MCKEAN-VLASOV SDES WITH DRIFTS DISCONTINUOUS UNDER WASSERSTEIN DISTANCE

XING HUANG

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

FENG-YU WANG*

Center for Applied Mathematics, Tianjin University, Tianjin 300072, China
Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, United Kingdom
(Communicated by Lorenzo Zambotti)

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 $W_0$ or $W_0 + W_\theta$ for some $\theta \geq 1$, where $W_0$ is the total variation distance and $W_\theta$ is the $L^\theta$-Wasserstein distance. This improves some existing results (see for instance [13]) derived for drifts continuous in the distribution variable with respect to the Wasserstein distance.

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

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

where $T > 0$ is a fixed time, $(W_t)_{t \in [0,T]}$ is an $n$-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 $X_t$,

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

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

This type of 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, 21, 23] and references within.

An adapted continuous process on $\mathbb{R}^d$ is called a (strong) solution of (1), if

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

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

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

We call (1) (strongly) well-posed for an $\mathcal{F}_0$-measurable initial value $X_0$, if (1) has a unique solution starting at $X_0$.

2020 Mathematics Subject Classification. Primary: 60H10; Secondary: 60G44.

Key words and phrases. Distribution dependent SDEs, total variation distance, Wasserstein distance, Krylov’s estimate, Zvonkin’s transform.

Supported in part by NNSFC (11771326, 11831014, 11801406, 11921001).

* Corresponding author: Feng-Yu Wang.
When a different probability measure \( \tilde{P} \) is concerned, we use \( L_\xi|\tilde{P} \) to denote the law of a random variable \( \xi \) under the probability \( \tilde{P} \), and use \( E_{\tilde{P}} \) to stand for the expectation under \( \tilde{P} \). For any \( \mu_0 \in \mathcal{P} \), \((\tilde{X}_t, \tilde{W}_t)_{t \in [0,T]} \) is called a weak solution to (1) starting at \( \mu_0 \), if \((\tilde{W}_t)_{t \in [0,T]} \) is an \( n \)-dimensional Brownian motion under a complete filtration probability space \((\tilde{\Omega}, \{\tilde{\mathcal{F}}_t\}_{t \in [0,T]}, \tilde{P}) \), \((\tilde{X}_t)_{t \in [0,T]} \) is a continuous \( \tilde{\mathcal{F}}_t \)-adapted process on \( \mathbb{R}^d \) with \( L_{\tilde{X}_0}|\tilde{P} = \mu_0 \), and (2)-(3) hold for \((\tilde{X}, \tilde{W}, \tilde{P}, \mathbb{E}) \) replacing \((X, W, \mathbb{P}, \mathbb{E}) \). We call (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 \((\tilde{\Omega}, \{\tilde{\mathcal{F}}_t\}_{t \in [0,T]}, \tilde{P}) \), and \( L_{\tilde{X}_0}|\tilde{P} = L_{\tilde{X}_0}|\tilde{P} \) 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 \((\tilde{\Omega}, \{\tilde{\mathcal{F}}_t\}_{t \in [0,T]}, \tilde{P}) \).

Recently, the (weak and strong) well-posedness is studied in [3, 4, 6, 13, 17, 18, 20] for (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, 17] for the case with memory. We briefly recall some conditions on \( b \) which together with a regular and non-degenerate condition on \( \sigma \) imply the well-posedness of (1). To this end, we recall the \( L^2 \)-Wasserstein distance \( W_\theta \) for \( \theta \in (0, \infty) \):

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

where \( \mathcal{C}(\gamma, \tilde{\gamma}) \) is the set of all couplings of \( \gamma \) and \( \tilde{\gamma} \). Under this metric,

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

is a Polish space. By the convention that \( r^0 = \mathbb{1}_{\{r > 0\}} \) for \( r \geq 0 \), we may regard \( W_0 \) as the total variation distance, i.e. set

\[
W_0(\gamma, \tilde{\gamma}) = \|\gamma - \tilde{\gamma}\|_{TV} := \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\gamma(A) - \tilde{\gamma}(A)|.
\]

References [3, 4] prove the well-posedness of (1) with a deterministic initial value \( X_0 \in \mathbb{R}^d \) for \( b_t(x, \gamma) \) of linear growth in \( x \) uniformly in \( t, \gamma \) with

\[
|b_t(x, \gamma) - b_t(x, \tilde{\gamma})| \leq \phi(W_1(\gamma, \tilde{\gamma})), \quad t \geq 0, \gamma, \tilde{\gamma} \in \mathcal{P}_1
\]

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

[18, Theorem 3] confirms the well-posedness of (1) for an exponentially integrable initial value \( X_0 \) and a drift \( b_t(x, \gamma) \) of type

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

where \( \tilde{b}_t(x, y) \) has linear growth in \( x \) uniformly in \( t \) and \( y \). Since for any \( x \), \( \tilde{b}_t(x, y) \) is bounded in \( y \), \( b_t(x, \cdot) \) is Lipschitz continuous in the total variation distance \( W_0 \).

[20] considers the same type of drift and proves the well-posedness of (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 \( h \in L^q([0, T]; \hat{L}^p(\mathbb{R}^d)) \) with \( p, q > 1 \) satisfying \( \frac{d}{p} + \frac{2}{q} < 1 \), where \( \hat{L}^p \) is a localized \( L^p \) space.

In [6] the well-posedness of (1) is proved for \( X_0 \) satisfying \( \mathbb{E}|X_0|^2 < \infty \), and for \( b \) given by

\[
b_t(x, \gamma) = \hat{b}_t(x, \gamma(\varphi)),
\]

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

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 \( \mathcal{W}_\theta \), and for any \( \mu \in C([0, T]; \mathcal{P}_b) \),

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

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

\[
K := \left\{(p, q) \in (1, \infty) \times (1, 0): \frac{d}{p} + \frac{2}{q} < 1\right\},
\]

and we write \( f \in L^q_p(T) \) for a Banach-valued measurable function \( f \) on \([0, T] \times \mathbb{R}^d\), if for \( |\cdot| \) being the Banach-norm we have

\[
\|f\|_{L^q_p(T)} := \left( \int_0^T \left( \int_{\mathbb{R}^d} |f_t(x)|^p \, dx \right)^\frac{q}{p} \, dt \right)^\frac{1}{q} < \infty.
\]

Moreover, we denote \( f \in L^q_p,loc(T) \), if \( \int_0^T \left( \int_{\mathbb{R}^d} |f_t(x)|^p \, dx \right)^\frac{q}{p} \, dt < \infty \) holds for any compact set \( K \subset \mathbb{R}^d \). We remark that the drift \( b \) in [13] is not necessarily to have the form as assumed in [20], but with the same integrability condition used therein.

Moreover, in [15] the well-posedness of (1) has been proved for

\[
b_t(x, \mu) = \hat{b}(\rho_\mu(x)), \quad \sigma_t(x, \mu) = \hat{\sigma}(\rho_\mu(x))
\]

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, \( \hat{b} \in C^2([0, \infty); \mathbb{R}^d) \) and \( \hat{\sigma} \in C^0([0, \infty); \mathbb{R}^d \otimes \mathbb{R}^m) \). As for the weak well-posedness, [14] assumes that \( b \) is bounded and \( \mathcal{W}_0 \)-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) for general type \( b \) with \( b_t(x, \gamma) \) Lipschitz continuous in \( \gamma \) under the metric \( \mathcal{W}_\theta \) or \( \mathcal{W}_0 + \mathcal{W}_\theta \) 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 [18, 20] where \( b \) is of the integral type as in (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

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

In this case, we write \( \xi = \nabla f \) and call it the weak gradient of \( f \). We will use the following conditions.
(Aσ) \( \sigma_t(x, \gamma) = \sigma_t(x) \) 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. \( t \in [0, T] \) satisfying \( |\nabla \sigma|^2 \in L^p_{\text{loc}}(T) \) for some \( (p, q) \in K; \) and there exists a constant \( K_1 \geq 1 \) such that

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

where \( I \) is the \( d \times d \) identity matrix.

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

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

for some constants \( \theta, K_2 \geq 1 \). Moreover, there exists \( (p, q) \in (A_\sigma) \) such that

\[
\sup_{t \in [0, T]} |\hat{b}_t(0, \delta_0)| + \sup_{\mu \in C([0, T]; \mathcal{P}_\theta)} \|\hat{b}^\theta\|^2_{L^q_{\text{loc}}(T)} < \infty,
\]

where \( \hat{b}^\theta(x) : = \hat{b}_t(x, \mu_t) \) 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 \( \mu \in \mathcal{B}([0, T]; \mathcal{P}) \), the class of measurable maps from \([0, T]\) to \( \mathcal{P} \), there exists \( (p, q) \in (A_\sigma) \) such that \( |b^\theta|^2 \in L^q_{p, \text{loc}}(T) \). Moreover, there exists an increasing function \( \Gamma : [0, \infty) \to (0, \infty) \) satisfying \( \int_1^\infty \frac{1}{\Gamma(x)} \, dx = \infty \) such that

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

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

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

When \((1)\) 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 \((1)\) is strongly and weakly well-posed for any initial values and any initial distribution. Moreover,

\[
\|P^*_t \mu_0 - P^*_t \nu_0\|_{TV}^2 \leq 2e^{-\frac{\kappa_1 \kappa_2 T}{2}} \|\mu_0 - \nu_0\|_{TV}^2, \quad t \in [0, T], \mu_0, \nu_0 \in \mathcal{P}.
\]

2. Let \((A_b)\) hold. Then \((1)\) is strongly well-posed for initial value \( X_0 \) with \( \mathbb{E}[X_0]^\theta < \infty \) and weakly well-posed for initial distribution \( \mu_0 \in \mathcal{P}_\theta \). Moreover, for any \( m \in (\frac{d}{2}, \infty) \cap [1, \infty) \), there exists a constant \( c > 0 \) such that

\[
\|P^*_t \mu_0 - P^*_t \nu_0\|_{TV} + \mathbb{W}_\theta(P^*_t \mu_0, P^*_t \nu_0) \\
\leq c \left\{ \|\mu_0 - \nu_0\|_{TV} + \mathbb{W}_{2m}(\mu_0, \nu_0) \right\}, \quad t \in [0, T], \mu_0, \nu_0 \in \mathcal{P}_\theta.
\]

We remark that estimate \((13)\) makes sense for \( \nu_0, \mu_0 \in \mathcal{P}_\theta \) since \( \mathbb{W}_{2m}(\mu_0, \nu_0) \) may be finite even though \( \nu_0, \mu_0 \notin \mathcal{P}_{2m} \). To illustrate this result comparing with earlier ones, we present an example of \( b \) which satisfies our conditions but is not of type \((4)-(6)\) 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 used 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 (8) and (9) for \( \bar{b} = 0 \) 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 (A0) but is not of type (4)-(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) 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

\[
dX_t = b_t(X_t, \mu_t)dt + \sigma_t(X_t, \mu_t)dW_t
\]

has strong uniqueness for some initial value \( X_0 \) with \( \mathcal{L}_{X_0} = \mu_0 \), then (1) has a strong solution starting at \( X_0 \). If moreover (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 (14) with initial distribution \( \mu_0 \). By the Yamada-Watanabe principle, the strong uniqueness of (14) with initial value \( X_0 \) implies the strong (resp. weak) well-posedness of (14) starting at \( X_0 \) (resp. \( \mu_0 \)). In particular, the weak uniqueness implies \( \mathcal{L}_{\bar{X}_t} = \mu_t, t \in [0, T] \), so that \( \bar{X}_t \) solves (1).

(b) Weak uniqueness. Let \( (\tilde{X}_t, \tilde{W}_t) \) under probability \( \tilde{\mathbb{P}} \) be another weak solution of (1) with initial distribution \( \mu_0 \). For any initial value \( X_0 \) with \( \mathcal{L}_{X_0} = \mu_0 \), the strong uniqueness of (14) 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 (14) proved in (a) yield

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

This and (15) imply \( \mathcal{L}_{\bar{X}_t} | \bar{\mathbb{P}} = \mu_t \), so that \( \bar{X}_t \) under \( \bar{\mathbb{P}} \) is a weak solution of (1) with \( \bar{X}_0 = \bar{X}_0 \). By the strong uniqueness of (1), we derive \( \bar{X}_{[0, T]} = \tilde{X}_{[0, T]} \). Combining this with (15) we obtain

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

i.e. (1) has weak uniqueness starting at \( \mu_0 \).

We will use the following result for the maximal operator:

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

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 $f$,
\[
|f(x) - f(y)| \leq C|x - y|(|\mathcal{M}\nabla f(x)| + |\mathcal{M}\nabla f(y)|), \quad \text{a.e. } x, y \in \mathbb{R}^d. \tag{17}
\]
Moreover, for any $p > 1$, there exists a constant $C_p > 0$ such that
\[
||\mathcal{M}f||_{L^p} \leq C_p ||f||_{L^p}, \quad f \in L^p(\mathbb{R}^d). \tag{18}
\]

To compare the distribution dependent SDE (1) with a classical one, for any $\mu \in \mathcal{B}([0,T];\mathcal{P})$, let $b^\mu_t(x) := b_t(x, \mu_t)$ and consider the classical SDE
\[
dX^\mu_t = b^\mu_t(X^\mu_t)dt + \sigma_t(X^\mu_t)dW_t, \quad t \in [0, T]. \tag{19}
\]
According to [27, Theorem 1.3], assumption $(A_\sigma)$ together with $(A'_\sigma)$ implies the strong well-posedness up to the explosion time. Moreover, $(A_\sigma)$ and $(A'_\sigma)$ yield that $\sigma_t$ is bounded and
\[
(b_t(x, \gamma), x) \leq (b_t(x, \delta_0), x) + K_3|x| \leq \Gamma(|x|^2) + K_3|x|, \quad t \in [0, T], x \in \mathbb{R}^d, \gamma \in \mathcal{P}
\]
holds for some positive increasing function $\Gamma$ with $\int_1^\infty \frac{dx}{\Gamma(x)} = \infty$, so that [22, Theorem 1.1(2)] ensures the non-explosion.

According to [27, Theorem 1.3], for any $\mu \in C([0,T];\mathcal{P}_b)$, assumption $(A_\sigma)$ together with $(A_\sigma)$ yields the strong well-posedness of (19) up to the explosion time. To see that the solution is also non-explosive, for any $n \geq 1$, let $\tau_n = \inf\{t > 0 : |X^\mu_t| \geq n\}$. By Krylov’s estimate [25, Theorem 3.1], we have the Khasminskii type estimate $[13, (2.2)]$. This together with (9) implies
\[
E \left( \int_0^{T \wedge \tau_n} |\tilde{b}^\mu_t(X^\mu_t)|dt \right)^\theta \leq C_1, \quad n \geq 1
\]
for some constant $C_1 > 0$ depending on $\theta, T$. Thus, by (8), (9), (7), Hölder’s inequality and the Burkholder-Davis-Gundy inequality, we find constants $C_2, C_3 > 0$ depending on $\theta, T$ and $\mu$, such that
\[
E \sup_{t \in [0,u \wedge \tau_n]} |X^\mu_t|^\theta \leq C_2 E|X_0^\mu|^\theta + C_2 u^{\theta - 1} E \int_0^{u \wedge \tau_n} (1 + |X^\mu_t|^\theta + \mu_t(\cdot)^\theta)dt
\]
\[
+ C_2 E \left( \int_0^{u \wedge \tau_n} |\tilde{b}^\mu_t(X^\mu_t)|dt \right)^\theta + C_2 E \sup_{s \in [0,u]} \left| \int_0^{s \wedge \tau_n} \sigma_t(X^\mu_t)dW_t \right|^\theta,
\]
\[
\leq C_3 E|X_0^\mu|^\theta + C_3 + C_3 E \int_0^u |X^\mu_{t \wedge \tau_n}|^\theta dt, \quad u \in [0, T].
\]
By Gronwall’s inequality, we get $\sup_{n \geq 1} E \sup_{t \in [0,T \wedge \tau_n]} |X^\mu_t|^\theta < \infty$ if $E|X_0^\mu|^\theta < \infty$. This implies that the solution is non-explosive before time $T$. See also [16] for the discussion with constant $\sigma_t$.

Therefore, in the two situations of the following Lemma 2.3, for any $\gamma \in \mathcal{P}$, (19) has a unique global solution $(X^\mu_t)_{t \in [0,T]}$ with $\mathcal{L}^\mu_{X^\mu_t} = \gamma$. We denote $\Phi^\gamma_t(\mu) = \mathcal{L}^\mu_{X^\mu_t}$ to emphasize the dependence on $\mu$ and $\gamma$.

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

(1) If $(A'_\sigma)$ holds, then for any $\mu, \nu \in \mathcal{B}([0,T];\mathcal{P})$,
\[
||\Phi^\gamma_t(\mu) - \Phi^\gamma_t(\nu)||^2_{TV} \leq \frac{K_1 K_3^2}{4} \int_0^t ||\mu_s - \nu_s||^2_{TV} ds, \quad t \in [0, T]. \tag{20}
\]
(2) If $(A_b)$ holds and $\gamma \in \mathcal{P}_\theta$, then for any $\mu \in C([0, T]; \mathcal{P}_\theta)$, we have $\Phi^\gamma_1(\mu) \in C([0, T]; \mathcal{P}_\theta)$. Moreover, for any $m \in (\frac{9}{2}, \infty) \cap [1, \infty)$, there exists a constant $C > 0$ such that for any $\mu, \nu \in C([0, T]; \mathcal{P}_\theta)$ and $\gamma_1, \gamma_2 \in \mathcal{P}_\theta$,

$$\{\mathcal{W}_\theta(\Phi_1^\gamma(\mu), \Phi_1^\gamma(\nu))\}^{2m} \leq C\mathcal{W}_{2m}(\gamma_1, \gamma_2)^{2m} + C \int_0^t \{||\mu_s - \nu_s||_{TV} + \mathcal{W}_\theta(\mu_s, \nu_s)\}^{2m} \, ds, \quad t \in [0, T].$$

Proof. (1) Let $(A'_b)$ hold and take $\mu, \nu \in \mathcal{B}([0, T]; \mathcal{P})$. To compare $\Phi_1^\gamma(\mu)$ with $\Phi_1^\gamma(\nu)$, we rewrite (19) as

$$dX^\mu_t = b_t(X^\mu_t, \nu_t)dt + \sigma_t(X^\mu_t)\, d\tilde{W}_t,$$

where

$$\tilde{W}_t = W_t + \int_0^t \xi_s \, ds, \quad \xi := \{s_{\sigma_s}^{-2}(\sigma_s\sigma_s^*)^{-1}\} (X^\mu_s) [b_s(X^\mu_s, \mu_s) - b_s(X^\mu_s, \nu_s)], \quad s, t \in [0, T].$$

Noting that (7) together with (11) implies

$$\mathbb{E}[e^{\frac{1}{2} \int_0^T |\xi|^2 \, ds}] < \infty,$$

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

By the weak uniqueness of (19) and $\mathcal{L}X^\mu_0|Q = \mathcal{L}X^\nu_0 = \gamma$, we conclude from (22) with $Q$-Brownian motion $\tilde{W}_t$ that

$$\Phi_1^\gamma(\nu) = \mathcal{L}X^\nu_0|Q, \quad t \in [0, T].$$

Combining this with $(A_\nu)$ and applying Pinsker’s inequality [19], we obtain

$$4\|\Phi_1^\gamma(\nu) - \Phi_1^\gamma(\mu)\|_{TV}^2 = \sup_{\|f\|_{\infty} \leq 1} (\mathbb{E}[f(X^\mu_t)(R_t - 1)])^2 = (\mathbb{E}[R_t - 1])^2$$

$$\leq 2\mathbb{E}[R_t \log R_t] = \mathbb{E}_Q \int_0^t |\xi_s|^2 \, ds$$

$$\leq K_1 \mathbb{E}_Q \int_0^t |b_s(X^\mu_s, \mu_s) - b_s(X^\mu_s, \nu_s)|^2 \, ds.$$

By $(A'_b)$, this implies (20).

(2) Let $(A_b)$ hold and let $m \in (\frac{9}{2}, \infty) \cap [1, \infty)$. Take $\mathcal{F}_0$-measurable random variables $X^\mu_0$ and $X^\nu_0$ such that $\mathcal{L}X^\mu_0 = \gamma_1, \mathcal{L}X^\nu_0 = \gamma_2$ and

$$\mathbb{E}|X^\mu_0 - X^\nu_0|^{2m} = \mathcal{W}_{2m}(\gamma_1, \gamma_2)^{2m}.$$

Let $X^\mu_t$ solve (19) with initial value $X^\mu_0$ and $X^\nu_t$ solve the same SDE for $\nu$ replacing $\mu$ and with initial value $X^\nu_0$. We need to find a constant $C > 0$ such that for any $t \in [0, T],

$$\{\mathcal{W}_\theta(\Phi_1^\gamma(\mu), \Phi_1^\gamma(\nu))\}^{2m} \leq C\mathbb{E}|X^\mu_0 - X^\nu_0|^{2m} + C \int_0^t \{||\mu_s - \nu_s||_{TV} + \mathcal{W}_\theta(\mu_s, \nu_s)\}^{2m} \, ds, \quad t \in [0, T].$$

To this end, we make a Zvonkin type transform as in [13], [26] and [25]. To introduce this transform, we need to recall some notations introduced in [26]. For $(\alpha, p_1) \in (0, \infty) \times (1, \infty)$, let $H^{\alpha, p_1} := (I - \Delta)^{-\alpha/2}(L^{p_1}(\mathbb{R}^d))$ be the usual Bessel potential space with norm

$$\|f\|_{\alpha, p_1} := \|(I - \Delta)^{\alpha/2} f\|_{L^{p_1}}.$$
Let $\chi \in C_0^\infty(\mathbb{R}^d)$ be a smooth function with $\chi(x) = 1$ for $|x| \leq 1$ and $\chi(x) = 0$ for $|x| > 2$. For $r > 0$ and $z \in \mathbb{R}^d$, define

$$\chi_r(x) := \chi(x/r), \quad \chi_r^z(x) := \chi_r(x - z), \quad x \in \mathbb{R}^d.$$ 

For $q_1 > 1, r > 0$, let $\tilde{H}_{r,q_1}(T)$ be the localized space consisting of measurable functions on $[0, T] \times \mathbb{R}^d$ such that

$$\|f\|_{\tilde{H}_{r,q_1}(T)} := \sup_{z \in \mathbb{R}^d} \left( \int_0^T \|\chi_r^z f_t\|_{L^q_{t,r}^p} dt \right)^{\frac{1}{q}} < \infty.$$ 

In view of [26, (2.4)], for any $r, r' > 0$, $\|\cdot\|_{\tilde{H}_{r,q_1}(T)}$ is equivalent to $\|\cdot\|_{\tilde{H}_{r',q_1}(T)}$. Therefore, we will fix $r > 0$ and omit subscript $r$ for simplicity. Moreover, a multidimensional valued function is said in the class $\tilde{H}_{r,q_1}(T)$, if so are its component functions.

Let $\tilde{b}_t^i = \tilde{b}_t(x, \mu_t)$. For any $\lambda \geq 1$, consider the following PDE for $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$: 

$$\frac{\partial u_t}{\partial t} + \frac{1}{2} \text{Tr}(\sigma_t \sigma_t^* \nabla^2 u_t) + \{\nabla u_t\} b_t^\mu + \tilde{b}_t^\mu = \lambda u_t, \quad u_T = 0. \quad (26)$$

Firstly, according to [25, Corollary 2.2], $(A_r)$ and $(A_b)$, (26) has a unique solution $u^{\lambda,\mu}$ satisfying

$$\|\nabla^2 u^{\lambda,\mu}\|_{L^{2^*}_{p,T}(T)} < \infty. \quad (27)$$

Next, to estimate $\|\nabla u^{\lambda,\mu}\|_{\infty}$, we will adopt the following procedure. By [25, Remark 2.1], there exist measurable maps $\tilde{b}_1^{\lambda,\mu}, \tilde{b}_2^{\lambda,\mu} : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfying $\tilde{b}_1^{\lambda,\mu} \in C^2(\mathbb{R}^d, \mathbb{R}^d), t \in [0, T]$ with

$$\sup_{\mu \in C([0,T];\mathbb{R}^d)} \left( \sup_{t \in [0,T]} \left( |\tilde{b}_1^{\lambda,\mu}(0)| + \|\nabla \tilde{b}_1^{\lambda,\mu}\|_{\infty} + \|\nabla^2 \tilde{b}_1^{\lambda,\mu}\|_{\infty} + \|\tilde{b}_2^{\lambda,\mu}\|_{\infty} \right) \right) < \infty \quad \text{such that} \frac{\partial v_t}{\partial t} + \frac{1}{2} \text{Tr}(\sigma_t \sigma_t^* \nabla^2 v_t) + \{\nabla v_t\} (\tilde{b}_1^{\lambda,\mu} + \tilde{b}_2^{\lambda,\mu}) + \tilde{b}_t^\mu = \lambda v_t, \quad v_T = 0. \quad (28)$$

Using [26, Theorem 3.1, (3.2)] with $\tilde{b}_1^{\mu}$ and $\tilde{b}_2^{\lambda,\mu}$ in place of $f$ and $b$ respectively and Sobolev’s embedding [26, (2.5)], we conclude that under assumptions $(A_r)$ and $(A_b)$, (28) has a unique strong solution $v^{\lambda,\mu} \in \tilde{H}^{2,\mu}_{q,T}(T)$ with

$$\lim_{\lambda \to \infty} \sup_{\mu \in C([0,T];\mathbb{R}^d)} (\|v^{\lambda,\mu}\|_{\infty} + \|\nabla v^{\lambda,\mu}\|_{\infty}) = 0. \quad (29)$$

On the other hand, thanks to [25, (2.25), (2.22)], we arrive at $u^{\lambda,\mu}(t, x) = v^{\lambda,\mu}(t, \psi^{-1}(t, x))$, where $\psi : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ satisfies

$$\frac{\text{d}\psi}{\text{d}t} = -\tilde{b}_1^{\lambda,\mu}(t, \psi), \quad \psi(0, x) = x, \quad x \in \mathbb{R}^d$$

and

$$\|\nabla \psi^{-1}\|_{\infty} < \infty.$$ 

This together with (29) implies that when $\lambda$ is large enough, it holds that

$$\sup_{\mu \in C([0,T];\mathbb{R}^d)} (\|u^{\lambda,\mu}\|_{\infty} + \|\nabla u^{\lambda,\mu}\|_{\infty}) \leq \frac{1}{5}. \quad (30)$$
Let \( \Theta_t^{\lambda, \mu}(x) = x + u_t^{\lambda, \mu}(x) \). It is easy to see that (26) and the Itô formula [25, Lemma 3.3] imply

\[
\begin{align*}
d\Theta_t^{\lambda, \mu}(X_t^\nu) &= (\lambda u_t^{\lambda, \mu} + \dot{b}_t^\mu)(X_t^\nu)dt + (\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t)(X_t^\nu)dW_t. \quad (31)
\end{align*}
\]

In particular, (30) and \( \mathbb{E}[|X_0^\nu|^2] < \infty \) imply that \( \mathbb{E}[|\Theta_0^{\lambda, \mu}(X_0^\nu)|^2] < \infty \) and (31) is an SDE for \( \xi_t := \Theta_t^{\lambda, \mu}(X_t^\nu) \) with coefficients of at most linear growth, so that \( \mathcal{L}_\xi \in C([0, T]; \mathcal{P}_0) \) and so does \( \mathcal{L}_{X^\nu} \) due to (30).

It remains to prove (21). To this end, we observe that (26) and the Itô formula [25, Lemma 3.3] yield

\[
\begin{align*}
d\Theta_t^{\lambda, \mu}(X_t^\nu) &= \lambda u_t^{\lambda, \mu}(X_t^\nu)dt + (\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t)(X_t^\nu)dW_t \\
&\quad + [\{\nabla u_t^{\lambda, \mu}\} (b_t^\nu - \dot{b}_t^\nu) + \dot{b}_t^\nu - \dot{\tilde{b}}_t^\nu](X_t^\nu)dt \\
&= [\lambda u_t^{\lambda, \mu} + \{\nabla \Theta_t^{\lambda, \mu}\} (b_t^\nu - \dot{b}_t^\nu) + \dot{\tilde{b}}_t^\nu](X_t^\nu)dt + (\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t)(X_t^\nu)dW_t.
\end{align*}
\]

Combining this with (31) and applying the Itô formula, we see that \( \eta_t := \Theta_t^{\lambda, \mu}(X_t^\nu) - \Theta_t^{\lambda, \mu}(X_t^\nu) \) satisfies

\[
\begin{align*}
d|\eta_t|^2 &= 2 \left< \eta_t, \lambda u_t^{\lambda, \mu}(X_t^\nu) - \lambda u_t^{\lambda, \mu}(X_t^\nu) + \dot{b}_t^\nu - \dot{\tilde{b}}_t^\nu \right> dt \\
&\quad + 2 \left< \eta_t, \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) - \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) \right> dt \\
&\quad + \left\| \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) - \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) \right\|^2_{HS} dt \\
&\quad - 2 \left< \eta_t, [\{\nabla \Theta_t^{\lambda, \mu}\} (b_t^\nu - \dot{\tilde{b}}_t^\nu)](X_t^\nu) \right> dt.
\end{align*}
\]

So, for any \( m \geq 1 \), it holds

\[
\begin{align*}
d|\eta_t|^2m &= 2m|\eta_t|^{2(m-1)} \left< \eta_t, \lambda u_t^{\lambda, \mu}(X_t^\nu) - \lambda u_t^{\lambda, \mu}(X_t^\nu) + \dot{b}_t^\nu - \dot{\tilde{b}}_t^\nu \right> dt \\
&\quad + 2m|\eta_t|^{2(m-1)} \left< \eta_t, \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) - \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) \right> dt \\
&\quad + m|\eta_t|^{2(m-1)} \left\| \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) - \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) \right\|^2_{HS} dt \\
&\quad + 2m(m-1)|\eta_t|^{2(m-2)} \left| \left< \eta_t, [\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) - \{\{\nabla \Theta_t^{\lambda, \mu}\} \sigma_t\} \Theta_t^{\lambda, \nu}(X_t^\nu) \right> \eta_t \right|^2 dt \\
&\quad - 2m|\eta_t|^{2(m-1)} \left< \eta_t, [\{\nabla \Theta_t^{\lambda, \mu}\} (b_t^\nu - \dot{\tilde{b}}_t^\nu)](X_t^\nu) \right> dt.
\end{align*}
\]

By (30) and (8) and the triangular inequality, we may find a constant \( c_0 > 0 \) such that

\[
|\eta_t|^{2(m-1)}|\eta_t| \cdot |\lambda u_t^{\lambda, \mu}(X_t^\nu) - \lambda u_t^{\lambda, \mu}(X_t^\nu) + \dot{b}_t^\nu - \dot{\tilde{b}}_t^\nu| \leq c_0|\eta_t|^{2m}, \quad (33)
\]

and

\[
\begin{align*}
|\eta_t|^{2(m-1)}|\eta_t| \cdot |\{\{\nabla \Theta_t^{\lambda, \mu}\} (b_t^\nu - \dot{\tilde{b}}_t^\nu)](X_t^\nu)| \\
\leq K_2 \|\nabla \Theta_t^{\lambda, \mu}\|_{\infty} |\eta_t|^{2(m-1)}|\eta_t| |\nabla \varphi(\nu, \mu)| + \|\nu_t - \nu_t\|_{TV}, \quad (34)
\end{align*}
\]

where in the last display, we have used the Young inequality \( a^p b^{1-p} \leq pa + (1-p)b \) for \( a, b \geq 0 \) and \( p \in [0, 1] \). According to [13, (4.19)-(4.20)], (32)-(34) imply

\[
|\eta_t|^{2m} \leq |\eta_0|^{2m} + c_1 \int_0^s |\eta_t|^{2m} dA_t + c_1 \int_0^s (\nabla \varphi(\nu, \mu)) + \|\nu_t - \nu_t\|_{TV})^{2m} dt + M_s. \quad (35)
\]
for some constant $c_1 > 0$, a local martingale $M_t$, and

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

Thanks to [25, Theorem 3.1], the Krylov estimate

$$E \left[ \int_s^t |f_r|(X_r^\nu)dr \right]_{\mathcal{F}_s} + E \left[ \int_s^t |f_r|(X_r^\nu)dr \right]_{\mathcal{F}_s}$$

\[ \leq C \left( \int_s^t \left( \int_{\mathbb{R}^d} |f_r(x)|^p dx \right)^{\frac{2}{p}} \right)^{\frac{1}{4}}, \quad 0 < s < t \leq T. \tag{36} \]

holds. As shown in [24, Lemma 3.5], (36), (18), (27) and ($A_\sigma$) imply

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

Let $m \in (\frac{\theta}{2}, \infty) \cap [1, \infty)$. Noting that $|x-y| \leq \frac{\delta}{2} |\Theta_s^\lambda(x) - \Theta_s^\lambda(y)|$, $t \in [0,T], x, y \in \mathbb{R}^d$ due to (30), it follows from (35) and the stochastic Gronwall lemma [24, Lemma 3.8] with $q = \frac{\theta}{2m} < 1$ and any $p \in (q, 1)$ that

$$\{\mathcal{W}_0(\Phi_t^1(\mu), \Phi_t^2(\nu))\}^{2m} \leq c_2 (E|\eta|^2)^{\frac{2m}{p}}$$

\[ \leq \frac{c_3 \left( \frac{p}{p-q} \right)^{\frac{1}{2}} \left( E \mathcal{E}^{\frac{2m}{p}} \right)^{\frac{1}{2}}}{(E X_0^m - X_0^m)^{2m} + \int_0^t \left( \mathcal{W}_0(\mu_s, \nu_s) + \|\mu_s - \nu_s\|_{TV} \right)^2 ds} \]

for all $t \in [0,T]$ and some constants $c_2, c_3 > 0$. Therefore, (25) holds for some constant $C > 0$ and the proof is thus finished. \hfill \Box

### 3. Proof of theorem 1.1.

Assume ($A_\sigma$). By Lemma 2.1, the strong well-posedness of (1) implies the weak well-posedness. Therefore, in the following we only consider the strong solution.

To prove the strong well-posedness of (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) 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 ($A_0^\lambda$) hold. Take $t_0 = \min\{T, \frac{1}{K_1 K_3} \}$ 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 (20) 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

$$\Phi_\gamma^t(\mu) = \mu_t, \quad t \in [0,t_0] \tag{37}$$

has a unique solution $\mu \in E_{t_0}$. By (37) and the definition of $\Phi_\gamma$, the unique solution of (14) with $X_0$ satisfying $\mathcal{L}X_0 = \mu_0 = \gamma$ is a strong solution of (1). On the other hand, $\mu_t := \mathcal{L}X_t$ for any strong solution to (1) is a solution to (37) with $\gamma = \mathcal{L}X_0$, hence the uniqueness of (37) and the strong uniqueness of (14) implies that of (1).

To prove (12), let $\mu_t = P_t^\mu \mu_0$ and $\nu_t = P_t^\mu \nu_0$. We have $P_t^\mu \mu_0 = \Phi_\mu^t(\mu)$ and $P_t^\mu \nu_0 = \Phi_\nu^t(\nu)$. So, (20) with $\gamma = \mu_0$ implies

$$\|P_t^\mu \mu_0 - \Phi_\mu^t(\nu)\|_{TV}^2 \leq \frac{K_1 K_3^2}{4} \int_0^t \|P_s^\mu \mu_0 - P_s^\mu \nu_0\|_{TV}^2 ds, \quad t \in [0,T]. \tag{38}$$
Next, taking Combining this with (39), (41) and triangular inequality, we find a constant $C > 0$ such that

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

This together with (38) yields

\[
\begin{align*}
&\| P_t^\nu \mu_0 - P_t^\nu \nu_0 \|_{TV} \leq 2 \| P_t^\nu \mu_0 - \Phi_t^\mu (\nu) \|_{TV} + 2 \| \Phi_t^\mu (\nu) - P_t^\nu \nu_0 \|_{TV} \\
&\leq 2 \| \mu_0 - \nu_0 \|_{TV} + \frac{K_1 K_2}{2} \int_0^t \| P_s^\nu \mu_0 - P_s^\nu \nu_0 \|_{TV}^2 ds, \quad t \in [0, T].
\end{align*}
\]

By Gronwall’s lemma, this implies (12).

(b) Assume (A_0), let $m \in (\frac{d}{2}, \infty) \cap [1, \infty)$ and $\gamma = \mathcal{L}_{X_0} \in \mathcal{P}_\theta$. For any $\mu, \nu \in C([0, T], \mathcal{P}_\theta)$, (8) implies (23) and then (24). By (24), (8) and (21) with $\gamma_1 = \gamma_2 = \gamma$, we find a constant $C > 0$ such that

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

Let $t_0 = \frac{1}{\nu^2}$. 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

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

Then $\Phi^\nu$ is strictly contractive in $\tilde{E}_{t_0}$. Since by Lemma 2.3(2) for a strong solution $X_t$ of (1) we have $\mu_t := \mathcal{L}_{X_t} \in \tilde{E}_{t_0}$. Similarly to the last part of (a), this implies the strong well-posedness of (1) with $\mathcal{L}_{X_0} = \gamma$ up to time $t_0$.

Let $\mu_t = P_t^\nu \mu_0$ and $\nu_t = P_t^\nu \nu_0$ be as in (a). By (40) with $\gamma = \mu_0$ we obtain

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

Next, taking $\gamma_1 = \nu_0$, $\gamma_2 = \mu_0$ and $\mu = \nu$ in (21), we derive

\[ \{ \mathcal{W}_\theta (P_t^\nu \nu_0, \Phi_t^\nu (\nu)) \}^{2m} \leq C \{ \mathcal{W}_{2m} (\mu_0, \nu_0) \}^{2m}. \]

Combining this with (39), (41) and triangular inequality, we find a constant $C' > 0$ such that

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

By Gronwall’s lemma, this implies (13) for some constant $c > 0$. 

Acknowledgments. The authors would like to thank the referee and the associate editor for helpful comments and corrections.

REFERENCES

[1] V. Barbu and M. Röckner, Probabilistic representation for solutions to non-linear Fokker-Planck equations, SIAM J. Math. Anal., 50 (2018), 4246–4260.

[2] V. Barbu and M. Röckner, From non-linear Fokker-Planck equations to solutions of distribution dependent SDE, Ann. Probab., 48 (2020), 1902–1920.

[3] M. Bauer and 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 and 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, Stoch. Process Appl., 130 (2020), 79–107.

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

[8] L. Campi and M. Fischer, N-player games and mean-field games with absorption, Ann. Appl. Probab., 28 (2016), 2188–2242.

[9] I. Csiszár and J. Körner, Information Theory: Coding Theorems for Discrete Memory-less Systems, Academic Press, New York, 1981.

[10] L. Desvilletes and 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. Desvilletes and 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 and 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 and 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 and S. Méloard, Propagation of chaos and fluctuations for a moderate model with smooth initial data, Ann. Inst. H. Poincaré Probab. Statist., 34 (1998), 727–766.

[16] N. V. Krylov and M. Röckner, Strong solutions of stochastic equations with singular time dependent drift, Probab. Theory Related Fields, 131 (2005), 154–196.

[17] D. Lacker, On a strong form of propagation of chaos for McKean-Vlasov equations, Electron. Commun. Probab., 23 (2018), 11 pp.

[18] Y. S. Mishura and A. Y. Veretennikov, Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations, preprint, arXiv:1603.02212.

[19] M. S. Pinsker, Information and Information Stability of Random Variables and Processes, Holden-Day, San Francisco, 1964.

[20] M. Röckner and X. Zhang, Well-posedness of distribution dependent SDEs with singular drifts, preprint, arXiv:1809.02216.

[21] A.-S. Sznitman, Topics in propagation of chaos, in École d’Été de Probabilités de Saint-Flour XIX-1989, Lecture Notes in Mathematics, 1681, Springer, Berlin, 1991, 165–251.

[22] 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.

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

[24] L. Xie and X. Zhang, Ergodicity of stochastic differential equations with jumps and singular coefficients, Ann. Inst. H. Poincaré Probab. Statist., 56 (2020), 175–229.

[25] C. G. Yuan and S.-Q. Zhang, A study on Zvonkin’s transformation for stochastic differential equations with singular drift and related applications, preprint, arXiv:1910.05903.

[26] P. Xia, L. Xie, X. Zhang and G. Zhao, $L^p(L^p)$-theory of stochastic differential equations, Stoch. Process Appl., 130 (2020), 5188–5211.
[27] X. Zhang, Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients, Electron. J. Probab., 16 (2011), 1096–1116.

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

Received March 2020; revised July 2020.

E-mail address: xinghuang@tju.edu.cn
E-mail address: wangfy@tju.edu.cn