{\gamma})} \left( \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^\theta \pi(dx, dy) \right)^{\frac{1}{\theta}}, \quad \gamma, \tilde{\gamma} \in \mathcal{P},$$

where  $\mathcal{C}(\gamma, \tilde{\gamma})$  is the set of all couplings of  $\gamma$  and  $\tilde{\gamma}$ . By the convention that  $r^0 = 1_{\{r > 0\}}$  for  $r \geq 0$ , we may regard  $\mathbb{W}_0$  as the total variation distance, i.e. set

$$\mathbb{W}_0(\gamma, \tilde{\gamma}) = \|\gamma - \tilde{\gamma}\|_{TV} := \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\gamma(A) - \tilde{\gamma}(A)|.$$

References [3, 4] give the well-posedness of (1.1) with a deterministic initial value  $X_0 \in \mathbb{R}^d$ , where the drift  $b_t(x, \gamma)$  is assumed to be linear growth in  $x$  uniformly in  $t, \gamma$ ,

and

$$|b_t(x, \gamma) - b_t(x, \tilde{\gamma})| \leq \phi(\mathbb{W}_1(\gamma, \tilde{\gamma}))$$

holds for some function  $\phi \in C((0, \infty); (0, \infty))$  with  $\int_0^\cdot \frac{1}{\phi(s)} ds = \infty$ . Note that for distribution dependent SDEs the well-posedness for deterministic initial values does not imply that for random ones.

[17, Theorem 3] presents the well-posedness of (1.1) with exponentially integrable  $X_0$  and a drift  $b$  of type

$$(1.4) \quad b_t(x, \gamma) := \int_{\mathbb{R}^d} \tilde{b}_t(x, y) \gamma(dy),$$

where  $\tilde{b}_t(x, y)$  has linear growth in  $x$  uniformly in  $t$  and  $y$ . Since  $\tilde{b}_t(x, y)$  is bounded in  $y$ ,  $b_t(x, \cdot)$  is Lipschitz continuous in the total variation distance  $\mathbb{W}_0$ . [19] considers the same type drift and proves the well-posedness of (1.1) under the conditions that  $\mathbb{E}|X_0|^\beta < \infty$  for some  $\beta > 0$  and

$$|\tilde{b}_t(x, y)| \leq h_t(x - y)$$

for some  $h \in L^q([0, T]; \tilde{L}^p(\mathbb{R}^d))$  for some  $p, q > 1$  with  $\frac{d}{p} + \frac{2}{q} < 1$ , where  $\tilde{L}^p$  is a localized  $L^p$  space.

In [6] the well-posedness of (1.1) is proved for  $X_0$  satisfying  $\mathbb{E}|X_0|^2 < \infty$ , and for  $b$  given by

$$(1.5) \quad b_t(x, \gamma) = \tilde{b}_t(x, \gamma(\varphi)),$$

where  $\gamma(\varphi) := \int_{\mathbb{R}^d} \varphi d\gamma$  for some  $\alpha$ -Hölder continuous function  $\varphi$ , and  $|\tilde{b}_t(x, r)| + |\partial_r \tilde{b}_t(x, r)|$  is bounded. Consequently,  $b_t(x, \gamma)$  is bounded and Lipschitz continuous in  $\gamma$  with respect to  $\mathbb{W}_\alpha$ .

In [13] the well-posedness is derived under the conditions that  $\mathbb{E}|X_0|^\theta < \infty$  for some  $\theta \geq 1$ ,  $b_t(x, \gamma)$  is Lipschitz continuous in  $\gamma$  with respect to  $\mathbb{W}_\theta$ , and for any  $\mu \in C([0, T]; \mathcal{P}_\theta)$ ,

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

satisfies  $|b^\mu|^2 \in L_p^q(T)$  for some  $(p, q) \in \mathcal{K}$ , where

$$L_p^q(T) := \left\{ f \in \mathcal{B}([0, T] \times \mathbb{R}^d) : \int_0^T \left( \int_{\mathbb{R}^d} |f_t(x)|^p dx \right)^{\frac{q}{p}} dt < \infty \right\},$$

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

Moreover, in [15] the well-posedness of (1.1) has been proved for

$$(1.6) \quad b_t(x, \mu) = \tilde{b}(\rho_\mu(x)), \quad \sigma_t(x, \mu) = \tilde{\sigma}(\rho_\mu(x))$$

with initial distribution having density function (with respect to the Lebesgue measure) in the class  $H^{2+\alpha}$  for some  $\alpha > 0$ , where  $\rho_\mu$  is the density function of  $\mu$  with respect to the

Lebesgue measure,  $\tilde{b} \in C^2([0, \infty); \mathbb{R}^d)$  and  $\tilde{\sigma} \in C^3([0, \infty); \mathbb{R}^d \otimes \mathbb{R}^d)$ . As for the weak well-posedness, [14] assumes that  $b$  is bounded and  $\mathbb{W}_0$ -Lipschitz continuous in distribution variable, and  $\sigma$  is Lipschitz continuous in space variable.

In this paper, we prove the (weak and strong) well-posedness of (1.1) for general type  $b$  with  $b_t(x, \gamma)$  Lipschitz continuous in  $\gamma$  under the metric  $\mathbb{W}_0$  or  $\mathbb{W}_0 + \mathbb{W}_\theta$  for some  $\theta \geq 1$ . This condition is weaker than those in [3, 4, 6, 13] in the sense that the drift is not necessarily continuous in the Wasserstein distance, but is incomparable with those in [17, 19] where  $b$  is of the integral type as in (1.4). Moreover, our result works for any initial value and initial distribution.

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

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

In this case, we write  $\xi = \nabla f$  and call it the weak gradient of  $f$ . For  $p, q \geq 1$ , let

$$L^q_{p,loc}(T) = \left\{ f \in \mathcal{B}([0, T] \times \mathbb{R}^d) : \int_0^T \left( \int_K |f_t(x)|^p dx \right)^{\frac{q}{p}} dt < \infty, K \subset \mathbb{R}^d \text{ is compact} \right\}.$$

We will use the following conditions.

(A $_\sigma$ )  $\sigma_t(x, \gamma) = \sigma_t(x)$  is uniformly continuous in  $x \in \mathbb{R}^d$  uniformly in  $t \in [0, T]$ ; the weak gradient  $\nabla \sigma_t$  exists for a.e.  $t \in [0, T]$  such that  $|\nabla \sigma|^2 \in L^q_p(T)$  for some  $(p, q) \in \mathcal{K}$ ; and there exists a constant  $K_1 \geq 1$  such that

$$(1.7) \quad K_1^{-1} I \leq (\sigma_t \sigma_t^*)(x) \leq K_1 I, \quad (t, x) \in [0, T] \times \mathbb{R}^d,$$

where  $I$  is the  $d \times d$  identity matrix.

(A $_b$ )  $b = \bar{b} + \hat{b}$ , where  $\bar{b}$  and  $\hat{b}$  satisfy

$$(1.8) \quad \begin{aligned} & |\hat{b}_t(x, \gamma) - \hat{b}_t(y, \tilde{\gamma})| + |\bar{b}_t(x, \gamma) - \bar{b}_t(x, \tilde{\gamma})| \\ & \leq K_2 (\|\gamma - \tilde{\gamma}\|_{TV} + \mathbb{W}_\theta(\gamma, \tilde{\gamma}) + |x - y|), \quad t \in [0, T], x, y \in \mathbb{R}^d, \gamma, \tilde{\gamma} \in \mathcal{P}_\theta \end{aligned}$$

for some constants  $\theta, K_2 \geq 1$ , and there exists  $(p, q) \in \mathcal{K}$  such that

$$(1.9) \quad \sup_{t \in [0, T]} |\hat{b}_t(0, \delta_0)| + \sup_{\mu \in C([0, T]; \mathcal{P}_\theta)} \|\bar{b}^\mu\|^2_{L^q_p(T)} < \infty,$$

where  $\bar{b}_t^\mu(x) := \bar{b}_t(x, \mu_t)$  for  $(t, x) \in [0, T] \times \mathbb{R}^d$ , and  $\delta_0$  stands for the Dirac measure at the point  $0 \in \mathbb{R}^d$ .

(A'\_b) For any  $\mu \in \mathcal{B}([0, T]; \mathcal{P})$ ,  $|b^\mu|^2 \in L^q_{p,loc}(T)$  for some  $(p, q) \in \mathcal{K}$ . Moreover, there exists a function  $\Gamma : [0, \infty) \rightarrow [0, \infty)$  satisfying  $\int_1^\infty \frac{1}{\Gamma(x)} = \infty$  such that

$$(1.10) \quad \langle b_t(x, \delta_0), x \rangle \leq \Gamma(|x|^2), \quad t \in [0, T], x \in \mathbb{R}^d.$$

In addition, there exists a constant  $K_3 \geq 1$  such that

$$(1.11) \quad |b_t(x, \gamma) - b_t(x, \tilde{\gamma})| \leq K_3 \|\gamma - \tilde{\gamma}\|_{TV}, \quad t \in [0, T], x \in \mathbb{R}^d, \gamma, \tilde{\gamma} \in \mathcal{P}.$$

When (1.1) is weakly well-posed for initial distribution  $\gamma$ , we denote  $P_t^* \gamma$  the distribution of the weak solution at time  $t$ .

**Theorem 1.1.** *Assume  $(A_\sigma)$ .*

- (1) *If  $(A'_b)$  holds, then (1.1) is strongly and weakly well-posed for any initial values and any initial distribution. Moreover,*

$$(1.12) \quad \|P_t^* \mu_0 - P_t^* \nu_0\|_{TV}^2 \leq 2e^{\frac{\kappa_1 \kappa_3^2 t}{2}} \|\mu_0 - \nu_0\|_{TV}^2, \quad t \in [0, T], \mu_0, \nu_0 \in \mathcal{P}.$$

- (2) *Let  $\mathbb{E}|X_0|^\theta < \infty$  and  $\mu_0(|\cdot|^\theta) < \infty$ . If  $(A_b)$  holds, then (1.1) is strongly well-posed for initial value  $X_0$  and weakly well-posed for initial distribution  $\mu_0$ . Moreover, there exists a constant  $c > 0$  such that for any  $\mu_0, \nu_0 \in \mathcal{P}_\theta$ ,*

$$(1.13) \quad \begin{aligned} & \|P_t^* \mu_0 - P_t^* \nu_0\|_{TV} + \mathbb{W}_\theta(P_t^* \mu_0, P_t^* \nu_0) \\ & \leq c \{ \|\mu_0 - \nu_0\|_{TV} + W_\theta(\mu_0, \nu_0) \}, \quad t \in [0, T]. \end{aligned}$$

To illustrate this result comparing with earlier ones, we present an example of  $b$  which satisfies our conditions but is not of type (1.4)-(1.6) and is discontinuous in both the space variable and the distribution variable under the weak topology. If one wants to control a stochastic system in terms of an ideal reference distribution  $\mu_0$ , it is natural to take a drift depending on a probability distance between  $\mu_0$  and the law of the system. As two typical probability distances, the total variation and Wasserstein distances have been widely applied in applications. So, we take for instance

$$b_t(x, \mu) = \bar{b}(t, x, \mu) + h(t, x, \mathbb{W}_\theta(\mu, \mu_0), \|\mu - \mu_0\|_{TV})$$

for some  $\theta \geq 1$ , where  $\bar{b}$  satisfies (1.8) and (1.9) for  $\hat{b} = 0$  which refers to the singularity in the space variable  $x$ , and  $h : [0, T] \times \mathbb{R}^d \times [0, \infty)^2 \rightarrow \mathbb{R}^d$  is measurable such that  $h(t, x, r, s)$  is bounded in  $t \in [0, T]$  and Lipschitz continuous in  $(x, r, s) \in \mathbb{R}^d \times [0, \infty)^2$  uniformly in  $t \in [0, T]$ . Obviously,  $b(t, x, \mu)$  satisfies condition  $(A_b)$  but is not of type (1.4)-(1.6) and can be discontinuous in  $x$  and  $\mu$  under the weak topology.

In the next section we make some preparations, which will be used in Section 3 for the proof of Theorem 1.1.

## 2 Preparations

We first present the following version of Yamada-Watanabe principle modified from [13, Lemma 3.4].

**Lemma 2.1.** *Assume that (1.1) has a weak solution  $(\bar{X}_t)_{t \in [0, T]}$  under probability  $\bar{\mathbb{P}}$ , and let  $\mu_t = \mathcal{L}_{\bar{X}_t} | \bar{\mathbb{P}}, t \in [0, T]$ . If the SDE*

$$(2.1) \quad dX_t = b_t(X_t, \mu_t)dt + \sigma_t(X_t, \mu_t)dW_t$$

*has strong uniqueness for some initial value  $X_0$  with  $\mathcal{L}_{X_0} = \mu_0$ , then (1.1) has a strong solution starting at  $X_0$ . If moreover (1.1) has strong uniqueness for any initial value  $X_0$  with  $\mathcal{L}_{X_0} = \mu_0$ , then it is weakly well-posed for the initial distribution  $\mu_0$ .*

*Proof.* (a) Strong existence. Since  $\mu_t = \mathcal{L}_{\bar{X}_t} | \bar{\mathbb{P}}, \bar{X}_t$  under  $\bar{\mathbb{P}}$  is also a weak solution of (2.1) with initial distribution  $\mu_0$ . By the Yamada-Watanabe principle, the strong uniqueness of (2.1) with initial value  $X_0$  implies the strong (resp. weak) well-posedness of (2.1) starting at  $X_0$  (resp.  $\mu_0$ ). In particular, the weak uniqueness implies  $\mathcal{L}_{X_t} = \mu_t, t \in [0, T]$ , so that  $X_t$  solves (1.1).

(b) Weak uniqueness. Let  $\tilde{X}_t$  under probability  $\tilde{\mathbb{P}}$  be another weak solution of (1.1) with initial distribution  $\mu_0$ . For any initial value  $X_0$  with  $\mathcal{L}_{X_0} = \mu_0$ , the strong uniqueness of (2.1) starting at  $X_0$  implies

$$X_{[0, T]} = F(X_0, W_{[0, T]})$$

for some measurable function  $F : \mathbb{R}^d \times C([0, T]; \mathbb{R}^d) \rightarrow C([0, T]; \mathbb{R}^d)$ . This and the weak uniqueness of (2.1) proved in (a) yield

$$(2.2) \quad \mathcal{L}_{\tilde{X}_{[0, T]}} | \tilde{\mathbb{P}} = \mathcal{L}_{X_{[0, T]}} | \mathbb{P}.$$

Let  $\hat{X}_{[0, T]} = F(\tilde{X}_0, \tilde{W}_{[0, T]})$ . We have  $\hat{X}_0 = \tilde{X}_0$  and

$$\mathcal{L}_{\hat{X}_{[0, T]}} | \tilde{\mathbb{P}} = \mathcal{L}_{X_{[0, T]}} | \mathbb{P}.$$

This and (2.2) imply  $\mathcal{L}_{\hat{X}_t} | \tilde{\mathbb{P}} = \mu_t$ , so that  $\hat{X}_t$  under  $\tilde{\mathbb{P}}$  is a weak solution of (1.1) with  $\hat{X}_0 = \tilde{X}_0$ . By the strong uniqueness of (1.1), we derive  $\hat{X}_{[0, T]} = \tilde{X}_{[0, T]}$ . Combining this with (2.2) we obtain

$$\mathcal{L}_{\tilde{X}_{[0, T]}} | \tilde{\mathbb{P}} = \mathcal{L}_{\hat{X}_{[0, T]}} | \tilde{\mathbb{P}} = \mathcal{L}_{X_{[0, T]}} | \mathbb{P} = \mathcal{L}_{\tilde{X}_{[0, T]}} | \tilde{\mathbb{P}},$$

i.e. (1.1) has weak uniqueness starting at  $\mu_0$ . □

We will use the following result for the maximal operator:

$$(2.3) \quad \mathcal{M}h(x) := \sup_{r>0} \frac{1}{|B(x, r)|} \int_{B(x, r)} h(y)dy, \quad h \in L^1_{loc}(\mathbb{R}^d), x \in \mathbb{R}^d,$$

where  $B(x, r) := \{y : |x - y| < r\}$ , see [7, Appendix A].

**Lemma 2.2.** *There exists a constant  $C > 0$  such that for any continuous and weak differentiable function  $f$ ,*

$$(2.4) \quad |f(x) - f(y)| \leq C|x - y|(\mathcal{M}|\nabla f|(x) + \mathcal{M}|\nabla f|(y)), \quad \text{a.e. } x, y \in \mathbb{R}^d.$$

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

$$(2.5) \quad \|\mathcal{M}f\|_{L^p} \leq C_p\|f\|_{L^p}, \quad f \in L^p(\mathbb{R}^d).$$

To compare the distribution dependent SDE (1.1) with a classical one, for any  $\mu \in \mathcal{B}([0, T]; \mathcal{P})$ , let  $b_t^\mu(x) := b_t(x, \mu_t)$  and consider the classical SDE

$$(2.6) \quad dX_t^\mu = b_t^\mu(X_t^\mu)dt + \sigma_t(X_t^\mu)dW_t, \quad t \in [0, T].$$

According to [25], assumption  $(A_\sigma)$  together with  $(A_b)$  or  $(A'_b)$  implies the strong well-posedness, where under  $(A'_b)$  the non-explosion is implied by (1.10). For any  $\gamma \in \mathcal{P}$ , Let  $\Phi_t^\gamma(\mu) = \mathcal{L}_{X_t^\mu}$  for  $(X_t^\mu)_{t \in [0, T]}$  solving (2.6) with  $\mathcal{L}_{X_0^\mu} = \gamma$ . We have the following result.

**Lemma 2.3.** *Assume  $(A_\sigma)$  and let  $\gamma \in \mathcal{P}$ .*

(1) *If  $(A'_b)$  holds, then for any  $\mu, \nu \in \mathcal{B}([0, T]; \mathcal{P})$ ,*

$$(2.7) \quad \|\Phi_t^\gamma(\mu) - \Phi_t^\gamma(\nu)\|_{TV}^2 \leq \frac{K_1 K_3^2}{4} \int_0^t \|\mu_s - \nu_s\|_{TV}^2 ds, \quad t \in [0, T].$$

(2) *If  $(A_b)$  holds and  $\gamma \in \mathcal{P}_\theta$ , then for any  $\mu \in C([0, T]; \mathcal{P}_\theta)$ , we have  $\Phi_t^\gamma(\mu) \in C([0, T]; \mathcal{P}_\theta)$ . Moreover, for any  $m \geq 1 \vee \frac{\theta}{2}$ , there exists a constant  $C > 0$  such that for any  $\mu, \nu \in C([0, T]; \mathcal{P}_\theta)$  and  $\gamma_1, \gamma_2 \in \mathcal{P}_\theta$ ,*

$$(2.8) \quad \begin{aligned} & \{\mathbb{W}_\theta(\Phi_t^{\gamma_1}(\mu), \Phi_t^{\gamma_2}(\nu))\}^{2m} \\ & \leq C\mathbb{W}_\theta(\gamma_1, \gamma_2)^{2m} + C \int_0^t \{\|\mu_s - \nu_s\|_{TV} + \mathbb{W}_\theta(\mu_s, \nu_s)\}^{2m} ds, \quad t \in [0, T]. \end{aligned}$$

*Proof.* (1) Let  $(A'_b)$  hold and take  $\mu, \nu \in \mathcal{B}([0, T]; \mathcal{P})$ . To compare  $\Phi_t^\gamma(\mu)$  with  $\Phi_t^\gamma(\nu)$ , we rewrite (2.6) as

$$(2.9) \quad dX_t^\mu = b_t(X_t^\mu, \nu_t)dt + \sigma_t(X_t^\mu)d\tilde{W}_t,$$

where

$$\tilde{W}_t = W_t + \int_0^t \xi_s ds, \quad \xi_s := \{\sigma_s^*(\sigma_s \sigma_s^*)^{-1}\}(X_s^\mu)[b_s(X_s^\mu, \mu_s) - b_s(X_s^\mu, \nu_s)], \quad s, t \in [0, T].$$

Noting that (1.7) together with (1.11) implies

$$(2.10) \quad \mathbb{E}[e^{\frac{1}{2} \int_0^T |\xi_s|^2 ds}] < \infty,$$

by the Girsanov theorem we see that  $R_T := e^{-\int_0^T \langle \xi_s, dW_s \rangle - \frac{1}{2} \int_0^T |\xi_s|^2 ds}$  is a probability density with respect to  $\mathbb{P}$ , and  $(\tilde{W}_t)_{t \in [0, T]}$  is a  $d$ -dimensional Brownian motion under the probability  $\mathbb{Q} := R_T \mathbb{P}$ .

By the weak uniqueness of (2.6) and  $\mathcal{L}_{X_0^\mu} | \mathbb{Q} = \mathcal{L}_{X_0^\mu} = \gamma$ , we conclude from (2.9) with  $\mathbb{Q}$ -Brownian motion  $\tilde{W}_t$  that

$$\Phi_t^\gamma(\nu) = \mathcal{L}_{X_t^\mu} | \mathbb{Q}, \quad t \in [0, T].$$

Combining this with  $(A_\sigma)$  and applying Pinker's inequality [18], we obtain

$$\begin{aligned} 2 \|\Phi_t^\gamma(\nu) - \Phi_t^\gamma(\mu)\|_{TV}^2 &\leq 2 \sup_{\|f\|_\infty \leq 1} (\mathbb{E}|f(X_t^\mu)(R_t - 1)|)^2 = 2(\mathbb{E}|R_t - 1|)^2 \\ (2.11) \quad &\leq \mathbb{E}[R_t \log R_t] = \frac{1}{2} \mathbb{E}_{\mathbb{Q}} \int_0^t |\xi_s|^2 ds \\ &\leq \frac{K_1}{2} \mathbb{E}_{\mathbb{Q}} \int_0^t |b_s(X_s^\mu, \mu_s) - b_s(X_s^\mu, \nu_s)|^2 ds. \end{aligned}$$

By  $(A'_b)$ , this implies (2.7).

(2) Let  $(A_b)$  hold and take  $m \geq 1 \vee \frac{\theta}{2}$ . Take  $\mathcal{F}_0$ -measurable random variables  $X_0^\mu$  and  $X_0^\nu$  such that  $\mathcal{L}_{X_0^\mu} = \gamma_1, \mathcal{L}_{X_0^\nu} = \gamma_2$  and

$$\mathbb{E}|X_0^\mu - X_0^\nu|^\theta = \{\mathbb{W}_\theta(\gamma_1, \gamma_2)\}^\theta.$$

Let  $X_t^\mu$  solve (2.6) and  $X_t^\nu$  solve the same SDE for  $\nu$  replacing  $\mu$ . We need to find a constant  $C > 0$  such that for any  $t \in [0, T]$ ,

$$\begin{aligned} (2.12) \quad &\{\mathbb{W}_\theta(\Phi_t^{\gamma_1}(\mu), \Phi_t^{\gamma_2}(\nu))\}^{2m} \\ &\leq C(\mathbb{E}|X_0^\mu - X_0^\nu|^\theta)^{\frac{2m}{\theta}} + C \int_0^t (\mathbb{W}_\theta(\mu_s, \nu_s) + \|\mu_s - \nu_s\|_{TV})^{2m} ds, \quad t \in [0, T]. \end{aligned}$$

To this end, we make a Zvokin type transform as in [13] and [24].

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

$$(2.13) \quad \frac{\partial u_t}{\partial t} + \frac{1}{2} \text{Tr}(\sigma_t \sigma_t^* \nabla^2 u_t) + \nabla_{b_t^\mu} u_t + \bar{b}_t^\mu = \lambda u_t, \quad u_T = 0.$$

According to [24, Remark 2.1, Proposition 2.3 (2)], under assumptions  $(A_\sigma)$  and  $(A_b)$ , when  $\lambda$  is large enough (2.13) has a unique solution  $\mathbf{u}^{\lambda, \mu}$  satisfying

$$(2.14) \quad \|\mathbf{u}^{\lambda, \mu}\|_\infty + \|\nabla \mathbf{u}^{\lambda, \mu}\|_\infty \leq \frac{1}{5},$$

and

$$(2.15) \quad \|\nabla^2 \mathbf{u}^{\lambda, \mu}\|_{L_{2p}^{2q}(T)} < \infty.$$

Let  $\Theta_t^{\lambda,\mu}(x) = x + \mathbf{u}_t^{\lambda,\mu}(x)$ . It is easy to see that (2.13) and the Itô formula imply

$$(2.16) \quad d\Theta_t^{\lambda,\mu}(X_t^\mu) = (\lambda \mathbf{u}_t^{\lambda,\mu} + \hat{b}_t^\mu)(X_t^\mu)dt + (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\mu) dW_t.$$

In particular, (2.14) and  $\mathbb{E}[|X_0^\mu|^\theta] < \infty$  imply that  $\mathbb{E}[|\Theta_0^{\lambda,\mu}(X_0^\mu)|^\theta] < \infty$  and (2.16) is an SDE for  $\xi_t := \Theta_t^{\lambda,\mu}(X_t^\mu)$  with coefficients of at most linear growth, so that  $\mathcal{L}_\xi \in C([0, T]; \mathcal{P}_\theta)$  and so does  $\mathcal{L}_{X_t^\mu}$  due to (2.14).

It remains to prove (2.8). To this end, we observe that (2.13) and the Itô formula yield

$$\begin{aligned} d\Theta_t^{\lambda,\mu}(X_t^\nu) &= \lambda \mathbf{u}_t^{\lambda,\mu}(X_t^\nu)dt + (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\nu) dW_t \\ &\quad + [\{\nabla \mathbf{u}_t^{\lambda,\mu}\}(b_t^\nu - b_t^\mu) + b_t^\nu - \bar{b}_t^\mu](X_t^\nu)dt \\ &= [\lambda \mathbf{u}_t^{\lambda,\mu} + \{\nabla \Theta_t^{\lambda,\mu}\}(b_t^\nu - b_t^\mu) + \hat{b}_t^\mu](X_t^\nu)dt + (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\nu) dW_t. \end{aligned}$$

Combining this with (2.16) and applying the Itô formula, we see that  $\eta_t := \Theta_t^{\lambda,\mu}(X_t^\mu) - \Theta_t^{\lambda,\mu}(X_t^\nu)$  satisfies

$$\begin{aligned} d|\eta_t|^2 &= 2 \left\langle \eta_t, \lambda \mathbf{u}_t^{\lambda,\mu}(X_t^\mu) - \lambda \mathbf{u}_t^{\lambda,\mu}(X_t^\nu) + \hat{b}_t^\mu(X_t^\mu) - \hat{b}_t^\mu(X_t^\nu) \right\rangle dt \\ &\quad + 2 \left\langle \eta_t, [(\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\mu) - (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\nu)] dW_t \right\rangle \\ &\quad + \left\| (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\mu) - (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\nu) \right\|_{HS}^2 dt \\ &\quad - 2 \left\langle \eta_t, [\{\nabla \Theta_t^{\lambda,\mu}\}(b_t^\nu - b_t^\mu)](X_t^\nu) \right\rangle dt. \end{aligned}$$

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

$$\begin{aligned} d|\eta_t|^{2m} &= 2m|\eta_t|^{2(m-1)} \left\langle \eta_t, \lambda \mathbf{u}_t^{\lambda,\mu}(X_t^\mu) - \lambda \mathbf{u}_t^{\lambda,\mu}(X_t^\nu) + \hat{b}_t^\mu(X_t^\mu) - \hat{b}_t^\mu(X_t^\nu) \right\rangle dt \\ &\quad + 2m|\eta_t|^{2(m-1)} \left\langle \eta_t, [(\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\mu) - (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\nu)] dW_t \right\rangle \\ (2.17) \quad &\quad + m|\eta_t|^{2(m-1)} \left\| (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\mu) - (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\nu) \right\|_{HS}^2 dt \\ &\quad + 2m(m-1)|\eta_t|^{2(m-2)} \left| [(\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\mu) - (\{\nabla \Theta_t^{\lambda,\mu}\} \sigma_t)(X_t^\nu)]^* \eta_t \right|^2 dt \\ &\quad - 2m|\eta_t|^{2(m-1)} \left\langle \eta_t, [\{\nabla \Theta_t^{\lambda,\mu}\}(b_t^\nu - b_t^\mu)](X_t^\nu) \right\rangle dt. \end{aligned}$$

By (2.14) and (1.8), we may find a constant  $c_0 > 0$  such that

$$(2.18) \quad |\eta_t|^{2(m-1)} |\eta_t| \cdot |\lambda \mathbf{u}_t^{\lambda,\mu}(X_t^\mu) - \lambda \mathbf{u}_t^{\lambda,\mu}(X_t^\nu) + \hat{b}_t^\mu(X_t^\mu) - \hat{b}_t^\mu(X_t^\nu)| \leq c_0 |\eta_t|^{2m},$$

and

$$\begin{aligned} &|\eta_t|^{2(m-1)} |\eta_t| \cdot |[\{\nabla \Theta_t^{\lambda,\mu}\}(b_t^\nu - b_t^\mu)](X_t^\nu)| \\ (2.19) \quad &\leq K_2 \|\nabla \Theta^{\lambda,\mu}\|_\infty |\eta_t|^{2(m-1)} |\eta_t| (\mathbb{W}_\theta(\mu_t, \nu_t) + \|\mu_t - \nu_t\|_{TV}) \\ &\leq c_0 (|\eta_t|^{2m} + \mathbb{W}_\theta(\mu_t, \nu_t)^{2m} + \|\mu_t - \nu_t\|_{TV}^{2m}), \end{aligned}$$

According to [13, (4.19)-(4.20)], we arrive at

$$(2.20) \quad d|\eta_t|^{2m} \leq c_1|\eta_t|^{2m}dA_t + c_1(\mathbb{W}_\theta(\mu_t, \nu_t)^{2m} + \|\mu_t - \nu_t\|_{TV}^{2m})dt + dM_t$$

for some constant  $c_1 > 0$ , a local martingale  $M_t$ , and

$$A_t := \int_0^t \left\{ 1 + (\mathcal{M}(\|\nabla^2\Theta_s^{\lambda,\mu}\| + \|\nabla\sigma_s\|)(X_s^\mu) + \mathcal{M}(\|\nabla^2\Theta_s^{\lambda,\mu}\| + \|\nabla\sigma_s\|)(X_s^\nu))^2 \right\} ds.$$

Thanks to [24, Theorem 3.1], the Krylov estimate

$$(2.21) \quad \begin{aligned} & \mathbb{E} \left[ \int_s^t |f_r|(X_r^\mu) dr \middle| \mathcal{F}_s \right] + \mathbb{E} \left[ \int_s^t |f_r|(X_r^\nu) dr \middle| \mathcal{F}_s \right] \\ & \leq C \left( \int_s^t \left( \int_{\mathbb{R}^d} |f_r(x)|^p dx \right)^{\frac{q}{p}} dr \right)^{\frac{1}{q}}, \quad 0 \leq s < t \leq T. \end{aligned}$$

holds. As shown in [23, Lemma 3.5], (2.21), (2.5), (2.15) and  $(A_\sigma)$  imply

$$\sup_{t \in [0, T]} \mathbb{E} e^{\delta A_t} = \mathbb{E} e^{\delta A_T} < \infty, \quad \delta > 0.$$

By (2.14) and the stochastic Gronwall lemma (see [23, Lemma 3.8]), (2.20) with  $2m > \theta$  implies

$$\begin{aligned} & \{\mathbb{W}_\theta(\Phi_t^{\gamma_1}(\mu), \Phi_t^{\gamma_2}(\nu))\}^{2m} \leq c_2(\mathbb{E}|\eta_t|^\theta)^{\frac{2m}{\theta}} \\ & \leq c_3(\mathbb{E}|X_0^\mu - X_0^\nu|^\theta)^{\frac{2m}{\theta}} + c_3(\mathbb{E}e^{\frac{c_1\theta}{2m-\theta}A_T})^{\frac{2m-\theta}{\theta}} \int_0^t (\mathbb{W}_\theta(\mu_s, \nu_s)^{2m} + \|\mu_s - \nu_s\|_{TV}^{2m}) ds \end{aligned}$$

holds for all  $t \in [0, T]$  and some constants  $c_2, c_3 > 0$ . Therefore, (2.12) holds for some constant  $C > 0$  and the proof is thus finished.  $\square$

### 3 Proof of Theorem 1.1

Assume  $(A_\sigma)$ . According to [25, Theorem 1.3], for any  $\mu \in \mathcal{B}([0, T]; \mathcal{P})$ , each of  $(A_b)$  and  $(A'_b)$  implies the strong existence and uniqueness up to life time of the SDE (2.1). Moreover, it is standard that in both cases a solution of (2.1) is non-explosive. So, by Lemma 2.1, the strong well-posedness of (1.1) implies the weak well-posedness. Therefore, in the following we need only consider the strong solution.

To prove the strong well-posedness of (1.1), it suffices to find a constant  $t_0 \in (0, T]$  independent of  $X_0$  such that in each of these two cases the SDE (1.1) has strong well-posedness up to time  $t_0$ . Indeed, once this is confirmed, by considering the SDE from time  $t_0$  we prove the same property up to time  $(2t_0) \wedge T$ . Repeating the procedure finite many times we derive the strong well-posedness.

Below we prove assertions (1) and (2) for strong solutions respectively.

(a) Let  $(A'_b)$  hold. Take  $t_0 = \min\{T, \frac{1}{K_1 K_3^2}\}$  and consider the space  $E_{t_0} := \{\mu \in \mathcal{B}([0, t_0]; \mathcal{P}) : \mu_0 = \gamma\}$  equipped with the complete metric

$$\rho(\nu, \mu) := \sup_{t \in [0, t_0]} \|\nu_t - \mu_t\|_{TV}.$$

Then (2.7) implies that  $\Phi^\gamma$  is a strictly contractive map on  $E_{t_0}$ , so that it has a unique fixed point, i.e. the equation

$$(3.1) \quad \Phi_t^\gamma(\mu) = \mu_t, \quad t \in [0, t_0]$$

has a unique solution  $\mu \in E_{t_0}$ . By (3.1) and the definition of  $\Phi^\gamma$  we see that the unique solution of (2.1) is a strong solution of (1.1). On the other hand,  $\mu_t := \mathcal{L}_{X_t}$  for any strong solution to (1.1) is a solution to (3.1), hence the uniqueness of (3.1) implies that of (1.1).

To prove (1.12), let  $\mu_t = P_t^* \mu_0$  and  $\nu_t = P_t^* \nu_0$ . We have  $P_t^* \mu_0 = \Phi_t^{\mu_0}(\mu)$  and  $P_t^* \nu_0 = \Phi_t^{\nu_0}(\nu)$ . So, (2.7) with  $\gamma = \mu_0$  implies

$$(3.2) \quad \|P_t^* \mu_0 - \Phi_t^{\mu_0}(\nu)\|_{TV}^2 \leq \frac{K_1 K_3^2}{4} \int_0^t \|P_s^* \mu_0 - P_s^* \nu_0\|_{TV}^2 ds, \quad t \in [0, T].$$

On the other hand, by the Markov property for the solution to (2.6) with  $\nu$  replacing  $\mu$ , we have

$$\Phi_t^\gamma(\nu) = \int_{\mathbb{R}^d} \Phi_t^{\delta_x}(\nu) \gamma(dx), \quad \gamma \in \mathcal{P}.$$

Combining this with  $P_t^* \nu_0 = \Phi_t^{\nu_0}(\nu)$ , we obtain

$$\begin{aligned} |\{\Phi_t^{\mu_0}(\nu)\}(A) - \{P_t^* \nu_0\}(A)| &= \left| \int_{\mathbb{R}^d} \{\Phi_t^{\delta_x}(\nu)\}(A) (\mu_0 - \nu_0)(dx) \right| \\ &\leq \|\mu_0 - \nu_0\|_{TV}, \quad A \in \mathcal{B}(\mathbb{R}^d). \end{aligned}$$

Hence,

$$(3.3) \quad \|\Phi_t^{\mu_0}(\nu) - P_t^* \nu_0\|_{TV} \leq \|\mu_0 - \nu_0\|_{TV}, \quad t \in [0, T].$$

This together with (3.2) yields

$$\begin{aligned} \|P_t^* \mu_0 - P_t^* \nu_0\|_{TV}^2 &\leq 2\|P_t^* \mu_0 - \Phi_t^{\mu_0}(\nu)\|_{TV}^2 + 2\|\Phi_t^{\mu_0}(\nu) - P_t^* \nu_0\|_{TV}^2 \\ &\leq 2\|\mu_0 - \nu_0\|_{TV}^2 + \frac{K_1 K_3^2}{2} \int_0^t \|P_s^* \mu_0 - P_s^* \nu_0\|_{TV}^2 ds, \quad t \in [0, T]. \end{aligned}$$

By Gronwall's lemma, this implies (1.12).

(b) Let  $(A_b)$  hold and let  $\gamma = \mathcal{L}_{X_0} \in \mathcal{P}_\theta$ . For any  $\mu, \nu \in C([0, T], \mathcal{P}_\theta)$ , (1.8) implies (2.11). By (2.11), (1.8) and (2.8) with  $\gamma_1 = \gamma_2 = \gamma$ , we find a constant  $C > 0$  such that

$$(3.4) \quad \begin{aligned} &\{\|\Phi_t^\gamma(\mu) - \Phi_t^\gamma(\nu)\|_{TV} + \mathbb{W}_\theta(\Phi_t^\gamma(\mu), \Phi_t^\gamma(\nu))\}^{2m} \\ &\leq C \int_0^t \{\|\mu_s - \nu_s\|_{TV} + \mathbb{W}_\theta(\mu_s, \nu_s)\}^{2m} ds, \quad t \in [0, T], \gamma \in \mathcal{P}_\theta. \end{aligned}$$

Let  $t_0 = \frac{1}{2C}$ . We consider the space  $\tilde{E}_{t_0} := \{\mu \in C([0, t_0]; \mathcal{P}_\theta) : \mu_0 = \gamma\}$  equipped with the complete metric

$$\tilde{\rho}(\nu, \mu) := \sup_{t \in [0, t_0]} \{\|\nu_t - \mu_t\|_{TV} + \mathbb{W}_\theta(\nu_t, \mu_t)\}.$$

Then  $\Phi^\gamma$  is strictly contractive in  $\tilde{E}_{t_0}$ , so that the same argument in (a) proves the strong well-posedness of (1.1) with  $\mathcal{L}_{X_0} = \gamma$  up to time  $t_0$ .

Let  $\mu_t$  and  $\nu_t$  be in (a). By (3.4) with  $\gamma = \mu_0$  we obtain

$$(3.5) \quad \begin{aligned} & \left\{ \|P_t^* \mu_0 - \Phi_t^{\mu_0}(\nu)\|_{TV} + \mathbb{W}_\theta(P_t^* \mu_0, \Phi_t^{\mu_0}(\nu)) \right\}^{2m} \\ & \leq C \int_0^t \left\{ \|P_s^* \mu_0 - P_s^* \nu_0\|_{TV} + \mathbb{W}_\theta(P_s^* \mu_0, P_s^* \nu_0) \right\}^{2m} ds, \quad t \in [0, T]. \end{aligned}$$

Next, taking  $\gamma_1 = \nu_0, \gamma_2 = \mu_0$  and  $\mu = \nu$  in (2.8), we derive

$$\left\{ \mathbb{W}_\theta(P_t^* \nu_0, \Phi_t^{\mu_0}(\nu)) \right\}^{2m} \leq C \left\{ \mathbb{W}_\theta(\mu_0, \nu_0) \right\}^{2m}.$$

Combining this with (3.3) and (3.5), we find a constant  $C' > 0$  such that

$$\begin{aligned} & \left\{ \|P_t^* \mu_0 - P_t^* \nu_0\|_{TV} + \mathbb{W}_\theta(P_t^* \mu_0, P_t^* \nu_0) \right\}^{2m} \\ & \leq C' \left\{ \|\mu_0 - \nu_0\|_{TV} + \mathbb{W}_\theta(\mu_0, \nu_0) \right\}^{2m} \\ & + C' \int_0^t \left\{ \|P_s^* \mu_0 - P_s^* \nu_0\|_{TV} + \mathbb{W}_\theta(P_s^* \mu_0, P_s^* \nu_0) \right\}^{2m} ds, \quad t \in [0, T]. \end{aligned}$$

By Gronwall's lemma, this implies (1.13) for some constant  $c > 0$ .

## References

- [1] V. Barbu, M. Röckner, *Probabilistic representation for solutions to non-linear Fokker-Planck equations*, SIAM J. Math. Anal. 50(2018), 4246-4260.
- [2] V. Barbu, M. Röckner, *From non-linear Fokker-Planck equations to solutions of distribution dependent SDE*, arXiv:1808.10706.
- [3] M. Bauer, T. M-Brandis, *Existence and regularity of solutions to multi-dimensional mean-field stochastic differential equations with irregular drift*, arXiv:1912.05932.
- [4] M. Bauer, T. M-Brandis, F. Proske, *Strong solutions of mean-field stochastic differential equations with irregular drift*, Electron. J. Probab. 23(2018), 35 pp.
- [5] K. Carrapatoso, *Exponential convergence to equilibrium for the homogeneous Landau equation with hard potentials*, Bull. Sci. Math. 139(2015), 777-805.

- [6] P. E. Chaudru de Raynal, *Strong well-posedness of McKean-Vlasov stochastic differential equation with Hölder drift*, DOI: 10.1016/j.spa.2019.01.006.
- [7] G. Crippa, C. De Lellis, *Estimates and regularity results for the DiPerna-Lions flow*, J. Reine Angew. Math. 616(2008), 15-46.
- [8] L. Campi, M. Fischer, *N-player games and mean-field games with absorption*, Ann. Appl. Probab. 28(2016), 2188-2242.
- [9] I. Csiszár, J. Körne, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.
- [10] L. Desvillettes, C. Villani, *On the spatially homogeneous Landau equation for hard potentials, Part I: existence, uniqueness and smoothness*, Comm. Part. Diff. Equat. 25(2000), 179-259.
- [11] L. Desvillettes, C. Villani, *On the spatially homogeneous Landau equation for hard potentials, Part II: H-Theorem and Applications*, Comm. Part. Diff. Equat. 25(2000), 261-298.
- [12] X. Huang, M. Röckner, F.-Y. Wang, *Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs*, Discrete Contin. Dyn. Syst. 39(2019), 3017-3035.
- [13] X. Huang, F.-Y. Wang, *Distribution dependent SDEs with singular coefficients*, Stoch. Process Appl. 129(2019), 4747-4770.
- [14] B. Jourdain, *Diffusions with a nonlinear irregular drift coefficient and probabilistic interpretation of generalized Burgers' equations*, ESAIM Probab. Statist. 1(1997), 339-355.
- [15] B. Jourdain, S. Méléard, *Propagation of chaos and fluctuations for a moderate model with smooth initial data*, Ann. Inst. H. Poincaré Probab. Statist. 34(1998), 727-766.
- [16] D. Lacker, *On a strong form of propagation of chaos for McKean-Vlasov equations*, Electron. Commun. Probab. 23(2018), 11 pp.
- [17] Yu. S. Mishura, A. Yu. Veretennikov, *Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations*, arXiv:1603.02212.
- [18] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco, 1964.
- [19] M. Röckner, X. Zhang, *Well-posedness of distribution dependent SDEs with singular drifts*, arXiv:1809.02216.

- [20] A.-S. Sznitman, *Topics in propagation of chaos*, In “École d’Été de Probabilités de Sain-Flour XIX-1989”, Lecture Notes in Mathematics 1464, p. 165-251, Springer, Berlin, 1991.
- [21] F.-Y. Wang, *Gradient estimate and applications for SDEs in Hilbert space with multiplicative noise and Dini continuous drift*, J. Differential Equations, 260(2016), 2792-2829.
- [22] F.-Y. Wang, *Distribution-dependent SDEs for Landau type equations*, Stoch. Proc. Appl. 128(2018), 595-621.
- [23] L. Xie, X. Zhang, *Ergodicity of stochastic differential equations with jumps and singular coefficients*, Ann. Inst. H. Poincaré Probab. Statist. 56(2020), 175-229.
- [24] C. G. Yuan, S.-Q. Zhang, *A study on Zvonkin’s transformation for stochastic differential equations with singular drift and related applications*, arXiv:1910.05903.
- [25] X. Zhang, *Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients*, Electron. J. Probab. 16(2011), 1096-1116.
- [26] A. K. Zvonkin, *A transformation of the phase space of a diffusion process that removes the drift*, Math. Sb. 93(1974), 129-149.