McKean-Vlasov SDEs with Drifts Discontinuous under Wasserstein Distance*

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

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

xinghuang@tju.edu.cn

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

wangfy@tju.edu.cn

April 22, 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^θ-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: 60H10, 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 $\mathbb{R}^d$:

$$(1.1)\qquad \mathrm{d}X_t = b_t(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma_t(X_t, \mathscr{L}_{X_t})\mathrm{d}W_t, \quad t \in [0, T],$$

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 $(\Omega, \{\mathscr{F}_t\}_{t\in[0,T]}, \mathbb{P})$, $\mathscr{L}_{X_t}$ is the law of X_t,

$$b : [0, T] \times \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}^d, \;\; \sigma : [0, T] \times \mathbb{R}^d \times \mathscr{P} \to \mathbb{R}^d \otimes \mathbb{R}^m$$

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

are measurable, and $\mathscr{P}$ 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

$$\mathbb{E}\int_0^T \big\{|b_t(X_t,\mathscr{L}_{X_t})| + \|\sigma_t(X_t,\mathscr{L}_{X_t})\|^2\big\}\mathrm{d}t < \infty, \tag{1.2}$$

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

$$X_t = X_0 + \int_0^t b_s(X_s,\mathscr{L}_{X_s})\mathrm{d}s + \int_0^t \sigma_s(X_s,\mathscr{L}_{X_s})\mathrm{d}W_s, \quad t\in[0,T]. \tag{1.3}$$

We call (1.1) (strongly) well-posed for an $\mathscr{F}_0$-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 $\mathscr{L}_\xi|\tilde{\mathbb{P}}$ to denote the law of a random variable ξ under the probability $\tilde{\mathbb{P}}$, and use $\mathbb{E}_{\tilde{\mathbb{P}}}$ to stand for the expectation under $\tilde{\mathbb{P}}$. For any $\mu_0\in\mathscr{P}$, $(\tilde X_t,\tilde W_t)_{t\in[0,T]}$ is called a weak solution to (1.1) starting at μ_0, if $(\tilde W_t)_{t\in[0,T]}$ is the m-dimensional Brownian motion under a complete filtration probability space $(\tilde\Omega,\{\tilde{\mathscr{F}}_t\}_{t\in[0,T]},\tilde{\mathbb{P}})$, $(\tilde X_t)_{t\in[0,T]}$ is a continuous $\tilde{\mathscr{F}}_t$-adapted process on $\mathbb{R}^d$ with $\mathscr{L}_{\tilde X_0}|\tilde{\mathbb{P}}=\mu_0$, and (1.2)-(1.3) hold for $(\tilde X,\tilde W,\tilde{\mathbb{P}},\mathbb{E}_{\tilde P})$ replacing $(X,W,\mathbb{P},\mathbb{E})$. We call (1.1) weakly well-posed for an initial distribution μ_0, if it has a unique weak solution starting at μ_0; i.e. it has a weak solution $(\tilde X_t,\tilde W_t)_{t\in[0,T]}$ with initial distribution μ_0 under some complete filtration probability space $(\tilde\Omega,\{\tilde{\mathscr{F}}_t\}_{t\in[0,T]},\tilde{\mathbb{P}})$, and $\mathscr{L}_{\tilde X_{[0,T]}}|\tilde{\mathbb{P}}=\mathscr{L}_{\bar X_{[0,T]}}|\bar{\mathbb{P}}$ holds for any other weak solution with the same initial distribution $(\bar X_t,\bar W_t)_{t\in[0,T]}$ under some complete filtration probability space $(\bar\Omega,\{\bar{\mathscr{F}}_t\}_{t\in[0,T]},\bar{\mathbb{P}})$.

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 γ, 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 σ implies the well-posedness of (1.1). To this end, we recall the L^θ-Wasserstein distance $\mathbb{W}_\theta$ for $\theta>0$:

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

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

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

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,γ, and

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

holds for some function $\phi \in C((0,\infty);(0,\infty))$ with $\int_0^\cdot \frac{1}{\phi(s)}\mathrm{d}s = \infty$. 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

$$b_t(x,\gamma) := \int_{\mathbb{R}^d} \tilde b_t(x,y)\gamma(\mathrm{d}y), \tag{1.4}$$

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 Lipschtiz 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

$$|\tilde b_t(x,y)| \le h_t(x-y)$$

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

$$b_t(x,\gamma) = \tilde b_t(x,\gamma(\varphi)), \tag{1.5}$$

where $\gamma(\varphi) := \int_{\mathbb{R}^d} \varphi \mathrm{d}\gamma$ for some α-Hölder continuous function φ, 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 γ 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 $\theta \ge 1$, $b_t(x,\gamma)$ is Lipschitz continuous in γ with respect to $\mathbb{W}_\theta$, and for any $\mu \in C([0,T];\mathscr{P}_\theta)$,

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

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

$$L^q_p(T) := \Big\{ f \in \mathscr{B}([0,T]\times\mathbb{R}^d) : \int_0^T \Big(\int_{\mathbb{R}^d} |f_t(x)|^p \mathrm{d}x \Big)^{\frac{q}{p}} \mathrm{d}t < \infty \Big\},$$
$$\mathscr{K} := \Big\{ (p,q) \in (1,\infty)\times(1,\infty) : \ \frac{d}{p} + \frac{2}{q} < 2 \Big\}.$$

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

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

with initial distribution having density function (with respect to the Lebesgue mseaure) in the class $H^{2+\alpha}$ for some $\alpha > 0$, where ρ_μ is the density function of μ 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 σ 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 γ 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

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

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

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

We will use the following conditions.

(A_σ) $\sigma_t(x,\gamma) = \sigma_t(x)$ 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 $(p,q) \in \mathscr{K}$; and there exists a constant $K_1 \geq 1$ such that

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

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

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

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

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

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

where $\bar{b}^\mu_t(x) := \bar{b}_t(x,\mu_t)$ for $(t,x) \in [0,T]\times\mathbb{R}^d$, and δ_0 stands for the Dirac measure at the point $0 \in \mathbb{R}^d$.

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

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

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

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

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

Theorem 1.1. *Assume* (A_σ).

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

$$\|P_t^*\mu_0 - P_t^*\nu_0\|_{TV}^2 \le 2\mathrm{e}^{\frac{K_1 K_3^2 t}{2}} \|\mu_0 - \nu_0\|_{TV}^2, \quad t \in [0,T], \mu_0, \nu_0 \in \mathscr{P}. \tag{1.12}$$

(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* μ_0. *Moreover, there exists a constant* $c > 0$ *such that for any* $\mu_0, \nu_0 \in \mathscr{P}_\theta$,

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

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 μ_0, it is natural to take a drift depending on a probability distance between μ_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 $\theta \ge 1$, where $\bar b$ satisfies (1.8) and (1.9) for $\hat b = 0$ which refers to the singularity in the space variable x, and $h : [0,T] \times \mathbb{R}^d \times [0,\infty)^2 \to \mathbb{R}^d$ 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 μ 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* $\mu_t = \mathscr{L}_{\bar{X}_t}|\bar{\mathbb{P}}, t \in [0,T]$*. If the SDE*

$$\mathrm{d}X_t = b_t(X_t, \mu_t)\mathrm{d}t + \sigma_t(X_t, \mu_t)\mathrm{d}W_t \tag{2.1}$$

has strong uniqueness for some initial value X_0 *with* $\mathscr{L}_{X_0} = \mu_0$*, 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* $\mathscr{L}_{X_0} = \mu_0$*, then it is weakly well-posed for the initial distribution* μ_0*.*

Proof. (a) Strong existence. Since $\mu_t = \mathscr{L}_{\bar{X}_t}|\bar{\mathbb{P}}$, $\bar{X}_t$ under $\bar{\mathbb{P}}$ is also a weak solution of (2.1) with initial distribution μ_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. μ_0). In particular, the weak uniqueness implies $\mathscr{L}_{X_t} = \mu_t, t \in [0,T]$, 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 μ_0. For any initial value X_0 with $\mathscr{L}_{X_0} = \mu_0$, the strong uniqueness of (2.1) starting at X_0 implies

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

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

$$\mathscr{L}_{\bar{X}_{[0,T]}}|\bar{\mathbb{P}} = \mathscr{L}_{X_{[0,T]}}|\mathbb{P}. \tag{2.2}$$

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

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

This and (2.2) imply $\mathscr{L}_{\hat{X}_t}|\tilde{\mathbb{P}} = \mu_t$, 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

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

i.e. (1.1) has weak uniqueness starting at μ_0. □

We will use the following result for the maximal operator:

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

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,*

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

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

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

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

$$\mathrm{d}X_t^\mu = b_t^\mu(X_t^\mu)\mathrm{d}t + \sigma_t(X_t^\mu)\mathrm{d}W_t, \quad t \in [0,T]. \tag{2.6}$$

According to [25], assumption (A_σ) 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 $\gamma \in \mathscr{P}$, Let $\Phi_t^\gamma(\mu) = \mathscr{L}_{X_t^\mu}$ for $(X_t^\mu)_{t\in[0,T]}$ solving (2.6) with $\mathscr{L}_{X_0^\mu} = \gamma$. We have the following result.

Lemma 2.3. *Assume (A_σ) and let $\gamma \in \mathscr{P}$.*

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

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

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

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

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

$$\mathrm{d}X_t^\mu = b_t(X_t^\mu, \nu_t)\mathrm{d}t + \sigma_t(X_t^\mu)\mathrm{d}\tilde{W}_t, \tag{2.9}$$

where

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

Noting that (1.7) together with (1.11) implies

$$\mathbb{E}[\mathrm{e}^{\frac{1}{2}\int_0^T |\xi_s|^2 \mathrm{d}s}] < \infty, \tag{2.10}$$

by the Girsanov theorem we see that $R_T := \mathrm{e}^{-\int_0^T \langle \xi_s, \mathrm{d}W_s\rangle - \frac{1}{2}\int_0^T |\xi_s|^2 \mathrm{d}s}$ 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 $\mathscr{L}_{X_0^\mu}|\mathbb{Q} = \mathscr{L}_{X_0^\mu} = \gamma$, we conclude from (2.9) with $\mathbb{Q}$-Brownian motion $\tilde{W}_t$ that

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

Combining this with (A_σ) and applying Pinker's inequality [18], we obtain

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

By (A_b'), this implies (2.7).

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

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

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

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

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

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

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

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

$$\|\mathbf{u}^{\lambda,\mu}\|_\infty + \|\nabla \mathbf{u}^{\lambda,\mu}\|_\infty \le \frac{1}{5}, \tag{2.14}$$

and

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

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

$$\mathrm{d}\Theta_t^{\lambda,\mu}(X_t^\mu) = (\lambda \mathbf{u}_t^{\lambda,\mu} + \hat{b}_t^\mu)(X_t^\mu)\mathrm{d}t + (\{\nabla\Theta_t^{\lambda,\mu}\}\sigma_t)(X_t^\mu)\,\mathrm{d}W_t. \tag{2.16}$$

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 $\mathscr{L}_{\xi_\cdot} \in C([0,T];\mathscr{P}_\theta)$ and so does $\mathscr{L}_{X_\cdot^\mu}$ due to (2.14).

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

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

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

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

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

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

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

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

and

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

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

$$\mathrm{d}|\eta_t|^{2m} \le c_1|\eta_t|^{2m}\mathrm{d}A_t + c_1(\mathbb{W}_\theta(\mu_t,\nu_t)^{2m} + \|\mu_t-\nu_t\|_{TV}^{2m})\mathrm{d}t + \mathrm{d}M_t \tag{2.20}$$

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

$$A_t := \int_0^t \Big\{1 + \big(\mathscr{M}\big(\|\nabla^2\Theta_s^{\lambda,\mu}\| + \|\nabla\sigma_s\|\big)(X_s^\mu) + \mathscr{M}\big(\|\nabla^2\Theta_s^{\lambda,\mu}\| + \|\nabla\sigma_s\|\big)(X_s^\nu)\big)^2\Big\}\mathrm{d}s.$$

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

$$\begin{aligned}&\mathbb{E}\bigg[\int_s^t |f_r|(X_r^\mu)\mathrm{d}r\bigg|\mathscr{F}_s\bigg] + \mathbb{E}\bigg[\int_s^t |f_r|(X_r^\nu)\mathrm{d}r\bigg|\mathscr{F}_s\bigg]\\ &\le C\bigg(\int_s^t\Big(\int_{\mathbb{R}^d}|f_r(x)|^p\mathrm{d}x\Big)^{\frac{q}{p}}\mathrm{d}r\bigg)^{\frac{1}{q}},\ \ 0\le s<t\le T.\end{aligned} \tag{2.21}$$

holds. As shown in [23, Lemma 3.5], (2.21), (2.5), (2.15) and (A_σ) imply

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

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

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

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.

□

3 Proof of Theorem 1.1

Assume (A_σ). According to [25, Theorem 1.3], for any $\mu_\cdot \in \mathscr{B}([0,T];\mathscr{P})$, 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 cosnider 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 $t_0 = \min\{T, \frac{1}{K_1K_3^2}\}$ and consider the space $E_{t_0} := \{\mu \in \mathscr{B}([0,t_0];\mathscr{P}) : \mu_0 = \gamma\}$ equipped with the complete metric

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

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

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

has a unique solution $\mu \in E_{t_0}$. By (3.1) and the definition of Φ^γ we see that the unique solution of (2.1) is a strong solution of (1.1). On the other hand, $\mu_t := \mathscr{L}_{X_t}$ 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

$$\|P_t^*\mu_0 - \Phi_t^{\mu_0}(\nu)\|_{TV}^2 \le \frac{K_1K_3^2}{4}\int_0^t \|P_s^*\mu_0 - P_s^*\nu_0\|_{TV}^2 \mathrm{d}s, \quad t \in [0,T]. \tag{3.2}$$

On the other hand, by the Markov property for the solution to (2.6) with ν replacing μ, we have

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

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

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

Hence,

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

This together with (3.2) yields

$$\begin{aligned}&\|P_t^*\mu_0 - P_t^*\nu_0\|_{TV}^2 \le 2\|P_t^*\mu_0 - \Phi_t^{\mu_0}(\nu)\|_{TV}^2 + 2\|\Phi_t^{\mu_0}(\nu) - P_t^*\nu_0\|_{TV}^2 \\ &\le 2\|\mu_0-\nu_0\|_{TV}^2 + \frac{K_1K_3^2}{2}\int_0^t \|P_s^*\mu_0 - P_s^*\nu_0\|_{TV}^2\mathrm{d}s, \quad t \in [0,T].\end{aligned}$$

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

(b) Let (A_b) hold and let $\gamma = \mathscr{L}_{X_0} \in \mathscr{P}_\theta$. For any $\mu,\nu \in C([0,T],\mathscr{P}_\theta)$, (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

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

Let $t_0 = \frac{1}{2C}$. We consider the space $\tilde{E}_{t_0} := \{\mu \in C([0,t_0];\mathscr{P}_\theta) : \mu_0 = \gamma\}$ 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 Φ^γ is strictly contractive in $\tilde{E}_{t_0}$, so that the same argument in (a) proves the strong well-posedness of (1.1) with $\mathscr{L}_{X_0} = \gamma$ up to time t_0.

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

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

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

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

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

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

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 smothness,* 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.