Bernoulli **27**(2), 2021, 1131–1158
https://doi.org/10.3150/20-BEJ1268

Well-posedness of distribution dependent SDEs with singular drifts

MICHAEL RÖCKNER[1,2] and XICHENG ZHANG[3]

[1]*Fakultät für Mathematik, Universität Bielefeld, 33615, Bielefeld, Germany*
[2]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences (CAS), Beijing, 100190, P.R. China.*
E-mail: roeckner@math.uni-bielefeld.de
[3]*School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R. China.*
E-mail: XichengZhang@gmail.com

Consider the following distribution dependent SDE:

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

where μ_{X_t} stands for the distribution of X_t. In this paper for non-degenerate σ, we show the strong well-posedness of the above SDE under some integrability assumptions in the spatial variable and Lipschitz continuity in μ about b and σ. In particular, we extend the results of Krylov–Röckner (*Probab. Theory Related Fields* **131** (2005) 154–196) to the distribution dependent case.

Keywords: Distribution dependent SDEs; Zvonkin's transformation; singular drifts; superposition principle; McKean–Vlasov system

1. Introduction

Let $\mathcal{P}(\mathbb{R}^d)$ be the space of all probability measures over $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, which is endowed with the weak convergence topology. Consider the following distribution dependent stochastic differential equation (abbreviated as DDSDEs):

$$\mathrm{d}X_t = b_t(X_t, \mu_{X_t})\,\mathrm{d}t + \sigma_t(X_t, \mu_{X_t})\,\mathrm{d}W_t, \tag{1}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d$ are two Borel measurable functions, W is a d-dimensional standard Brownian motion on some filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$, and $\mu_{X_t} := \mathbf{P} \circ X_t^{-1}$ is the time marginal of X_t at time t. By Itô's formula, it is easy to see that μ_{X_t} satisfies the following non-linear Fokker–Planck equation (abbreviated as FPE) in the distributional sense:

$$\partial_t \mu_{X_t} = \big(\mathscr{L}_t^{\sigma^X}\big)^* \mu_{X_t} + \mathrm{div}\big(b_t^X \mu_{X_t}\big), \tag{2}$$

where $\sigma_t^X(x) := \sigma_t(x, \mu_{X_t})$, $b_t^X(x) := b_t(x, \mu_{X_t})$, and $(\mathscr{L}_t^{\sigma^X})^*$ is the adjoint operator of the following second order partial differential operator

$$\mathscr{L}_t^{\sigma^X} f(x) := \frac{1}{2} \sum_{i,j,k=1}^{d} \big(\sigma_t^{ik}\sigma_t^{jk}\big)(x, \mu_{X_t})\partial_i\partial_j f(x). \tag{3}$$

We note that if

$$\sigma_t^X(x) = \int_{\mathbb{R}^d} \sigma_t(x, y)\mu_{X_t}(\mathrm{d}y), \qquad b_t^X(x) = \int_{\mathbb{R}^d} b_t(x, y)\mu_{X_t}(\mathrm{d}y),$$

1350-7265

then DDSDE (1) is also called mean-field SDE or McKean–Vlasov SDE in the literature, which naturally appears in the studies of interacting particle systems and mean-field games (see [3,5,13,20,24,27], in particular, [4] and references therein).

Up to now, there are numerous papers devoted to the study of this type of nonlinear FPEs and DDSDE (1). In [9], Funaki showed the existence of martingale solutions for (1) under broad conditions of Lyapunov's type and also the uniqueness under global Lipschitz assumptions. His method is based on a suitable time discretization. Thus, the well-posedness of FPE (2) is also obtained. More recently, under some one-side Lipschitz assumptions, Wang [28] showed the strong well-posedness and some functional inequalities to DDSDE (1). In [10], Hammersley, Siska and Szpruch proved the existence of weak solutions to SDE (1) on a domain $D \subset \mathbb{R}^d$ with continuous and unbounded coefficients under Lyapunov-type conditions. Moreover, uniqueness is also obtained under some functional Lyapunov conditions. Notice that all the above results require the continuity of coefficients. In [7], Chiang obtained the existence of weak solutions for time-independent SDE (1) with drifts that have some discontinuities. When the diffusion matrix is uniformly non-degenerate and b, σ are only measurable and of at most linear growth, by using the classical Krylov estimates, Mishura and Veretennikov [21] showed the existence of weak solutions. The uniqueness is also proved when σ does not depend on μ and is Lipschitz continuous in x and b is Lipschitz continuous with respect to μ with Lipschitz constant linearly depending on x (see also [16]). It should be noted that by Schauder's fixed point theorem and Girsanov's theorem, Li and Min [17] also obtained the existence and uniqueness of weak solutions when b is bounded measurable and σ is nondegenerate and Lipschitz continuous. On the other hand, by a purely analytic argument, Manita and Shaposhnikov [19] and Manita, Romanov and Shaposhnikov [18] showed the existence and uniqueness of solutions to the nonlinear FPE (2) under quite general assumptions. As observed in [1], by a result of Trevisan [25] (see Theorem 5.1 below), one in fact can obtain the well-posedness of DDSDE (1) from [19] and [18]. In [1], a technique is developed to prove weak existence of solutions to (1) by first solving (2) which works also for coefficients whose dependence on μ_{X_t} is of "Nemytskii-type", that is, are not continuous in μ_{X_t} in the weak topology.

In this work we are interested in extending Krylov–Röckner's result [14] to the singular distribution dependent case, that is not covered by all of the above results. More precisely, we want to show the well-posedness of the following DDSDE:

$$\mathrm{d}X_t = \left(\int_{\mathbb{R}^d} b_t(X_t, y)\mu_{X_t}(\mathrm{d}y)\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \tag{4}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is a Borel measurable function and satisfies

($\mathbf{H}^b$) $|b_t(x, y)| \le h_t(x - y)$ for some $h \in L^q_{\mathrm{loc}}(\mathbb{R}_+; \widetilde{L}^p(\mathbb{R}^d))$, where $p, q \in (2, \infty)$ satisfy $\frac{d}{p} + \frac{2}{q} < 1$, and $\widetilde{L}^p(\mathbb{R}^d)$ is the localized L^p-space defined by (13) below.

Here the advantage of using the localized space $\widetilde{L}^p(\mathbb{R}^d)$ is that for any $1 \le p \le p' \le \infty$,

$$L^\infty(\mathbb{R}^d) + L^{p'}(\mathbb{R}^d) \subset \widetilde{L}^{p'}(\mathbb{R}^d) \subset \widetilde{L}^p(\mathbb{R}^d) \overset{p>d}{\subset} \mathbb{K}_{d-1},$$

where $\mathbb{K}_{d-1}$ is the usual Kato's class defined by

$$\mathbb{K}_{d-1} := \left\{ f : \lim_{\varepsilon \to 0} \sup_{x \in \mathbb{R}^d} \int_{|x-y| \le \varepsilon} |x - y|^{1-d} f(y)\,\mathrm{d}y = 0 \right\}.$$

We note that the above DDSDE is not covered by Huang and Wang's recent results [11] since $\mu \mapsto \int_{\mathbb{R}^d} b_t(x,y)\mu(\mathrm{d}y)$ is not weakly continuous. In fact, if we let

$$B_t(x,\mu) := \int_{\mathbb{R}^d} b_t(x,y)\mu(\mathrm{d}y) =: \mu\big(b_t(x,\cdot)\big), \quad \mu \in \mathcal{P}\big(\mathbb{R}^d\big), \tag{5}$$

then by $|b_t(x,y)| \le h_t(x-y)$, we only have

$$\big\|\big| B_t(\cdot,\mu) - B_t\big(\cdot,\mu'\big)\big\|\big|_p \le \|\!| h_t \|\!|_p \big\|\mu - \mu'\big\|_{\mathrm{TV}}, \tag{6}$$

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation distance, and $\|\!|\cdot\|\!|_p$ is defined by (13) below.

Throughout this paper, we assume $d \ge 2$. One of the main results of this paper is stated as follows (but see also Section 4 for corresponding results when the diffusion matrix σ is non-degenerate, but not constant).

Theorem 1.1. *Under* ($\mathbf{H}^b$), *for any $\beta > 2$ and initial random variable X_0 with finite β-order moment, there is a unique strong solution to SDE* (4). *Moreover, the following assertions hold*:

(i) *The time marginal law μ_t of X_t uniquely solves the following nonlinear FPE in the distributional sense*:

$$\partial_t \mu_t = \Delta\mu_t + \mathrm{div}\big(\mu_t\big(b_t(x,\cdot)\big)\mu_t\big), \qquad \lim_{t\downarrow 0}\mu_t(\mathrm{d}y) = \mathbf{P}\circ X_0^{-1}(\mathrm{d}y) \tag{7}$$

in the class of all measures such that $t \mapsto \mu_t$ is weakly continuous and

$$\int_0^T \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \big|b_t(x,y)\big|\mu_t(\mathrm{d}y)\mu_t(\mathrm{d}x)\,\mathrm{d}t < \infty, \quad \forall T > 0.$$

(ii) *$\mu_t(\mathrm{d}y) = \rho_t^X(y)\,\mathrm{d}y$ and $(t,y) \mapsto \rho_t^X(y)$ is continuous on $(0,\infty)\times\mathbb{R}^d$ and satisfies the following two-sided estimate*: *for any $T > 0$, there are constants $\gamma_0, c_0 \ge 1$ such that for all $t \in (0,T]$ and $y \in \mathbb{R}^d$*,

$$c_0^{-1} P_{t/\gamma_0}\mu_0(y) \le \rho_t^X(y) \le c_0 P_{\gamma_0 t}\mu_0(y),$$

where $P_t\mu_0(y) := (2\pi t)^{-d/2}\int_{\mathbb{R}^d} \mathrm{e}^{-|x-y|^2/(2t)}\mu_0(\mathrm{d}x)$ is the Gaussian heat semigroup.

(iii) *If $\mathrm{div}_x b = 0$, then for each $t > 0$, $\rho_t^X(\cdot) \in C^1(\mathbb{R}^d)$ and we have the following gradient estimate*: *for any $T > 0$, there are constants $\gamma_1, c_1 \ge 1$ such that for all $t \in (0,T]$ and $y \in \mathbb{R}^d$*,

$$\big|\nabla\rho_t^X(y)\big| \le c_1 t^{-1/2} P_{\gamma_1 t}\mu_0(y).$$

Example 1. Let $b_t(x,y) := a_t(x,y)/|x-y|^\alpha$ for some $\alpha \in [1,2)$, where $a_t(x,y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ satisfies that for some $\kappa > 0$,

$$\big|a_t(x,y)\big| \le \kappa|x-y|.$$

Then it is easy to see that b satisfies ($\mathbf{H}^b$) for some $p > d$ and $q = \infty$.

Remark 1. Here an open question is to show the following propagation of chaos (see [24]): Given $N \in \mathbb{N}$, let $X^{N,j}$, $j = 1,\dots,N$ solve the following SDEs

$$\mathrm{d}X_t^{N,j} = \frac{1}{N}\sum_{i=1}^N b_t\big(X_t^{N,j}, X_t^{N,i}\big)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t^j, \quad j = 1,\dots,N,$$

where $W^j_\cdot$, $j = 1, \ldots, N$ are N-independent d-dimensional Brownian motions. Let X be the unique solution of SDE (4) in Theorem 1.1. Is it possible to show that

$$X^{N,1}_\cdot \to X_\cdot \quad \text{in distribution as } N \to \infty?$$

It should be noticed that when b is bounded measurable, the above propagation of chaos has been shown by Lacker in [16]. However, for singular drift b, it seems to be open.

To show the existence of a solution to DDSDE (4), by the well-known result for bounded measurable drift b obtained in [21] (see also [16,17] and [31]), for each $n \in \mathbb{N}$, there is a solution to the following distribution dependent SDE:

$$\mathrm{d}X^n_t = \left(\int_{\mathbb{R}^d} b^n_t\big(X^n_t, y\big)\mu_{X^n_t}(\mathrm{d}y)\right)\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \qquad X^n_0 = X_0, \tag{8}$$

where $b^n_t(x, y) := (-n) \vee b_t(x, y) \wedge n$. By the well-known results in [29], one can show the following uniform Krylov estimate: For any $p_1, q_1 \in (1, \infty)$ with $\frac{d}{p_1} + \frac{2}{q_1} < 2$ and $T > 0$, there is a constant $C > 0$ such that for any $f \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$,

$$\sup_n \mathbf{E}\left(\int_0^T f_t\big(X^n_t\big)\,\mathrm{d}t\right) \le C_T |||f|||_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}. \tag{9}$$

By this estimate and Zvonkin's technique, we can further show the tightness of $X^n_\cdot$ in the space of continuous functions. However, since b is allowed to be singular, it is not obvious by taking the limit $n \to \infty$ to obtain the existence of a solution. Indeed, one needs the following Krylov estimate: for suitable $p_0, q_0 \in (1, \infty)$ and any $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$,

$$\sup_n \mathbf{E}\left(\int_0^t f_s\big(X^n_s, \tilde{X}^n_s\big)\,\mathrm{d}s\right) \le |||f|||_{\widetilde{\mathbb{L}}^{p_0}_{q_0}(T)},$$

where $\tilde{X}^n_\cdot$ is an independent copy of X^n. When b is bounded measurable, such an estimate is easy to get by considering $(X^n, \tilde{X}^n)$ as an $\mathbb{R}^{2d}$-dimensional Itô process and using the classical Krylov estimates (see [21]). While for singular b, such simple observation fails in order to obtain best integrability index p. We overcome this difficulty by a simple duality argument (see Lemma 2.6 below). Moreover, concerning the uniqueness, under assumption (6), we shall employ Girsanov's transformation as usual.

This paper is organized as follows: In Section 2, we prepare some well-known results and tools for later use. In Section 3, we show the existence of weak and strong solutions to DDSDE (1) when the drift satisfies $(\mathbf{H}^b)$, and the diffusion coefficient is uniformly nondegenerate and bounded Hölder continuous. In Section 4, we prove the uniqueness of weak and strong solutions to (1) in two cases: the coefficients b and σ are Lipschitz continuous in the third variable with respect to the Wasserstein metric; drift b is Lipschitz continuous in the third variable with respect to the total variation distance and the diffusion coefficient does not depend on the distribution. In Section 5, we present some applications to nonlinear FPE (2) and prove Theorem 1.1.

Finally, we collect some frequently used notations and conventions for later use.

- For $\theta > 0$, $\mathcal{P}_\theta(\mathbb{R}^d) := \{\mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^\theta \mu(\mathrm{d}x) < \infty\}$.
- For $R > 0$, set $B_R := \{x \in \mathbb{R}^d : |x| < R\}$.
- For a function $f : \mathbb{R}^d \to \mathbb{R}$, $\mathcal{M}_R f(x) := \sup_{r \in (0,R)} \frac{1}{|B_r|} \int_{B_r} |f|(x + y)\,\mathrm{d}y$.
- Let $\mathbf{S}_{\mathrm{toch}}$ be the set of all measurable stochastic processes on $(\Omega, \mathscr{F}, \mathbf{P})$ that are stochastically continuous.

- Let $b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ be a measurable vector field. For $X \in \mathbf{S}_{\text{toch}}$, define

$$b_t^X(x) := b_t(x, \mu_{X_t}), \quad \mu_{X_t} := \mathbf{P} \circ X_t^{-1}. \tag{10}$$

- For a signed measure μ, we denote by $\|\mu\|_{\text{TV}} := \sup_{\|f\|_\infty \le 1} |\mu(f)|$ the total variation of μ.
- For $j = 1, 2$, we introduce the index set $\mathscr{I}_j$ as following:

$$\mathscr{I}_j := \left\{(p, q) \in (1, \infty) : \frac{d}{p} + \frac{2}{q} < j\right\}. \tag{11}$$

- For a matrix σ, we use $\|\sigma\|_{\text{HS}}$ to denote the Hilbert–Schmidt norm of σ.
- We use $A \lesssim B$ (resp. $\asymp$) to denote $A \le CB$ (resp. $C^{-1}B \le A \le CB$) for some unimportant constant $C \ge 1$, whose dependence on the parameters can be traced from the context.

2. Preliminaries

In this section, we recall some well-known results. We first introduce the following spaces and notations for later use. For $(\alpha, p) \in \mathbb{R}_+ \times (1, \infty)$, the usual Bessel potential space $H^{\alpha,p}$ is defined by

$$H^{\alpha,p} := \left\{f \in L^1_{\text{loc}}(\mathbb{R}^d) : \|f\|_{\alpha,p} := \left\|(\mathbb{I} - \Delta)^{\alpha/2} f\right\|_p < \infty\right\},$$

where $\|\cdot\|_p$ is the usual L^p-norm, and $(\mathbb{I} - \Delta)^{\alpha/2} f$ is defined by Fourier transform

$$(\mathbb{I} - \Delta)^{\alpha/2} f := \mathcal{F}^{-1}\left(\left(1 + |\cdot|^2\right)^{\alpha/2} \mathcal{F} f\right).$$

Notice that for $n \in \mathbb{N}$, an equivalent norm in $H^{n,p}$ is given by

$$\|f\|_{n,p} = \|f\|_p + \left\|\nabla^n f\right\|_p.$$

For $T > S \ge 0$, $p, q \in (1, \infty)$ and $\alpha \in \mathbb{R}_+$, we introduce space-time function spaces

$$\mathbb{L}^p_q(S, T) := L^q\left([S, T]; L^p\right), \qquad \mathbb{H}^{\alpha,p}_q(S, T) := L^q\left([S, T]; H^{\alpha,p}\right).$$

Let $\chi \in C_c^\infty(\mathbb{R}^d)$ be a smooth function with $\chi(x) = 1$ for $|x| \le 1$ and $\chi(x) = 0$ for $|x| > 2$. For $r > 0$ and $z \in \mathbb{R}^d$, define

$$\chi_r^z(x) := \chi\left((x - z)/r\right). \tag{12}$$

Fix $r > 0$. We introduce the following localized $H^{\alpha,p}$-space:

$$\widetilde{H}^{\alpha,p} := \left\{f \in H^{\alpha,p}_{\text{loc}}\left(\mathbb{R}^d\right), |||f|||_{\alpha,p} := \sup_z \left\|f \chi_r^z\right\|_{\alpha,p} < \infty\right\}, \tag{13}$$

and the localized space-time function space $\widetilde{\mathbb{H}}^{\alpha,p}_q(S, T)$ with norm

$$|||f|||_{\widetilde{\mathbb{H}}^{\alpha,p}_q(S,T)} := \sup_{z \in \mathbb{R}^d} \left\|\chi_r^z f\right\|_{\mathbb{H}^{\alpha,p}_q(S,T)} < \infty. \tag{14}$$

For simplicity we shall write

$$\widetilde{\mathbb{H}}^{\alpha,p}_q(T) := \widetilde{\mathbb{H}}^{\alpha,p}_q(0, T), \qquad \widetilde{\mathbb{L}}^p_q(T) := \widetilde{\mathbb{H}}^{0,p}_q(0, T),$$

and

$$\widetilde{\mathbb{H}}_q^{\alpha,p} := \bigcap_{T>0} \widetilde{\mathbb{H}}_q^{\alpha,p}(T), \qquad \widetilde{\mathbb{L}}_q^p := \bigcap_{T>0} \widetilde{\mathbb{L}}_q^p(T).$$

The following lemma list some easy properties of $\widetilde{\mathbb{H}}_q^{\alpha,p}$ (see [34] and [29]).

Proposition 2.1. *Let $p, q \in (1,\infty)$, $\alpha \in \mathbb{R}_+$ and $T > 0$.*

(i) *For $r \neq r' > 0$, there is a $C = C(d, \alpha, r, r', p, q) \geq 1$ such that*

$$C^{-1} \sup_z \|f\chi_{r'}^z\|_{\mathbb{H}_q^{\alpha,p}(T)} \leq \sup_z \|f\chi_r^z\|_{\mathbb{H}_q^{\alpha,p}(T)} \leq C \sup_z \|f\chi_{r'}^z\|_{\mathbb{H}_q^{\alpha,p}(T)}. \tag{15}$$

In other words, the definition of $\widetilde{\mathbb{H}}_q^{\alpha,p}$ does not depend on the choice of r.

(ii) *Let $\alpha > 0$, $p, q \in [1,\infty)$ and $p' \in [p, \frac{pd}{d-p\alpha}\mathbf{1}_{p\alpha<d} + \infty \cdot \mathbf{1}_{p\alpha>d}]$. It holds that for some $C = C(d, \alpha, p, p') > 0$,*

$$|||f|||_{\widetilde{\mathbb{L}}_q^{p'}(T)} \leq C |||f|||_{\widetilde{\mathbb{H}}_q^{\alpha,p}(T)}. \tag{16}$$

(iii) *For any $k \in \mathbb{N}$, there is a constant $C = C(d, k, \alpha, p, q) \geq 1$ such that*

$$C^{-1} |||f|||_{\widetilde{\mathbb{H}}_q^{\alpha+k,p}(T)} \leq |||f|||_{\widetilde{\mathbb{H}}_q^{\alpha,p}(T)} + |||\nabla^k f|||_{\widetilde{\mathbb{H}}_q^{\alpha,p}(T)} \leq C |||f|||_{\widetilde{\mathbb{H}}_q^{\alpha+k,p}(T)}.$$

(iv) *Let $(\rho_\varepsilon)_{\varepsilon\in(0,1)}$ be a family of mollifiers in $\mathbb{R}^d$ and $f_\varepsilon(t,x) := f(t,\cdot) * \rho_\varepsilon(x)$. For any $f \in \widetilde{\mathbb{H}}_q^{\alpha,p}$, it holds that $f_\varepsilon \in L^q_{\mathrm{loc}}(\mathbb{R}; C_b^\infty(\mathbb{R}^d))$ and for some $C = C(d, \alpha, p, q) > 0$,*

$$|||f_\varepsilon|||_{\widetilde{\mathbb{H}}_q^{\alpha,p}(T)} \leq C |||f|||_{\widetilde{\mathbb{H}}_q^{\alpha,p}(T)}, \quad \forall \varepsilon \in (0,1), \tag{17}$$

and for any $\varphi \in C_c^\infty(\mathbb{R}^d)$,

$$\lim_{\varepsilon\to 0} \|(f_\varepsilon - f)\varphi\|_{\mathbb{H}_q^{\alpha,p}(T)} = 0. \tag{18}$$

(v) *Let $\mathbb{Z}^d$ be the d-dimensional integer lattice. For $r := \frac{p}{p-1}$ and $s := \frac{q}{q-1}$, define*

$$|||f|||'_{\widetilde{\mathbb{L}}_q^p(T)} := \sup_{z\in\mathbb{Z}^d} \|\mathbf{1}_{Q_z} f\|_{\mathbb{L}_q^p(T)}, \qquad |||g|||^*_{\widetilde{\mathbb{L}}_s^r(T)} := \sum_{z\in\mathbb{Z}^d} \|\mathbf{1}_{Q_z} g\|_{\mathbb{L}_s^r(T)},$$

where

$$Q_z := \Pi_{i=1}^d (z_i, z_i + 1], \quad z = (z_1, \ldots, z_d) \in \mathbb{Z}^d. \tag{19}$$

Then we have

$$|||f|||_{\widetilde{\mathbb{L}}_q^p(T)} \asymp |||f|||'_{\widetilde{\mathbb{L}}_q^p(T)} = \sup_{|||g|||^*_{\widetilde{\mathbb{L}}_s^r(T)} \leq 1} \left| \int_0^T \int_{\mathbb{R}^d} f_t(x) g_t(x)\,\mathrm{d}x\,\mathrm{d}t \right| \tag{20}$$

and

$$|||g|||^*_{\widetilde{\mathbb{L}}_s^r(T)} = \sup_{|||f|||'_{\widetilde{\mathbb{L}}_q^p(T)} \leq 1} \left| \int_0^T \int_{\mathbb{R}^d} f_t(x) g_t(x)\,\mathrm{d}x\,\mathrm{d}t \right|. \tag{21}$$

Proof. The first four conclusions can be found in [34], Proposition 4.1. We only prove (v). The equivalence between $|||f|||_{\widetilde{\mathbb{L}}_q^p(T)}$ and $|||f|||'_{\widetilde{\mathbb{L}}_q^p(T)}$ is obvious by definition. Concerning the others, we note that by Hölder's inequality,

$$\int_0^T\int_{\mathbb{R}^d} f_t(x)g_t(x)\,\mathrm{d}x\,\mathrm{d}t = \sum_{z\in\mathbb{Z}^d}\int_0^T\int_{\mathbb{R}^d}\mathbf{1}_{Q_z}(x)f_t(x)g_t(x)\,\mathrm{d}x\,\mathrm{d}t$$
$$\leq \sum_{z\in\mathbb{Z}^d}\|\mathbf{1}_{Q_z}f\|_{\mathbb{L}_q^p(T)}\|\mathbf{1}_{Q_z}g\|_{\mathbb{L}_s^r(T)} \leq |||f|||'_{\widetilde{\mathbb{L}}_q^p(T)}|||g|||^*_{\widetilde{\mathbb{L}}_s^r(T)}. \tag{22}$$

On the other hand, assume that z_n is a sequence in $\mathbb{Z}^d$ so that for $Q_n := Q_{z_n}$,

$$\lim_{n\to\infty}\|\mathbf{1}_{Q_n}f\|_{\mathbb{L}_q^p(T)} = |||f|||'_{\widetilde{\mathbb{L}}_q^p(T)}. \tag{23}$$

If we take

$$g_t(x) := \frac{\mathbf{1}_{Q_n}(x)|f_t(x)|^{p-1}}{\|\mathbf{1}_{Q_n}f_t\|_p^{p-q}}\left(\int_0^T\|\mathbf{1}_{Q_n}f_t\|_p^q\,\mathrm{d}t\right)^{1/q-1}$$

with the convention $0/0=0$, then by easy calculations, we have $|||g|||^*_{\widetilde{\mathbb{L}}_s^r(T)}=1$ and

$$\int_0^T\int_{\mathbb{R}^d} f_t(x)g_t(x)\,\mathrm{d}x\,\mathrm{d}t = \left(\int_0^T\|\mathbf{1}_{Q_n}f_t\|_p^q\,\mathrm{d}t\right)^{1/q} = \|\mathbf{1}_{Q_n}f\|_{\mathbb{L}_q^p(T)},$$

which together with (22) and (23) yields (20). Similarly, if we take

$$f_t(x) := \sum_{z\in\mathbb{Z}^d}\frac{\mathbf{1}_{Q_z}(x)|g_t(x)|^{r-1}}{\|\mathbf{1}_{Q_z}g_t\|_r^{r-s}}\cdot\left(\int_0^T\|\mathbf{1}_{Q_z}g_t\|_r^s\,\mathrm{d}t\right)^{1/s-1},$$

then $|||f|||'_{\widetilde{\mathbb{L}}_q^p(T)}=1$ and

$$\int_0^T\int_{\mathbb{R}^d} f_t(x)g_t(x)\,\mathrm{d}x\,\mathrm{d}t = \sum_{z\in\mathbb{Z}^d}\left(\int_0^T\|\mathbf{1}_{Q_z}g_t\|_r^s\,\mathrm{d}t\right)^{1/s} = |||g|||^*_{\widetilde{\mathbb{L}}_s^r(T)},$$

which together with (22) yields (21). □

We now recall the following result about $L^q(L^p)$-solvability of PDE (see [29]).

Theorem 2.2. *Let $(p,q)\in\mathscr{I}_1$ (see (11)) and $T>0$. Assume that $\sigma_t(x,\mu)=\sigma_t(x)$ and $b_t(x,\mu)=b_t(x)$ are independent of μ, and satisfy that for some $c_0\geq 1$, $\gamma\in(0,1]$ and for all $t\geq 0$, $x,y,\xi\in\mathbb{R}^d$,*

$$c_0^{-1}|\xi|\leq\left|\sigma_t^*(x)\xi\right|\leq c_0|\xi|, \qquad \left\|\sigma_t(x)-\sigma_t(y)\right\|_{\mathrm{HS}}\leq c_0|x-y|^\gamma, \tag{24}$$

and $|||b|||_{\widetilde{\mathbb{L}}_q^p(T)}\leq\kappa_0$ for some $\kappa_0>0$, Then for any $\lambda\geq 1$ and $f\in\widetilde{\mathbb{L}}_q^p(T)$, there exists a unique solution $u\in\widetilde{\mathbb{H}}_q^{2,p}(T)$ to the following backward parabolic equation:

$$\partial_t u+\left(\mathscr{L}_t^\sigma-\lambda\right)u+b\cdot\nabla u=f, \qquad u(T,x)=0. \tag{25}$$

Moreover, letting $\Theta:=(\gamma,c_0,d,p,q,\kappa_0,T)$, we have the following:

(i) *For any* $\alpha \in [0, 2 - \frac{2}{q})$, *there is a* $c_1 = c_1(\alpha, \Theta) > 0$ *such that for all* $\lambda \geq 1$,

$$\lambda^{1-\frac{\alpha}{2}-\frac{1}{q}} |||u|||_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} + |||u|||_{\widetilde{\mathbb{H}}^{2,p}_q(T)} \leq c_1 |||f|||_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{26}$$

(ii) *Let* (σ', b', f') *be another set of coefficients satisfying the same assumptions as* (σ, b, f) *with the same parameters* (γ, c_0, κ_0). *Let* u' *be the solution of* (25) *corresponding to* (σ', b', f'). *For any* $\alpha \in [0, 2 - \frac{2}{q})$, *there is a constant* $c_2 = c_2(\alpha, \Theta) > 0$ *such that for all* $\lambda \geq 1$,

$$\begin{aligned}\lambda^{1-\frac{\alpha}{2}-\frac{1}{q}} \big|\big|\big|u - u'\big|\big|\big|_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} &\leq c_2 \big|\big|\big|f - f'\big|\big|\big|_{\widetilde{\mathbb{L}}^p_q(T)} \\ &\quad + c_2 |||f|||_{\widetilde{\mathbb{L}}^p_q(T)} \big(\big\|\sigma - \sigma'\big\|_{\mathbb{L}^\infty(T)} + \big|\big|\big|b - b'\big|\big|\big|_{\widetilde{\mathbb{L}}^p_q(T)}\big).\end{aligned} \tag{27}$$

Proof. The existence and uniqueness of $u \in \widetilde{\mathbb{H}}^{2,p}_q(T)$ as well as the first conclusion are proved in [29], Theorem 3.1. We only show (ii). Let $w = u' - u$. Then

$$\partial_t w + \big(\mathscr{L}^{\sigma'}_t - \lambda\big) w + b' \cdot \nabla w = \big(\mathscr{L}^{\sigma}_t - \mathscr{L}^{\sigma'}_t\big) u + \big(b - b'\big) \cdot \nabla u + f' - f.$$

By (26) and Hölder's inequality, we have

$$\begin{aligned}\lambda^{1-\frac{\alpha}{2}-\frac{1}{q}} |||w|||_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} &\lesssim \big|\big|\big|\big(\mathscr{L}^{\sigma}_t - \mathscr{L}^{\sigma'}_t\big) u + \big(b - b'\big) \cdot \nabla u + f' - f\big|\big|\big|_{\widetilde{\mathbb{L}}^p_q(T)} \\ &\lesssim \big\|\sigma' - \sigma\big\|_{\mathbb{L}^\infty(T)} \big|\big|\big|\nabla^2 u\big|\big|\big|_{\widetilde{\mathbb{L}}^p_q(T)} + \big|\big|\big|b' - b\big|\big|\big|_{\widetilde{\mathbb{L}}^p_q(T)} \cdot \|\nabla u\|_{\mathbb{L}^\infty(T)} + \big|\big|\big|f' - f\big|\big|\big|_{\widetilde{\mathbb{L}}^p_q(T)}.\end{aligned}$$

Estimate (27) now follows by Sobolev's embedding (16) due to $\frac{d}{p} + \frac{2}{q} < 1$ and (26). □

Remark 2. It should be noted that if b is bounded measurable, then the assertions in Theorem 2.2 hold for all $p, q \in (1, \infty)$.

The following stochastic Gronwall inequality for continuous martingales was proved by Scheutzow [22], and for general discontinuous martingales in [30].

Lemma 2.3 (Stochastic Gronwall's inequality). *Let* $\xi(t)$ *and* $\eta(t)$ *be two nonnegative càdlàg* $\mathscr{F}_t$-*adapted processes,* A_t *a continuous nondecreasing* $\mathscr{F}_t$-*adapted process with* $A_0 = 0$, M_t *a local martingale with* $M_0 = 0$. *Suppose that*

$$\xi(t) \leq \eta(t) + \int_0^t \xi(s)\, \mathrm{d}A_s + M_t, \quad \forall t \geq 0. \tag{28}$$

Then for any $0 < q < p < 1$ *and* $\tau > 0$, *we have*

$$\big[\mathbf{E}\big(\xi(\tau)^*\big)^q\big]^{1/q} \leq \left(\frac{p}{p-q}\right)^{1/q} \big(\mathbf{E}\mathrm{e}^{pA_\tau/(1-p)}\big)^{(1-p)/p} \mathbf{E}\big(\eta(\tau)^*\big), \tag{29}$$

where $\xi(t)^* := \sup_{s\in[0,t]} \xi(s)$.

We also recall the following result about maximal functions (see [29], Lemma 2.1).

Lemma 2.4.

(i) *For any $R > 0$, there exists a constant $C = C(d, R)$ such that for any $f \in L^\infty(\mathbb{R}^d)$ with $\nabla f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ and Lebesgue-almost all $x, y \in \mathbb{R}^d$,*

$$\big|f(x) - f(y)\big| \le C|x - y|\big(\mathcal{M}_R|\nabla f|(x) + \mathcal{M}_R|\nabla f|(y) + \|f\|_\infty\big), \tag{30}$$

where $\mathcal{M}_R$ is defined at the end of the introduction.

(ii) *For any $p > 1$ and $R > 0$, there is a constant $C = C(R, d, p)$ such that for any $T > 0$ and all $f \in \widetilde{\mathbb{L}}^p_q(T)$,*

$$|||\mathcal{M}_R f|||_{\widetilde{\mathbb{L}}^p_q(T)} \le C|||f|||_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{31}$$

We introduce the following notion about Krylov's estimates.

Definition 2.5. Let $p, q \in (1, \infty)$ and $T, \kappa > 0$. We say a stochastic process $X \in \mathbf{S}_{\mathrm{toch}}$ satisfies Krylov's estimate with index p, q and constant κ if for any $f \in \widetilde{\mathbb{L}}^p_q(T)$,

$$\mathbf{E}\bigg(\int_0^T f_t(X_t)\,\mathrm{d}t\bigg) \le \kappa|||f|||_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{32}$$

The set of all such X will be denoted by $\mathbf{K}^{p,q}_{T,\kappa}$.

For a space-time function $f_t(x, y) : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $p_1, p_2, q_0 \in [1, \infty]$, we also introduce the norm

$$|||f|||_{\widetilde{\mathbb{L}}^{p_1,p_2}_{q_0}(T)} := \sup_{z,z' \in \mathbb{R}^d}\bigg(\int_0^T \bigg(\int_{Q_{z'}} \big\|\mathbf{1}_{Q_z} f_t(\cdot, y)\big\|^{p_2}_{p_1}\,\mathrm{d}y\bigg)^{\frac{q_0}{p_2}}\bigg)^{\frac{1}{q_0}}.$$

The following lemma is an easy consequence of Proposition 2.1(v).

Lemma 2.6. *Let $p_1, p_2, q_0, q_1, q_2 \in (1, \infty)$ with $\frac{1}{q_1} + \frac{1}{q_2} = 1 + \frac{1}{q_0}$ and $T, \kappa_1, \kappa_2 > 0$. Let $X \in \mathbf{K}^{p_1,q_1}_{T,\kappa_1}$ and $Y \in \mathbf{K}^{p_2,q_2}_{T,\kappa_2}$ be two independent processes. Then for any $f_t(x, y) \in \widetilde{\mathbb{L}}^{p_1,p_2}_{q_0}(T)$,*

$$\mathbf{E}\bigg(\int_0^T f_t(X_t, Y_t)\,\mathrm{d}t\bigg) \le \kappa_1\kappa_2|||f|||_{\widetilde{\mathbb{L}}^{p_1,p_2}_{q_0}(T)}. \tag{33}$$

Proof. Let $Z^1 = X$ and $Z^2 = Y$. First of all, by Krylov's estimate (32), for each $i = 1, 2$, there is a function $\rho^{Z^i} \in \mathbb{L}^{r_i}_{s_i}(T)$ with $r_i = \frac{p_i}{p_i - 1}$, $s_i = \frac{q_i}{q_i - 1}$ so that

$$\int_0^T \int_{\mathbb{R}^d} f_t(x)\rho^{Z^i}_t(x)\,\mathrm{d}x\,\mathrm{d}t = \mathbf{E}\bigg(\int_0^T f_t\big(Z^i_t\big)\,\mathrm{d}t\bigg) \le \kappa_i|||f|||_{\widetilde{\mathbb{L}}^{p_i}_{q_i}(T)} \le \kappa_i\|f\|_{\mathbb{L}^{p_i}_{q_i}(T)}.$$

By Proposition 2.1(v), we further have

$$\big|\big|\big|\rho^{Z^i}\big|\big|\big|^*_{\widetilde{\mathbb{L}}^{r_i}_{s_i}(T)} := \sum_{z \in \mathbb{Z}^d} \big\|\mathbf{1}_{Q_z}\rho^{Z^i}\big\|_{\mathbb{L}^{r_i}_{s_i}(T)} \le \kappa_i, \quad i = 1, 2,$$

where Q_z is defined by (19). Now by the independence of X, Y and Hölder's inequality, we have

$$\begin{aligned}\mathbf{E}\left(\int_0^T f_t(X_t,Y_t)\,\mathrm{d}t\right)&=\int_0^T\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f_t(x,y)\rho_t^X(x)\rho_t^Y(y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t\\&=\sum_{z\in\mathbb{Z}^d}\sum_{z'\in\mathbb{Z}^d}\int_0^T\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{1}_{Q_z}(x)\mathbf{1}_{Q_{z'}}(y)f_t(x,y)\rho_t^X(x)\rho_t^Y(y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}t\\&\leq\sum_{z\in\mathbb{Z}^d}\sum_{z'\in\mathbb{Z}^d}\|\mathbf{1}_{Q_z\times Q_{z'}}f\|_{\mathbb{L}^{p_1,p_2}_{q_0}(T)}\left\|\mathbf{1}_{Q_z}\rho^X\right\|_{\mathbb{L}^{r_1}_{s_1}(T)}\left\|\mathbf{1}_{Q_{z'}}\rho^Y\right\|_{\mathbb{L}^{r_2}_{s_2}(T)}\\&\leq\kappa_1\kappa_2\sup_{z,z'\in\mathbb{Z}^d}\|\mathbf{1}_{Q_z\times Q_{z'}}f\|_{\mathbb{L}^{p_1,p_2}_{q_0}(T)}=\kappa_1\kappa_2|||f|||_{\widetilde{\mathbb{L}}^{p_1,p_2}_{q_0}(T)},\end{aligned}$$

which gives (33). The proof is complete. □

Now we prove the following convergence lemmas, which have independent interest and will be crucial for showing the existence of solutions in Section 3.

Lemma 2.7. *Let $X^n,Y^n,X,Y\in\mathbf{S}_{\mathrm{toch}}$ be such that for each $t\geq 0$, X_t^n converges to X_t almost surely and Y_t^n converges to Y_t in distribution. Let $p,q>1$ and $T,\beta,\kappa>0$. Suppose that $X^n\in\mathbf{K}^{p,q}_{T,\kappa}$ for each $n\in\mathbb{N}$, and for some $C_1>0$,*

$$\sup_n\sup_{t\in[0,T]}\mathbf{E}\left|X_t^n\right|^\beta\leq C_1. \tag{34}$$

If for each (t,x), $\mu\mapsto b_t(x,\mu)$ is continuous with respect to the weak convergence topology and for some $\gamma>1$ and $C_2>0$,

$$\sup_{Z\in\mathbf{S}_{\mathrm{toch}}}\left|\left|\left|b^Z\right|\right|\right|_{\mathbb{L}^{\gamma p}_{\gamma q}(T)}\leq C_2, \tag{35}$$

where b^Z is defined by (10), then

$$\lim_{n\to\infty}\mathbf{E}\left(\int_0^T\left|b_t^{Y_n}\left(X_t^n\right)-b_t^Y(X_t)\right|\mathrm{d}t\right)=0. \tag{36}$$

Proof. To prove (36), it suffices to show the following:

$$\lim_{n\to\infty}\mathbf{E}\left(\int_0^T\left|b_t^{Y^n}\left(X_t^n\right)-b_t^Y\left(X_t^n\right)\right|\mathrm{d}t\right)=0, \tag{37}$$

$$\lim_{n\to\infty}\mathbf{E}\left(\int_0^T\left|b_t^{Y}\left(X_t^n\right)-b_t^Y(X_t)\right|\mathrm{d}t\right)=0. \tag{38}$$

We first look at (37). Since $\mu_{Y_t^n}$ weakly converges to μ_{Y_t} for each $t\geq 0$, by the assumption we have

$$b_t^{Y_n}(x)\overset{n\to\infty}{\to}b_t^Y(x),\quad\forall(t,x)\in\mathbb{R}_+\times\mathbb{R}^d. \tag{39}$$

For fixed $R, M > 0$, since $X^n \in \mathbf{K}^{p,q}_{T,\kappa}$ (see (32)), by the definitions we have

$$\begin{aligned}
&\mathbf{E}\left(\int_0^T \mathbf{1}_{B_R}(X^n_t)\big|b^{Y_n}_t(X^n_t) - b^Y_t(X^n_t)\big|\,\mathrm{d}t\right)\\
&\quad\le \kappa \big|\!\big|\!\big|\mathbf{1}_{B_R}\big(b^{Y_n} - b^Y\big)\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^p_q(T)}\\
&\quad\lesssim \big\|\mathbf{1}_{B_R}\big(b^{Y_n} - b^Y\big)\mathbf{1}_{|b^{Y_n}-b^Y|\le M}\big\|_{\mathbb{L}^p_q(T)} + \big\|\mathbf{1}_{B_R}\big(b^{Y_n} - b^Y\big)\mathbf{1}_{|b^{Y_n}-b^Y|> M}\big\|_{\mathbb{L}^p_q(T)}\\
&\quad\le \big\|\mathbf{1}_{B_R}\big(b^{Y_n} - b^Y\big)\mathbf{1}_{|b^{Y_n}-b^Y|\le M}\big\|_{\mathbb{L}^p_q(T)} + \big\|\mathbf{1}_{B_R}\big|b^{Y_n} - b^Y\big|^\gamma\big\|_{\mathbb{L}^p_q(T)}/M^{\gamma-1}.
\end{aligned}$$

By the dominated convergence theorem and (39), the first term converges to zero as $n \to \infty$ for each $M > 0$. By (35), the second term converges to zero uniformly in n as $M \to \infty$. Thus, we obtain that for any $R > 0$,

$$\lim_{n\to\infty} \mathbf{E}\left(\int_0^T \mathbf{1}_{B_R}(X^n_t)\big|b^{Y_n}_t(X^n_t) - b^Y_t(X^n_t)\big|\,\mathrm{d}t\right) = 0. \tag{40}$$

On the other hand, by Hölder and Chebyshev's inequalities and (34), we have

$$\begin{aligned}
&\mathbf{E}\left(\int_0^T \mathbf{1}_{B^c_R}(X^n_t)\big|b^{Y_n}_t(X^n_t) - b^Y_t(X^n_t)\big|\,\mathrm{d}t\right)\\
&\quad\le \int_0^T \mathbf{P}\big(|X^n_t| > R\big)^{\frac{\gamma-1}{\gamma}}\big(\mathbf{E}\big|b^{Y_n}_t(X^n_t) - b^Y_t(X^n_t)\big|^\gamma\big)^{\frac{1}{\gamma}}\,\mathrm{d}t\\
&\quad\le \sup_{t\in[0,T]} \mathbf{P}\big(|X^n_t| > R\big)^{\frac{\gamma-1}{\gamma}} T^{\frac{\gamma-1}{\gamma}}\left(\int_0^T \mathbf{E}\big|b^{Y_n}_t(X^n_t) - b^Y_t(X^n_t)\big|^\gamma\,\mathrm{d}t\right)^{\frac{1}{\gamma}}\\
&\quad\le \left(\frac{C_1 T}{R^\beta}\right)^{\frac{\gamma-1}{\gamma}} \kappa^{\frac{1}{\gamma}} \big|\!\big|\!\big|b^{Y_n} - b^Y\big|\!\big|\!\big|_{\mathbb{L}^{\gamma p}_{\gamma q}(T)} \overset{(35)}{\le} \left(\frac{C_1 T}{R^\beta}\right)^{\frac{\gamma-1}{\gamma}} \kappa^{\frac{1}{\gamma}} \cdot 2C_2.
\end{aligned}$$

Combining this with (40), we obtain (37).

Next, we show (38). Let $b^{Y,\varepsilon}_t(x) := b^Y_t(\cdot) * \varrho_\varepsilon(x)$ be a mollifying approximation of b^Y. By Proposition 2.1(iv) and (34), as above one can derive that

$$\lim_{\varepsilon\to 0} \sup_{n\in\mathbb{N}\cup\{\infty\}} \mathbf{E}\left(\int_0^T \big|b^{Y,\varepsilon}_t(X^n_t) - b^Y_t(X^n_t)\big|\,\mathrm{d}t\right) = 0, \tag{41}$$

where we have used the convention $X^\infty := X$. On the other hand, since by (32),

$$\sup_n \mathbf{E}\left(\int_0^T \big|b^{Y,\varepsilon}_t(X^n_t) - b^{Y,\varepsilon}_t(X_t)\big|^\gamma\,\mathrm{d}t\right) \le C \big|\!\big|\!\big|b^{Y,\varepsilon}\big|\!\big|\!\big|^\gamma_{\widetilde{\mathbb{L}}^{\gamma p}_{\gamma q}(T)},$$

and for fixed $\varepsilon > 0$ and any $t > 0$, $x \mapsto b^{Y,\varepsilon}_t(x)$ is continuous, by the dominated convergence theorem, we have

$$\lim_{n\to\infty} \mathbf{E}\left(\int_0^T \big|b^{Y,\varepsilon}_t(X^n_t) - b^{Y,\varepsilon}_t(X_t)\big|\,\mathrm{d}t\right) = 0,$$

which together with (41) yields (38). $\square$

There are, of course, many examples where the *weak* continuity assumption of $\mu \mapsto b_t(x,\mu)$ in the above lemma is not satisfied, as in the following interesting case:

$$b_t(x,\mu) = \int_{\mathbb{R}^d} \bar{b}_t(x,y)\mu(\mathrm{d}y), \tag{42}$$

where $\bar{b} : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a bounded measurable function. Obviously the weak continuity of $\mu \mapsto b(t,x,\mu)$ does not hold. However, in this case we still have the following limiting result.

Lemma 2.8. *Let $X^n, Y^n, X, Y \in \mathbf{S}_{\rm toch}$ be such that for each $t \geq 0$, X^n_t converges to X_t almost surely and Y^n_t converges to Y_t in distribution. Let $p_1, p_2, q_0, q_1, q_2 \in (1,\infty)$ with $\frac{1}{q_1} + \frac{1}{q_2} = 1 + \frac{1}{q_0}$ and $T, \beta, \kappa > 0$. Suppose that $X^n \in \mathbf{K}^{p_1,q_1}_{T,\kappa}$ and $Y^n \in \mathbf{K}^{p_2,q_2}_{T,\kappa}$ for each $n \in \mathbb{N}$, and that there is a constant $C_1 > 0$ such that*

$$\sup_n \sup_{t\in[0,T]} \mathbf{E}\big(\big|X^n_t\big|^\beta + \big|Y^n_t\big|^\beta\big) \leq C_1. \tag{43}$$

Let $\gamma > 1$. Then for any $\bar{b} \in \widetilde{\mathbb{L}}^{\gamma p_1,\gamma p_2}_{\gamma q_0}(T)$, we have

$$\lim_{n\to\infty} \mathbf{E}\left(\int_0^T \big|b^{Y^n}_t\big(X^n_t\big) - b^Y_t(X_t)\big|\,\mathrm{d}t\right) = 0. \tag{44}$$

Proof. Let $\mathbb{N}_\infty := \mathbb{N} \cup \{\infty\}$ and $Y^\infty := Y$, $X^\infty := X$. Since b^{Y^n} only depends on the distribution of Y^n, by Skorohod's representation, without loss of generality we may assume that $(X^n)_{n\in\mathbb{N}_\infty}$ and $(Y^n)_{n\in\mathbb{N}_\infty}$ are independent, and $(X^n_t, Y^n_t) \to (X_t, Y_t)$ a.e. as $n \to \infty$ for each t. Notice that by the assumptions and (33),

$$\sup_{n\in\mathbb{N}_\infty} \mathbf{E}\left(\int_0^T \big|\bar{b}_t\big(X^n_t, Y^n_t\big)\big|^\gamma\,\mathrm{d}t\right) \leq \kappa^2 |||\bar{b}|||^\gamma_{\widetilde{\mathbb{L}}^{\gamma p_1,\gamma p_2}_{\gamma q_0}(T)} < \infty. \tag{45}$$

Let $\bar{b}^\varepsilon_t(x,y) = \bar{b}_t * \varrho_\varepsilon(x,y)$ be a mollifying approximation of $\bar{b}$. As in the proof of (37), we have

$$\lim_{\varepsilon\to 0} \sup_{n\in\mathbb{N}_\infty} \mathbf{E}\left(\int_0^T \big|\bar{b}^\varepsilon_t\big(X^n_t, Y^n_t\big) - \bar{b}_t\big(X^n_t, Y^n_t\big)\big|\,\mathrm{d}t\right) = 0. \tag{46}$$

Thus, to prove (44), it suffices to show that for fixed $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} \mathbf{E}\left(\int_0^T \big|\bar{b}^\varepsilon_t\big(X^n_t, Y^n_t\big) - \bar{b}^\varepsilon_t\big(X^n_t, Y_t\big)\big|\,\mathrm{d}t\right) = 0,$$

$$\lim_{n\to\infty} \mathbf{E}\left(\int_0^T \big|\bar{b}^\varepsilon_t\big(X^n_t, Y_t\big) - \bar{b}^\varepsilon_t(X_t, Y_t)\big|\,\mathrm{d}t\right) = 0,$$

which follows by (45) and the dominated convergence theorem. □

3. Existence of weak and strong solutions

In this section, we show the weak existence and strong existence of DDSDEs with singular drifts. First of all, we recall the notions of martingale solutions and weak solutions for (1). Let $\mathbb{C}$ be the space of all continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^d$, which is endowed with the usual Borel σ-field $\mathcal{B}(\mathbb{C})$. The set of all probability measures on $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$ is denoted by $\mathcal{P}(\mathbb{C})$. Let w_t be the coordinate process over $\mathbb{C}$, that is,

$$w_t(\omega) = \omega_t, \quad \omega \in \mathbb{C}.$$

For $t \geq 0$, let $\mathcal{B}_t(\mathbb{C}) = \sigma\{w_s : s \leq t\}$ be the natural filtration. For a probability measure $\mathbb{P} \in \mathcal{P}(\mathbb{C})$, the expectation with respect to $\mathbb{P}$ will be denoted by $\mathbb{E}$ if there is no confusion.

Definition 3.1 (Martingale solutions). We call a probability measure $\mathbb{P} \in \mathcal{P}(\mathbb{C})$ a martingale solution of DDSDE (1) with initial distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$ if $\mathbb{P} \circ w_0^{-1} = \nu$ and for any $f \in C^\infty(\mathbb{R}^d)$,

$$\int_0^t \left|\mathscr{L}_s^{\sigma^{\mathbb{P}}} f\right|(w_s)\,\mathrm{d}s + \int_0^t \left|b_s^{\mathbb{P}} \cdot \nabla f\right|(w_s)\,\mathrm{d}s < \infty, \quad \mathbb{P}\text{-a.s.}, \forall t > 0,$$

where $\sigma_t^{\mathbb{P}}(x) := \sigma_t(x, \mu_t^{\mathbb{P}})$ and $b_t^{\mathbb{P}}(x) := b_t(x, \mu_t^{\mathbb{P}})$, $\mu_t^{\mathbb{P}} := \mathbb{P} \circ w_t^{-1}$, and

$$M_t^f := f(w_t) - f(w_0) - \int_0^t \big(\mathscr{L}_s^{\sigma^{\mathbb{P}}} f\big)(w_s)\,\mathrm{d}s - \int_0^t \big(b_s^{\mathbb{P}} \cdot \nabla f\big)(w_s)\,\mathrm{d}s, \tag{47}$$

is a continuous local $\mathcal{B}_t(\mathbb{C})$-martingale under $\mathbb{P}$. All the martingale solutions of DDSDE (1) with coefficients σ, b and initial distribution ν are denoted by $\mathscr{M}_\nu^{\sigma,b}$.

Definition 3.2 (Weak solutions). Let (X, W) be two $\mathbb{R}^d$-valued continuous adapted processes on some filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P})$. We call

$$\big(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P}; X, W\big)$$

a weak solution of DDSDE (1) with initial distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$ if

(i) $\mathbf{P} \circ X_0^{-1} = \nu$ and W is a d-dimensional standard $\mathscr{F}_t$-Brownian motion.
(ii) For all $t > 0$, it holds that

$$\int_0^t |b_s|(X_s, \mu_{X_s})\,\mathrm{d}s + \int_0^t \left\|\sigma_s\sigma_s^*\right\|_{\mathrm{HS}}(X_s, \mu_{X_s})\,\mathrm{d}s < \infty, \quad \mathbf{P}\text{-a.s.}$$

and

$$X_t = X_0 + \int_0^t b_s(X_s, \mu_{X_s})\,\mathrm{d}s + \int_0^t \sigma_s(X_s, \mu_{X_s})\,\mathrm{d}W_s, \quad \mathbf{P}\text{-a.s.} \tag{48}$$

Remark 3. It is well known that weak solutions and martingale solutions are equivalent (cf. [23]), which means that for any $\mathbb{P} \in \mathscr{M}_\nu^{\sigma,b}$, there is a weak solution

$$\big(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P}; X, W\big)$$

to DDSDE (1) with initial distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$ such that

$$\mathbb{P} = \mathbf{P} \circ X^{-1}.$$

Now we make the following assumptions about σ and b:

($\mathbf{H}^{\sigma,b}$) For each t, x, the mapping $\mu \mapsto \sigma_t(x,\mu)$ is weakly continuous, and there are $c_0 \geq 1$ and $\gamma \in (0,1]$ such that for all $t \geq 0$, $x, x', \xi \in \mathbb{R}^d$ and $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$c_0^{-1}|\xi| \leq \big|\sigma_t(x,\mu)\xi\big| \leq c_0|\xi|, \qquad \big\|\sigma_t(x,\mu) - \sigma_t\big(x',\mu\big)\big\|_{\mathrm{HS}} \leq c_0\big|x - x'\big|^\gamma. \tag{49}$$

The drift b satisfies one of the following conditions:

(i) For each t, x, the mapping $\mu \mapsto b_t(x,\mu)$ is weakly continuous, and for some $(p,q) \in \mathscr{I}_1$ and $\kappa_0 > 0$,

$$\sup_{Z \in \mathbf{S}_{\mathrm{toch}}} \big\|\big|b^Z\big\|\big|_{\widetilde{\mathbb{L}}^p_q(T)} \leq \kappa_0 < \infty. \tag{50}$$

(ii) b has the form (42) with $\bar{b}$ satisfying ($\mathbf{H}^b$).

It should be noticed that under ($\mathbf{H}^b$), (50) holds. Indeed, by definition we have

$$\begin{aligned}
\big\|\big|b^Z\big\|\big|^q_{\widetilde{\mathbb{L}}^p_q(T)} &= \sup_{z\in\mathbb{R}^d} \int_0^T \bigg\|\chi_r^z \int_{\mathbb{R}^d} \bar{b}_s(\cdot,y)\mu_{Z_s}(\mathrm{d}y)\bigg\|_p^q \,\mathrm{d}s \\
&\leq \sup_{z\in\mathbb{R}^d} \int_0^T \bigg\|\chi_r^z \int_{\mathbb{R}^d} h_s(\cdot - y)\mu_{Z_s}(\mathrm{d}y)\bigg\|_p^q \,\mathrm{d}s \\
&\leq \sup_{z\in\mathbb{R}^d} \int_0^T \int_{\mathbb{R}^d} \big\|\chi_r^{z-y} h_s\big\|_p^q \mu_{Z_s}(\mathrm{d}y)\,\mathrm{d}s \\
&\leq \int_0^T \sup_{z\in\mathbb{R}^d} \big\|\chi_r^z h_s\big\|_p^q \mu_{Z_s}\big(\mathbb{R}^d\big)\,\mathrm{d}s = \int_0^T \|\!|h_s|\!\|_p^q \,\mathrm{d}s.
\end{aligned}$$

To show the existence of weak solutions, we first establish the following a priori estimates.

Lemma 3.3. *Let $\beta > 0$. Under ($\mathbf{H}^{\sigma,b}$), for any $\nu \in \mathcal{P}_\beta(\mathbb{R}^d)$ and $Z \in \mathbf{S}_{\mathrm{toch}}$, there is a unique weak solution $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P}; X, W)$ to the following SDE*:

$$\mathrm{d}X_t = b_t^Z(X_t)\,\mathrm{d}t + \sigma_t^Z(X_t)\,\mathrm{d}W_t, \qquad \mathbf{P} \circ X_0^{-1} = \nu.$$

Moreover, letting $\Theta = (d, p, q, c_0, \gamma, \kappa_0, \beta)$, we have

(i) *For any $T > 0$, there is a $C_1 = C_1(\Theta, T) > 0$ such that*

$$\mathbf{E}\Big(\sup_{t\in[0,T]} |X_t|^\beta\Big) \leq C_1\big(\mathbf{E}|X_0|^\beta + 1\big), \tag{51}$$

and for any $\delta \in (0,T)$,

$$\mathbf{E}\Big(\sup_{t,s\in[0,T],|t-s|\leq\delta} |X_t - X_s|^\beta\Big) \leq C_1\delta^{\beta/2}. \tag{52}$$

(ii) *For any $(p_1,q_1) \in \mathscr{I}_2$ and $T > 0$, there is a constant $C_2 = C_2(p_1,q_1,\Theta,T) > 0$ such that for all $0 \leq t_0 < t_1 \leq T$ and $f \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(t_0,t_1)$,*

$$\mathbf{E}\bigg(\int_{t_0}^{t_1} f_s(X_s)\,\mathrm{d}s \Big| \mathscr{F}_{t_0}\bigg) \leq C_2 \|\!|f|\!\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(t_0,t_1)}. \tag{53}$$

Proof. The proof of this lemma is essentially contained in [32]. For the reader's convenience, we sketch the proofs below. We use Zvonkin's transformation to kill the drift b^Z (cf. [35]). For $\lambda, T > 0$, consider the following backward PDE:

$$\partial_t u + \big(\mathscr{L}_t^{\sigma^Z} - \lambda\big)u + b^Z \cdot \nabla u + b^Z = 0, \qquad u(T,x) = 0.$$

Since $b^Z \in \widetilde{\mathbb{L}}_q^p(T)$ with $(p,q) \in \mathscr{I}_1$, by Theorem 2.2, for $\lambda \geq 1$, there is a unique solution $u \in \widetilde{\mathbb{H}}_q^{2,p}(T)$ solving the above PDE. Moreover, for any $\alpha \in [0, 2 - \frac{2}{q})$, there is a constant $c_1 = c_1(\alpha, \Theta, T) > 0$ such that for all $\lambda \geq 1$,

$$\lambda^{1-\frac{\alpha}{2}-\frac{2}{q}} \|u\|_{\widetilde{\mathbb{H}}_\infty^{\alpha,p}(T)} + \|u\|_{\widetilde{\mathbb{H}}_q^{2,p}(T)} \leq c_1 \big\|b^Z\big\|_{\widetilde{\mathbb{L}}_q^p(T)}. \tag{54}$$

In particular, since $\frac{d}{p} + \frac{2}{q} < 1$, by (16) we can choose λ large enough so that

$$\|u\|_{\mathbb{L}^\infty(T)} + \|\nabla u\|_{\mathbb{L}^\infty(T)} \leq 1/2.$$

Now if we define

$$\Phi_t(x) := x + u_t(x),$$

then it is easy to see that

$$|x-y|/2 \leq \big|\Phi_t(x) - \Phi_t(y)\big| \leq 2|x-y|, \tag{55}$$

and

$$\partial_t \Phi + \mathscr{L}_t^{\sigma^Z}\Phi + b^Z \cdot \nabla \Phi = \lambda u. \tag{56}$$

By the generalized Itô formula and (56), we have

$$\begin{aligned} Y_t := \Phi_t(X_t) &= \Phi_0(X_0) + \lambda \int_0^t u_s(X_s)\,\mathrm{d}s + \int_0^t \big(\sigma_s^Z \cdot \nabla \Phi_s\big)(X_s)\,\mathrm{d}W_s \\ &= \Phi_0(X_0) + \int_0^t \tilde{b}_s(Y_s)\,\mathrm{d}s + \int_0^t \tilde{\sigma}_s(Y_s)\,\mathrm{d}W_s, \end{aligned} \tag{57}$$

where

$$\tilde{\sigma} := \big(\sigma^Z \cdot \nabla \Phi\big) \circ \Phi^{-1}, \qquad \tilde{b} := \lambda u \circ \Phi^{-1}.$$

Moreover, by (54), (55) and the Sobolev embedding (16), it is easy to see that for some $c_2 = c_2(\Theta, T) > 0$ and $\gamma_0 = \gamma_0(\gamma, p, q) \in (0,1)$,

$$c_2^{-1}|\xi| \leq \big|\tilde{\sigma}_t(x)\xi\big| \leq c_2|\xi|, \qquad \big\|\tilde{\sigma}_t(x) - \tilde{\sigma}_t(y)\big\|_{\mathrm{HS}} \leq c_2|x-y|^{\gamma_0}, \tag{58}$$

and

$$\|\tilde{b}\|_{\mathbb{L}^\infty(T)} + \|\nabla \tilde{b}\|_{\mathbb{L}^\infty(T)} \leq 4\lambda. \tag{59}$$

By well-known results, SDE (57) admits a unique weak solution (cf. [23]). Moreover, as in [32], one can check that $X_t := \Phi_t^{-1}(Y_t)$ solves the original SDE.

(i) Let $\beta > 0$. By (58) and (59), estimate (51) directly follows by BDG's inequality. We prove (52). Fix $\delta \in (0,T)$. Let τ be any stopping time less than $T - \delta$. By equation (57) and BDG's inequality, we have

$$\mathbf{E}\Big(\sup_{r\in[0,\delta]}|Y_{\tau+r}-Y_\tau|^\beta\Big)\lesssim \mathbf{E}\bigg(\int_\tau^{\tau+\delta}\big|\tilde b_s(X_s)\big|\,\mathrm{d}s\bigg)^\beta+\mathbf{E}\bigg|\sup_{r\in[0,\delta]}\int_\tau^{\tau+r}\tilde\sigma_s(X_s)\,\mathrm{d}W_s\bigg|^\beta$$
$$\lesssim \|\tilde b\|^\beta_{\mathbb{L}^\infty(T)}\delta^\beta+\|\tilde\sigma\|^\beta_{\mathbb{L}^\infty(T)}\delta^{\beta/2}\le C\delta^{\beta/2},$$

which yields (52) by [33], Lemma 2.7 and (55).

(ii) It was proved in [29], Lemma 4.1, that for any $(p_1,q_1)\in\mathscr{I}_2$, there is a constant $C_2=C_2(p_1,q_1,\Theta,T)>0$ such that for all $0\le t_0<t_1\le T$ and $f\in\widetilde{\mathbb{L}}^{p_1}_{q_1}(t_0,t_1)$,

$$\mathbf{E}\bigg(\int_{t_0}^{t_1}f_s(Y_s)\,\mathrm{d}s|\mathscr{F}_{t_0}\bigg)\le C_2|||f|||_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(t_0,t_1)}.$$

By a change of variable and (55) again, we obtain (53). □

Remark 4. An important conclusion of (ii) above is the following Khasminskii's type estimate (see [30], Lemma 3.5): For any $\lambda,T>0$ and $f\in\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$ with $(p_1,q_1)\in\mathscr{I}_2$,

$$\mathbf{E}\exp\bigg(\lambda\int_0^T\big|f_s(X_s)\big|\,\mathrm{d}s\bigg)\le C_3, \tag{60}$$

where C_3 only depends on λ, Θ, p_1, q_1, T and $|||f|||_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}$.

Now we can show the following weak existence result.

Theorem 3.4. *Let $\beta>0$. Under $(\mathbf{H}^{\sigma,b})$, for any $\nu\in\mathcal{P}_\beta(\mathbb{R}^d)$, there exists a weak solution*

$$\big(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\ge0},\mathbf{P};X,W\big)$$

to DDSDE (1) with $\mathbf{P}\circ X_0^{-1}=\nu$.

Proof. For $n\in\mathbb{N}$, consider the following approximating SDE:

$$X^n_t=X^n_0+\int_0^t b^n_s\big(X^n_s,\mu_{X^n_s}\big)\,\mathrm{d}s+\int_0^t\sigma_s\big(X^n_s,\mu_{X^n_s}\big)\,\mathrm{d}W_s, \tag{61}$$

where

$$b^n_s(x,\mu):=(-n)\vee b_s(x,\mu)\wedge n\quad\text{in case (i) of }\big(\mathbf{H}^{\sigma,b}\big),$$

and

$$\bar b^n_s(x,y):=(-n)\vee\bar b_s(x,y)\wedge n\quad\text{in case (ii) of }\big(\mathbf{H}^{\sigma,b}\big).$$

Since b^n is bounded measurable, by [21] or [31], Theorem 1.2, there is a weak solution

$$\big(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\ge0},\mathbf{P};X^n,W\big)$$

to DDSDE (61) with $\mathbf{P}\circ(X^n_0)^{-1}=\nu$. Moreover, since

$$\sup_{Z\in\mathbf{S}_{\rm toch}}|||b^{n,Z}|||_{\widetilde{\mathbb{L}}^p_q(T)}\le\sup_{Z\in\mathbf{S}_{\rm toch}}|||b^Z|||_{\widetilde{\mathbb{L}}^p_q(T)}\le\kappa_0,$$

by Lemma 3.3, the following uniform estimates hold:

(i) For any $T > 0$, there is a constant $C_1 > 0$ such that

$$\sup_n \mathbf{E}\Big(\sup_{t\in[0,T]} \big|X_t^n\big|^\beta\Big) \le C_1\big(\mathbf{E}|X_0|^\beta + 1\big),$$

and for all $\delta \in (0, T)$,

$$\sup_n \mathbf{E}\Big(\sup_{t,s\in[0,T],|t-s|\le\delta} \big|X_t^n - X_s^n\big|^\beta\Big) \le C_1\delta^{\beta/2}.$$

(ii) Let $(p_1, q_1) \in \mathscr{I}_2$. For any $T > 0$, there is a $C_2 > 0$ such that for all $f \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(T)$,

$$\sup_n \mathbf{E}\Big(\int_0^T f_s\big(X_s^n\big)\,\mathrm{d}s\Big) \le C_2 |||f|||_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}.$$

Now by (i), the laws $\mathbb{Q}^n$ of (X^n, W) in $\mathbb{C} \times \mathbb{C}$ are tight. Let $\mathbb{Q}$ be any accumulation point of $\mathbb{Q}^n$. Without loss of generality, we assume that $\mathbb{Q}^n$ weakly converges to some probability measure $\mathbb{Q}$. By Skorokhod's representation theorem, there are a probability space $(\tilde\Omega, \tilde{\mathscr{F}}, \tilde{\mathbf{P}})$ and random variables $(\tilde X^n, \tilde W^n)$ and $(\tilde X, \tilde W)$ defined on it such that

$$\big(\tilde X^n, \tilde W^n\big) \to (\tilde X, \tilde W), \quad \tilde{\mathbf{P}}\text{-a.s.} \tag{62}$$

and

$$\tilde{\mathbf{P}} \circ \big(\tilde X^n, \tilde W^n\big)^{-1} = \mathbb{Q}^n = \mathbf{P} \circ \big(X^n, W\big)^{-1}, \qquad \tilde{\mathbf{P}} \circ (\tilde X, \tilde W)^{-1} = \mathbb{Q}. \tag{63}$$

Define $\tilde{\mathscr{F}}^n_t := \sigma(\tilde W^n_s, \tilde X^n_s; s \le t)$. We note that

$$\mathbf{P}(W_t - W_s \in \cdot|\mathscr{F}_s) = \mathbf{P}(W_t - W_s \in \cdot) \quad \Rightarrow \quad \tilde{\mathbf{P}}\big(\tilde W^n_t - \tilde W^n_s \in \cdot|\tilde{\mathscr{F}}^n_s\big) = \tilde{\mathbf{P}}\big(\tilde W^n_t - \tilde W^n_s \in \cdot\big).$$

In other words, $\tilde W^n$ is an $\tilde{\mathscr{F}}^n_t$-Brownian motion. Thus, by (61) and (63) we have

$$\tilde X^n_t = \tilde X^n_0 + \int_0^t b^n_s\big(\tilde X^n_s, \mu_{\tilde X^n_s}\big)\,\mathrm{d}s + \int_0^t \sigma_s\big(\tilde X^n_s, \mu_{\tilde X^n_s}\big)\,\mathrm{d}\tilde W^n_s.$$

By (ii), (62), Lemmas 2.7, 2.8 and [12], Theorem 6.22, page 383, one can take limits as $n \to \infty$ to obtain

$$\tilde X_t = \tilde X_0 + \int_0^t b_s(\tilde X_s, \mu_{\tilde X_s})\,\mathrm{d}s + \int_0^t \sigma_s(\tilde X_s, \mu_{\tilde X_s})\,\mathrm{d}\tilde W_s.$$

Here we only check that the assumptions of Lemma 2.8 are satisfied in the case that b takes the form (42) with $\bar b$ satisfying $(\mathbf{H}^b)$. Clearly, by (ii) above, for any $(p_1, q_1) \in \mathscr{I}_2$, there is a $\kappa > 0$ such that for each $n \in \mathbb{N}$,

$$\tilde X^n \in \mathbf{K}^{p_1,q_1}_{T,\kappa}.$$

We note that $|\bar b_t(x, y)| \le h_t(x - y)$, where for some $(p, q) \in \mathscr{I}_1$, $h \in \widetilde{\mathbb{L}}^p_q$. One can choose $\gamma > 1$ so that $\frac{d\gamma}{p} + \frac{2\gamma}{q} < 1$. Now if we take $p_1 = \frac{p}{\gamma}$, $q_0 = \frac{q}{\gamma}$, $q_1 = \frac{2q}{q+2\gamma}$, then it is easy to see that $(p_1, q_1) \in \mathscr{I}_2$

and

$$\bar{b} \in \widetilde{\mathbb{L}}_q^{p,\infty} = \widetilde{\mathbb{L}}_{\gamma q_0}^{\gamma p_1,\infty} \subset \bigcap_{p' \geq 1} \widetilde{\mathbb{L}}_{\gamma q_0}^{\gamma p_1,p'}.$$

Thus one can apply Lemma 2.8 to conclude that

$$\lim_{n\to\infty} \tilde{\mathbf{E}}\left(\int_0^t \left| b_s\left(\tilde{X}_s^n, \mu_{\tilde{X}_s^n}\right) - b_s(\tilde{X}_s, \mu_{\tilde{X}_s}) \right| \mathrm{d}s \right) = 0.$$

Moreover, as in showing (46), we also have

$$\lim_{m\to\infty} \sup_n \tilde{\mathbf{E}}\left(\int_0^t \left| \bar{b}_s^m - \bar{b}_s \right| \left(\tilde{X}_s^n, Y_s^n\right) \mathrm{d}s \right) = 0,$$

where $Y_\cdot^n$ is an independent copy of $\tilde{X}_\cdot^n$. The proof is thus complete. □

About the existence of strong solutions, we have the following corollary.

Corollary 3.5. *Let $\beta > 0$. Under $(\mathbf{H}^{\sigma,b})$, if for some $(p_1, q_1) \in \mathscr{I}_1$,*

$$\sup_{Z \in \mathbf{S}_{\mathrm{toch}}} \left\| \nabla \sigma^Z \right\|_{\widetilde{\mathbb{L}}_{p_1}^{q_1}(T)} < \infty,$$

then for any initial random variable X_0 with finite β-order moment, there exists a strong solution to DDSDE (1).

Proof. Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbf{P}; X, W)$ be a weak solution of DDSDE (1). Define

$$b_t^X(x) := b_t(x, \mu_{X_t}), \qquad \sigma_t^X(x) := \sigma_t(x, \mu_{X_t}), \qquad \mu_{X_t} := \mathbf{P} \circ X_t^{-1}.$$

Consider the following SDE:

$$\mathrm{d}Z_t = b_t^X(Z_t)\,\mathrm{d}t + \sigma_t^X(Z_t)\,\mathrm{d}W_t.$$

Under the assumption of the theorem, it has been shown in [29] that there is a unique strong solution to this equation. Since X also satisfies the above equation, by strong uniqueness, we obtain that $X = Z$ is a strong solution. □

Remark 5. Although we have shown the existence of strong or weak solutions, the uniqueness of strong solutions or weak solutions is a more difficult problem.

4. Uniqueness of strong and weak solutions

In this section, we study the uniqueness of strong and weak solutions. We introduce the following assumptions about the dependence on the third variable μ:

$(\mathbf{A}_\theta^{\sigma,b})$ Let $\theta \geq 1$. We assume (49) and for some $(p,q), (p_1,q_1) \in \mathscr{I}_1$,

$$\sup_{Z \in \mathbf{S}_{\mathrm{toch}}} \left\| b^Z \right\|_{\widetilde{\mathbb{L}}_q^p(T)} < \infty, \qquad \sup_{Z \in \mathbf{S}_{\mathrm{toch}}} \left\| \nabla \sigma^Z \right\|_{\widetilde{\mathbb{L}}_{q_1}^{p_1}(T)} < \infty,$$

and there are $\ell \in L^q_{loc}(\mathbb{R}_+)$ and a constant $c_1 \geq 1$ such that for any two random variables X, Y with finite θ-order moments,

$$\begin{aligned} \left|\!\left|\!\left| b_t(\cdot, \mu_X) - b_t(\cdot, \mu_Y) \right|\!\right|\!\right|_p &\leq \ell_t \|X - Y\|_\theta, \\ \left\| \sigma_t(\cdot, \mu_X) - \sigma_t(\cdot, \mu_Y) \right\|_\infty &\leq c_1 \|X - Y\|_\theta, \end{aligned} \tag{64}$$

where $\|\cdot\|_\theta$ stands for the L^θ-norm in the probability space $(\Omega, \mathscr{F}, \mathbf{P})$.

Notice that (64) is equivalent to that for all $\mu, \mu' \in \mathcal{P}_\theta(\mathbb{R}^d)$,

$$\begin{aligned} \left\| b_t(\cdot, \mu) - b_t\left(\cdot, \mu'\right) \right\|_p &\leq \ell_t \mathcal{W}_\theta\left(\mu, \mu'\right), \\ \left\| \sigma_t(\cdot, \mu) - \sigma_t\left(\cdot, \mu'\right) \right\|_\infty &\leq c_0 \mathcal{W}_\theta\left(\mu, \mu'\right), \end{aligned}$$

where $\mathcal{W}_\theta$ is the usual Wasserstein metric of θ-order. In particular, $(\mathbf{A}^{\sigma,b}_\theta) \Rightarrow (\mathbf{H}^{\sigma,b})$. For convenience, we would like to use (64) rather than introducing the Wasserstein metric.

Remark 6. We note that in [11], (64) is assumed to hold for $p = \infty$.

We first show the following strong uniqueness result.

Theorem 4.1. *Let $\theta \geq 1$ and $\beta > \theta$. Under $(\mathbf{A}^{\sigma,b}_\theta)$, for any initial random variable X_0 with finite β-order moment, there is a unique strong solution to DDSDE* (1).

Proof. Below we fix $p, q \in \mathscr{I}_1$, and without loss of generality, we consider the time interval $[0, 1]$ and assume that for some $\gamma > 1$,

$$\|\ell\|_{L^{\gamma q}(0,1)} + \sup_{Z \in \mathbf{S}_{\text{toch}}} \left\| b^Z \right\|_{\mathbb{L}^{\gamma q}_{\gamma p}(1)} < \infty. \tag{65}$$

Otherwise, we may choose $\gamma > 1$ so that $\frac{2\gamma}{q} + \frac{d\gamma}{p} < 1$ holds and replace (p, q) with $(p/\gamma, q/\gamma)$. The existence of strong solutions has been shown in Corollary 3.5. We only need to prove the pathwise uniqueness. Let X, Y be two strong solutions defined on the same probability space with same starting points $X_0 = Y_0$ a.s. We divide the proof into three steps and use the convention that all the constants below will be independent of $T \in [0, 1]$.

(i) Let $T \in (0, 1)$ and $\lambda > 0$. We consider the following backward PDE:

$$\partial_t u^X + \left(\mathscr{L}^{\sigma^X}_t - \lambda\right) u + b^X \cdot \nabla u^X + b^X = 0, \qquad u^X_T(x) = 0. \tag{66}$$

By Theorem 2.2, for $\lambda \geq 1$, there is a unique solution $u^X \in \widetilde{\mathbb{H}}^{2,p}_q(T)$ solving the above PDE. Moreover, for any $\alpha \in [0, 2 - \frac{2}{q})$, there is a constant $c_1 > 0$ such that for all $\lambda \geq 1$ and $T \in [0, 1]$,

$$\lambda^{1 - \frac{\alpha}{2} - \frac{2}{q}} \left|\!\left|\!\left| u^X \right|\!\right|\!\right|_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} + \left|\!\left|\!\left| u^X \right|\!\right|\!\right|_{\widetilde{\mathbb{H}}^{2,p}_q(T)} \leq c_1 \left|\!\left|\!\left| b^X \right|\!\right|\!\right|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{67}$$

In particular, since $\frac{d}{p} + \frac{2}{q} < 1$, by (16), we can choose λ large enough so that

$$\left\| u^X \right\|_{\mathbb{L}^\infty(T)} + \left\| \nabla u^X \right\|_{\mathbb{L}^\infty(T)} \leq 1/2, \quad \forall T \in [0, 1]. \tag{68}$$

Below we shall fix such a λ and define

$$\Phi_t^X(x) := x + u_t^X(x).$$

It is easy to see that

$$\partial_t \Phi^X + \mathscr{L}_t^{\sigma^X} \Phi^X + b^X \cdot \nabla \Phi^X = \lambda u^X.$$

(ii) By the generalized Itô formula (cf. [29], Lemma 4.1), we have

$$\tilde{X}_t := \Phi_t^X(X_t) = \Phi_0^X(X_0) + \lambda \int_0^t u_s^X(X_s)\,\mathrm{d}s + \int_0^t \tilde{\sigma}_s^X(X_s)\,\mathrm{d}W_s, \tag{69}$$

where

$$\tilde{\sigma}^X := \sigma^X \cdot \nabla \Phi^X.$$

Similarly, we define $\tilde{Y}_t := \Phi_t^Y(Y_t)$, and for simplicity write

$$\xi_t := X_t - Y_t, \qquad \tilde{\xi}_t := \tilde{X}_t - \tilde{Y}_t.$$

Noting that by (68),

$$|x-y| \le 2\big|\Phi_t^X(x) - \Phi_t^X(y)\big| \le 2\big|\Phi_t^X(x) - \Phi_t^Y(y)\big| + 2\big\|u^X - u^Y\big\|_{\mathbb{L}^\infty(T)}$$

and

$$\big|\Phi_t^X(x) - \Phi_t^Y(y)\big| \le 2|x-y| + \big\|u^X - u^Y\big\|_{\mathbb{L}^\infty(T)},$$

we have

$$|\xi_t| \le 2|\tilde{\xi}_t| + 2\big\|u^X - u^Y\big\|_{\mathbb{L}^\infty(T)}, \qquad |\tilde{\xi}_t| \le 2|\xi_t| + \big\|u^X - u^Y\big\|_{\mathbb{L}^\infty(T)}. \tag{70}$$

By (69) and again Itô's formula, we have for any $\beta \ge 1$,

$$\begin{aligned}
|\tilde{\xi}_t|^\beta &= |\tilde{\xi}_0|^\beta + \beta\lambda \int_0^t |\tilde{\xi}_s|^{\beta-2}\big\langle \tilde{\xi}_s, u_s^X(X_s) - u_s^Y(Y_s)\big\rangle\,\mathrm{d}s \\
&\quad + \beta \int_0^t |\tilde{\xi}_s|^{\beta-2}\big\langle \big(\tilde{\sigma}_s^X(X_s) - \tilde{\sigma}_s^Y(Y_s)\big)^* \tilde{\xi}_s, \mathrm{d}W_s\big\rangle \\
&\quad + \beta\left(\frac{\beta}{2} - 1\right) \int_0^t |\tilde{\xi}_s|^{\beta-4}\big|\big(\tilde{\sigma}_s^X(X_s) - \tilde{\sigma}_s^Y(Y_s)\big)^* \tilde{\xi}_s\big|^2\,\mathrm{d}s \\
&\quad + \frac{\beta}{2} \int_0^t |\tilde{\xi}_s|^{\beta-2}\big\|\tilde{\sigma}_s^X(X_s) - \tilde{\sigma}_s^Y(Y_s)\big\|_{\mathrm{HS}}^2\,\mathrm{d}s \\
&:= I_1 + I_2 + I_3 + I_4 + I_5.
\end{aligned}$$

Since by (68),

$$\big|u_t^X(x) - u_t^Y(y)\big| \le |x-y| + \big\|u^X - u^Y\big\|_{\mathbb{L}^\infty(T)},$$

by Young's inequality we obtain

$$I_2 \lesssim \int_0^t |\tilde{\xi}_s|^\beta \, \mathrm{d}s + \lambda \int_0^t \big|u_s^X(X_s) - u_s^Y(Y_s)\big|^\beta \, \mathrm{d}s$$
$$\lesssim \int_0^t \big(|\tilde{\xi}_s|^\beta + \lambda|\xi_s|^\beta\big) \, \mathrm{d}s + \lambda^\beta T \big\|u^X - u^Y\big\|^\beta_{\mathbb{L}^\infty(T)}.$$

Let

$$g_s^X(x) := \big|\nabla^2 u_s^X(x)\big| + \big|\nabla \sigma_s^X(x)\big| + \big\|\nabla u^X\big\|_{\mathbb{L}^\infty(T)} + \big\|\sigma^X\big\|_{\mathbb{L}^\infty(T)}.$$

By the definition of $\tilde{\sigma}^X$, we also have that

$$\begin{aligned}
&\big|\tilde{\sigma}_s^X(x) - \tilde{\sigma}_s^Y(y)\big| \\
&\quad \le \big\|\sigma^Y\big\|_{\mathbb{L}^\infty(T)} \big|\nabla \Phi_s^X(x) - \nabla \Phi_s^Y(y)\big| + \big|\sigma_s^X(x) - \sigma_s^Y(y)\big| \cdot \big\|\nabla \Phi^X\big\|_{\mathbb{L}^\infty(T)} \\
&\quad \le \big\|\sigma^Y\big\|_{\mathbb{L}^\infty(T)} \big(\big|\nabla u_s^X(x) - \nabla u_s^X(y)\big| + \big|\nabla u_s^X(y) - \nabla u^Y(s,y)\big|\big) \\
&\qquad + \big(\big|\sigma_s^X(x) - \sigma_s^X(y)\big| + \big|\sigma_s^X(y) - \sigma_s^Y(y)\big|\big) \cdot \big\|\nabla \Phi^X\big\|_{\mathbb{L}^\infty(T)} \\
&\quad \overset{(30)}{\lesssim} |x-y|\big(\mathcal{M}_1 g_s^X(x) + \mathcal{M}_1 g_s^X(y)\big) + \big\|\nabla u^X - \nabla u^Y\big\|_{\mathbb{L}^\infty(T)} + \big\|\sigma_s^X - \sigma_s^Y\big\|_\infty.
\end{aligned}$$

Hence,

$$I_4 + I_5 \lesssim \int_0^t \big(|\xi_s|^\beta + |\tilde{\xi}_s|^\beta\big)\big(\mathcal{M} g_s^X(X_s) + \mathcal{M} g_s^X(Y_s)\big)^2 \, \mathrm{d}s$$
$$+ T \big\|\nabla u^X - \nabla u^Y\big\|^\beta_{\mathbb{L}^\infty(T)} + \int_0^t \big\|\sigma_s^X - \sigma_s^Y\big\|^\beta_\infty \, \mathrm{d}s.$$

Combining the above calculations and noting that $|\tilde{\xi}_0| \le \|u_0^X - u_0^Y\|_\infty$, we obtain

$$\begin{aligned}
|\tilde{\xi}_t|^\beta \lesssim{}& \big\|u^X - u^Y\big\|^\beta_{\mathbb{H}^{1,\infty}_\infty(T)} + \int_0^t \big(|\tilde{\xi}_s|^\beta + |\xi_s|^\beta + \|\xi_s\|^\beta_\theta\big) \, \mathrm{d}s \\
&+ \int_0^t \big(|\xi_s|^\beta + |\tilde{\xi}_s|^\beta\big)\big(\mathcal{M}_1 g_s^X(X_s) + \mathcal{M}_1 g_s^X(Y_s)\big)^2 \, \mathrm{d}s + M_t,
\end{aligned} \tag{71}$$

where M_t is a continuous local martingale.

(iii) Now we define

$$A_t := t + \int_0^t \big(\mathcal{M}_1 g_s^X(X_s) + \mathcal{M}_1 g_s^X(Y_s)\big)^2 \, \mathrm{d}s.$$

By (71) and (70), we obtain that for all $t \in [0, T]$,

$$|\xi_s|^\beta + |\tilde{\xi}_s|^\beta \lesssim \big\|u^X - u^Y\big\|^\beta_{\mathbb{H}^{1,\infty}_\infty(T)} + \int_0^t \|\xi_s\|^\beta_\theta \, \mathrm{d}s + \int_0^t \big(|\xi_s|^\beta + |\tilde{\xi}_s|^\beta\big) \, \mathrm{d}A_s + M_t.$$

Note that by the assumption and (31),

$$(s,x) \mapsto \big(\mathcal{M}_1 \big|\nabla^2 u_s^X(x)\big|\big)^2 \in \widetilde{\mathbb{L}}^{p/2}_{q/2}(T)$$

and

$$(s,x)\mapsto \big(\mathcal{M}_1\big|\nabla\sigma^X_s(x)\big|\big)^2\in\widetilde{\mathbb{L}}^{p_1/2}_{q_1/2}(T).$$

Since $(\frac{p}{2},\frac{q}{2}),(\frac{p_1}{2},\frac{q_1}{2})\in\mathscr{I}_2$, by Khasminskii's estimate (60), we have

$$\mathbb{E}\exp\gamma A_T<\infty,\quad \forall\gamma>0,\forall T\in[0,1].$$

Thus we can use the stochastic Gronwall inequality (29) to derive that

$$\sup_{s\in[0,T]}\|\xi_s\|_\theta^\beta=\Big(\sup_{s\in[0,T]}\mathbf{E}|\xi_s|^\theta\Big)^{\beta/\theta}\lesssim\big\|u^X-u^Y\big\|^\beta_{\mathbb{H}^{1,\infty}_\infty(T)}+\int_0^T\|\xi_s\|_\theta^\beta\,\mathrm{d}s. \tag{72}$$

Noticing that by (64),

$$\big|\!\big|\!\big|b^X-b^Y\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^p_q(T)}\le\Big(\int_0^T\ell_t^q\,\|X_t-Y_t\|_\theta^q\,\mathrm{d}t\Big)^{1/q}\le\|\ell\|_{L^q(0,T)}\sup_{t\in[0,T]}\|\xi_t\|_\theta$$

and

$$\big\|\sigma^X-\sigma^Y\big\|_{\mathbb{L}^\infty(T)}\le c_0\sup_{t\in[0,T]}\|X_t-Y_t\|_\theta=c_0\sup_{t\in[0,T]}\|\xi_t\|_\theta,$$

we have by (27),

$$\begin{aligned}\big\|u^X-u^Y\big\|_{\mathbb{H}^{1,\infty}_\infty(T)}&\lesssim\big|\!\big|\!\big|b^X-b^Y\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^p_q(T)}+\big|\!\big|\!\big|b^X\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^p_q(T)}\big(\big\|\sigma^X-\sigma^Y\big\|_{\mathbb{L}^\infty(T)}+\big|\!\big|\!\big|b^X-b^Y\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^p_q(T)}\big)\\&\lesssim\big(\|\ell\|_{L^q(0,T)}+\big|\!\big|\!\big|b^X\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^p_q(T)}\big)\sup_{t\in[0,T]}\|\xi_t\|_\theta\overset{(65)}{\lesssim}T^{\frac{\gamma-1}{\gamma q}}\sup_{t\in[0,T]}\|\xi_t\|_\theta.\end{aligned}$$

Substituting this into (72), we obtain

$$\sup_{s\in[0,T]}\|\xi_s\|_\theta^\beta\le CT^{\frac{\beta(\gamma-1)}{\gamma q}}\sup_{t\in[0,T]}\|\xi_t\|_\theta^\beta,\quad T\in(0,1),$$

where C does not depend on $T\in(0,1)$. By choosing T small enough, we get $\|\xi_t\|_\theta^\beta=0$ for all $t\in[0,T]$. By shifting the time T, we obtain the uniqueness. □

It is obvious that b defined in (42) does not satisfy (64). Below we shall relax it to the weighted total variation norm by Girsanov's transformation. The price we have to pay is that we need to assume that the diffusion coefficient does not depend on the time marginal law of X. For $\theta\ge1$, let

$$\phi_\theta(x):=1+|x|^\theta.$$

($\widetilde{\mathbf{A}}^{\sigma,b}_\theta$) Let $\theta\ge1$. We assume (49), $\sigma_t(x,\mu)=\sigma_t(x)$ and for some $(p,q),(p_1,q_1)\in\mathscr{I}_1$,

$$\sup_{Z\in\mathbf{S}_{\rm toch}}\big|\!\big|\!\big|b^Z\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^p_q(T)}<\infty,\qquad \big|\!\big|\!\big|\nabla\sigma\big|\!\big|\!\big|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(T)}<\infty,$$

and there is an $\ell\in L^q_{\rm loc}(\mathbb{R}_+)$ such that for all $\mu,\mu'\in\mathcal{P}(\mathbb{R}^d)$ and $t\ge0$,

$$\big|\!\big|\!\big|b(t,\cdot,\mu)-b\big(t,\cdot,\mu'\big)\big|\!\big|\!\big|_p\le\ell_t\big\|\phi_\theta\cdot\big(\mu-\mu'\big)\big\|_{\rm TV}. \tag{73}$$

It should be noted that [26], Theorem 6.15, implies,

$$\mathcal{W}_\theta(\mu, \mu') \le c \|\phi_\theta \cdot (\mu - \mu')\|_{\mathrm{TV}}^{1/\theta}.$$

Theorem 4.2. *Let $\theta \ge 1$ and $\beta > 2\theta$. Under $(\widetilde{\mathbf{A}}_\theta^{\sigma,b})$, for any initial random variable X_0 with finite β-order moment, there is a unique weak solution to DDSDE (1), which is also a unique strong solution.*

Proof. We use the Girsanov transform in the same way as in [21] to show the weak uniqueness, and so also the strong uniqueness. Since under the assumptions of the theorem, weak solutions are also strong solutions (see Corollary 3.5), without loss of generality, let $X^{(i)}, i = 1, 2$ be two solutions of SDE (1) defined on the same probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and with the same Brownian motion and starting point ξ. That is,

$$\mathrm{d}X_t^{(i)} = \sigma_t\big(X_t^{(i)}\big)\,\mathrm{d}W_t + b_t\big(X_t^{(i)}, \mu_t^{(i)}\big)\,\mathrm{d}t, \qquad X_0^{(i)} = \xi, \tag{74}$$

where $\mu_t^{(i)} = \mathbf{P} \circ (X_t^{(i)})^{-1}$. We want to show $\mu_t^{(1)} = \mu_t^{(2)}$.

Since $\sigma_t(x, \mu) = \sigma_t(x)$ satisfies (24) under our assumptions, it is well known that there is a unique weak solution to SDE

$$\mathrm{d}Z_t = \sigma_t(Z_t)\,\mathrm{d}W_t, \qquad Z_0 = \xi.$$

Let $\beta > 2\theta$. Since σ is bounded, it is easy to see that

$$\sup_{t\in[0,T]} \mathbf{E}|Z_t|^\beta \le C\big(\mathbf{E}|\xi|^\beta + 1\big). \tag{75}$$

Define

$$\tilde{b}_s^{(i)}(x) := \sigma_s^{-1}(x) \cdot b_s^{X^{(i)}}(x), \qquad \tilde{W}_t^{(i)} := W_t - \int_0^t \tilde{b}_s^{(i)}(Z_s)\,\mathrm{d}s$$

and

$$\mathscr{E}_T^{(i)} := \exp\bigg\{\int_0^T \tilde{b}_s^{(i)}(Z_s) \cdot \mathrm{d}W_s - \frac{1}{2}\int_0^T \big|\tilde{b}_s^{(i)}(Z_s)\big|^2\,\mathrm{d}s\bigg\}.$$

Since $|||\tilde{b}^{(i)}|||_{\widetilde{\mathbb{L}}_q^p(T)} \le |||b^{X^{(i)}}|||_{\widetilde{\mathbb{L}}_q^p(T)} < \infty$ for some $(p, q) \in \mathscr{I}_1$, by Khasminskii's estimate (60), we have

$$\mathbf{E}\exp\bigg\{\gamma\int_0^T \big|\tilde{b}_s^{(i)}(Z_s)\big|^2\,\mathrm{d}s\bigg\} \le C_{T,\gamma}, \quad \forall \gamma > 0, \tag{76}$$

and for any $\gamma \in \mathbb{R}$,

$$\mathbf{E}\big(\mathscr{E}_T^{(i)}\big)^\gamma \le C_{T,\gamma} < \infty. \tag{77}$$

Hence, for each $i = 1, 2$, $\mathbf{E}\mathscr{E}_T^{(i)} = 1$, and $\tilde{W}^{(i)}$ is still a Brownian motion under $\mathscr{E}_T^{(i)}\mathbf{P}$, and

$$\mathrm{d}Z_t = \sigma_t(Z_t)\,\mathrm{d}\tilde{W}_t^{(i)} + b_t^{X^{(i)}}(Z_t)\,\mathrm{d}t, \qquad Z_0 = \xi.$$

Since the above SDE admits a unique strong solution (see also (74)), we have

$$\big(\mathscr{E}_T^{(i)}\mathbf{P}\big) \circ Z_T^{-1} = \mathbf{P} \circ \big(X_T^{(i)}\big)^{-1} = \mu_T^{(i)}, \quad i = 1, 2.$$

Therefore, for $\delta = \frac{\beta}{\beta-\theta} < 2$, by Hölder's inequality, we get

$$\begin{aligned}\|\phi_\theta \cdot (\mu_T^{(1)} - \mu_T^{(2)})\|_{\mathrm{TV}} &= \|\phi_\theta \cdot ((\mathscr{E}_T^{(1)}\mathbf{P}) \circ Z_T^{-1} - (\mathscr{E}_T^{(2)}\mathbf{P}) \circ Z_T^{-1})\|_{\mathrm{TV}} \\ &\leq \mathbf{E}(\phi_\theta(Z_T)|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}|) \leq \|\phi_\theta(Z_T)\|_{\delta/(\delta-1)}\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta \\ &= \|1 + |Z_T|^\theta\|_{\beta/\theta}\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta \overset{(75)}{\leq} C\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta. \end{aligned} \tag{78}$$

Noting that

$$\mathrm{d}\mathscr{E}_t^{(i)} = \mathscr{E}_t^{(i)}\tilde{b}_t^{(i)}(Z_t) \cdot \mathrm{d}W_t,$$

we have

$$\mathrm{d}(\mathscr{E}_t^{(1)} - \mathscr{E}_t^{(2)}) = (\mathscr{E}_t^{(1)}\tilde{b}_t^{(1)}(Z_t) - \mathscr{E}_t^{(2)}\tilde{b}_t^{(2)}(Z_t)) \cdot \mathrm{d}W_t.$$

By Itô's formula, we have

$$\begin{aligned}\mathrm{d}|\mathscr{E}_t^{(1)} - \mathscr{E}_t^{(2)}|^2 &= |\mathscr{E}_t^{(1)}\tilde{b}_t^{(1)}(Z_t) - \mathscr{E}_t^{(2)}\tilde{b}_t^{(2)}(Z_t)|^2\,\mathrm{d}t + M_t \\ &\leq 2|\mathscr{E}_t^{(1)} - \mathscr{E}_t^{(2)}|^2|\tilde{b}_t^{(1)}(Z_t)|^2\,\mathrm{d}t + 2|\mathscr{E}_t^{(2)}(\tilde{b}_t^{(1)}(Z_t) - \tilde{b}_t^{(2)}(Z_t))|^2\,\mathrm{d}t + M_t,\end{aligned}$$

where M is a continuous local martingale. Since $\delta < 2$, by the stochastic Gronwall inequality (29) and (76), we obtain

$$\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta^2 \lesssim \int_0^T \mathbf{E}|\mathscr{E}_t^{(2)}(\tilde{b}_t^{(1)}(Z_t) - \tilde{b}_t^{(2)}(Z_t))|^2\,\mathrm{d}t.$$

Since $(p, q) \in \mathscr{I}_1$, one can choose $\gamma \in (1, 1/(d/p + 2/q))$ so that

$$(p/(2\gamma), q/(2\gamma)) \in \mathscr{I}_2.$$

Thus by Hölder's inequality and Krylov's estimate (53), we further have

$$\begin{aligned}\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta^2 &\overset{(77)}{\lesssim} \left(\int_0^T \mathbf{E}|\tilde{b}_t^{(1)}(Z_t) - \tilde{b}_t^{(2)}(Z_t)|^{2\gamma}\,\mathrm{d}t\right)^{\frac{1}{\gamma}} \\ &\lesssim \||\tilde{b}^{(1)} - \tilde{b}^{(2)}|^{2\gamma}\|^{1/\gamma}_{\tilde{\mathbb{L}}^{p/(2\gamma)}_{q/(2\gamma)}(T)} = \|\tilde{b}^{(1)} - \tilde{b}^{(2)}\|^2_{\tilde{\mathbb{L}}^p_q(T)} \\ &\lesssim \left(\int_0^T \|b_t(\cdot, \mu_t^{(1)}) - b_t(\cdot, \mu_t^{(2)})\|_p^q\,\mathrm{d}t\right)^{\frac{2}{q}} \\ &\overset{(73)}{\lesssim} \left(\int_0^T \ell_t^q\|\phi_\theta \cdot (\mu_t^{(1)} - \mu_t^{(2)})\|_{\mathrm{TV}}^q\,\mathrm{d}t\right)^{\frac{2}{q}},\end{aligned}$$

which together with (78) yields

$$\|\phi_\theta \cdot (\mu_T^{(1)} - \mu_T^{(2)})\|_{\mathrm{TV}}^q \leq C\int_0^T \ell_t^q\|\phi_\theta \cdot (\mu_t^{(1)} - \mu_t^{(2)})\|_{\mathrm{TV}}^q\,\mathrm{d}t.$$

By Gronwall's inequality, we obtain

$$\|\phi_\theta \cdot (\mu_T^{(1)} - \mu_T^{(2)})\|_{\mathrm{TV}}^q = 0 \quad \Rightarrow \quad \mu_T^{(1)} = \mu_T^{(2)}.$$

The proof is thus complete. □

5. Application to nonlinear Fokker–Planck equations

In this section, we present some applications to nonlinear Fokker–Planck equations. First of all we recall the following superposition principle: one-to-one correspondence between DDSDE (1) and nonlinear Fokker–Planck equation (2), which was first proved in [1,2], and is based on a result for linear Fokker–Planck equations due to Trevisan [25] (see also [8,15] for the special linear case where the coefficients are bounded). We repeat the argument from [1,2] here.

Theorem 5.1 (Superposition principle). *Let $\mu_t : \mathbb{R}_+ \to \mathcal{P}(\mathbb{R}^d)$ be a continuous curve such that for each $T > 0$,*

$$\int_0^T \int_{\mathbb{R}^d} \big(\big|(\sigma_t^{ik}\sigma_t^{jk})(x, \mu_t)\big| + \big|b_t(x, \mu_t)\big|\big)\mu_t(\mathrm{d}x)\,\mathrm{d}t < \infty. \tag{79}$$

Then μ_t solves the nonlinear Fokker–Planck equation (2) in the distributional sense if and only if there exists a martingale solution $\mathbb{P} \in \mathscr{M}_\nu^{\sigma,b}$ to DDSDE (1) so that for each $t > 0$,

$$\mu_t = \mathbb{P} \circ w_t^{-1}.$$

In particular, if there is at most one element in $\mathcal{M}_\nu^{\sigma,b}$ with time marginal $\mu_t := \mu_{X_t}, t \geq 0$, satisfying (79), then there is at most one solution to (2) satisfying (79).

Proof. If $\mathbb{P} \in \mathscr{M}_\nu^{\sigma,b}$ and $\mu_t = \mathbb{P} \circ w_t^{-1}$, then by (79) and Itô's formula, it is easy to see that μ_t solves (2). Now we assume μ_t solves (2). Consider the following linear Fokker–Planck equation:

$$\partial_t \tilde{\mu}_t = \big(\mathscr{L}_t^{\sigma^\mu}\big)^* \tilde{\mu}_t + \mathrm{div}\big(b_t^\mu \cdot \tilde{\mu}_t\big),$$

where $b_t^\mu(x) := b_t(x, \mu_t)$ and $\sigma_t^\mu(x) := \sigma_t(x, \mu_t)$. Since μ_t is a solution of the above linear Fokker–Planck equation, by [25], Theorem 2.5, there is a martingale solution $\mathbb{P} \in \mathscr{M}_\nu^{\sigma^\mu, b^\mu}$ so that

$$\mu_t = \mathbb{P} \circ w_t^{-1}.$$

In particular, $\mathbb{P} \in \mathscr{M}_\nu^{\sigma,b}$. The last assertion is then obvious and thus the proof is complete. □

From the above superposition principle and our well-posedness results, we can obtain the following well-posedness result about the nonlinear Fokker–Planck equations.

Theorem 5.2. *In the situations of Theorems 4.1 and 4.2, there is a unique continuous curve μ_t solving the nonlinear Fokker–Planck equation (2).*

Now we turn to the proof of Theorem 1.1.

Proof of Theorem 1.1. The existence and uniqueness of solutions to the nonlinear FPE (7) are consequences of Theorem 4.2 and Theorem 5.1. We now aim to show the existence and smoothness of the density $\rho_t^X(y)$. Let μ_t be the solution of the Fokker–Planck equation (7). We consider the following SDE:

$$\mathrm{d}X_t = b_t^\mu(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W_t, \qquad X_0 = \xi, \tag{80}$$

where $b_t^\mu(x) := \int_{\mathbb{R}^d} b_t(x,y)\mu_t(\mathrm{d}y)$. Since $b^\mu \in \widetilde{\mathbb{L}}_q^p$, where $\frac{d}{p} + \frac{2}{q} < 1$, it is well known that the operator $\Delta + b^\mu \cdot \nabla$ admits a heat kernel $\rho_{b^\mu}(s,x;t,y)$ (see [6], Theorems 1.1 and 1.3), which is continuous in $(s,x;t,y)$ on $\{(s,x;t,y) : 0 \le s < t < \infty, x, y \in \mathbb{R}^d\}$ and satisfies the following two-sided estimate: For any $T > 0$, there are constants $c_0, \gamma_0 > 1$ such that for all $0 \le s < t \le T$ and $x, y \in \mathbb{R}^d$

$$c_0^{-1}(t-s)^{-d/2}\mathrm{e}^{-\gamma_0|x-y|^2/(t-s)} \le \rho_{b^\mu}(s,x;t,y) \le c_0(t-s)^{-d/2}\mathrm{e}^{-|x-y|^2/(\gamma_0(t-s))},$$

and the gradient estimate: for some $c_1, \gamma_1 > 1$,

$$\left|\nabla_x \rho_{b^\mu}(s,x;t,y)\right| \le c_1(t-s)^{-(d+1)/2}\mathrm{e}^{-|x-y|^2/(\gamma_1(t-s))}.$$

If $\mathrm{div}_x\, b \equiv 0$, then $\rho_{b^\mu}(s,x;t,y) = \rho_{-b^\mu}(s,y;t,x)$, and so in this case,

$$\left|\nabla_y \rho_{b^\mu}(s,x;t,y)\right| \le c_1(t-s)^{-(d+1)/2}\mathrm{e}^{-|x-y|^2/(\gamma_1(t-s))}.$$

In particular, the density of the law of X_t is just given by

$$\rho_t^X(y) = \int_{\mathbb{R}^d} \rho(0,x;t,y)\big(\mathbf{P} \circ X_0^{-1}\big)(\mathrm{d}x) = \mathbf{E}\rho(0,X_0;t,y).$$

Strong uniqueness of SDE (80) ensures that $\rho_t^X(y)\,\mathrm{d}y = \mu_t(\mathrm{d}y)$. The desired estimates now follow from the above estimates. □

Acknowledgement

This work is supported by NNSFC grant of China (No. 11731009) and the DFG through the CRC 1283 “Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications.” The authors thank Dr. Xing Huang for pointing out an error in the earlier version. The referee’s useful comments are also acknowledged.

References

[1] Barbu, V. and Röckner, M. (2018). Probabilistic representation for solutions to nonlinear Fokker–Planck equations. *SIAM J. Math. Anal.* **50** 4246–4260. MR3835244 https://doi.org/10.1137/17M1162780

[2] Barbu, V. and Röckner, M. (2020). From nonlinear Fokker–Planck equations to solutions of distribution dependent SDE. *Ann. Probab.* **48** 1902–1920. MR4124528 https://doi.org/10.1214/19-AOP1410

[3] Carmona, R. and Delarue, F. (2013). Probabilistic analysis of mean-field games. *SIAM J. Control Optim.* **51** 2705–2734. MR3072222 https://doi.org/10.1137/120883499

[4] Carmona, R. and Delarue, F. (2018). *Probabilistic Theory of Mean Field Games with Applications. II. Mean Field Games with Common Noise and Master Equations. Probability Theory and Stochastic Modelling* **84**. Cham: Springer. MR3753660

[5] Carrillo, J.A., Gvalani, R.S., Pavliotis, G.A. and Schlichting, A. (2020). Long-time behaviour and phase transitions for the Mckean–Vlasov equation on the torus. *Arch. Ration. Mech. Anal.* **235** 635–690. MR4062483 https://doi.org/10.1007/s00205-019-01430-4

[6] Chen, Z.-Q., Hu, E., Xie, L. and Zhang, X. (2017). Heat kernels for non-symmetric diffusion operators with jumps. *J. Differential Equations* **263** 6576–6634. MR3693184 https://doi.org/10.1016/j.jde.2017.07.023
[7] Chiang, T.S. (1994). McKean–Vlasov equations with discontinuous coefficients. *Soochow J. Math*. **20** 507–526. MR1309485
[8] Figalli, A. (2008). Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal*. **254** 109–153. MR2375067 https://doi.org/10.1016/j.jfa.2007.09.020
[9] Funaki, T. (1984). A certain class of diffusion processes associated with nonlinear parabolic equations. *Z. Wahrsch. Verw. Gebiete* **67** 331–348. MR0762085 https://doi.org/10.1007/BF00535008
[10] Hammersley, W., Siska, D. and Szpruch, L. (2018). Mckean–Vlasov SDEs under measure dependent Lyapunov conditions. Preprint. Available at arXiv:1802.03974v2.
[11] Huang, X. and Wang, F.-Y. (2019). Distribution dependent SDEs with singular coefficients. *Stochastic Process. Appl*. **129** 4747–4770. MR4013879 https://doi.org/10.1016/j.spa.2018.12.012
[12] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Berlin: Springer. MR1943877 https://doi.org/10.1007/978-3-662-05265-5
[13] Kac, M. (1956). Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. III* 171–197. Berkeley: Univ. California Press. MR0084985
[14] Krylov, N.V. and Röckner, M. (2005). Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Related Fields* **131** 154–196. MR2117951 https://doi.org/10.1007/s00440-004-0361-z
[15] Kurtz, T.G. (2014). Weak and strong solutions of general stochastic models. *Electron. Commun. Probab.* **19** no. 58, 16. MR3254737 https://doi.org/10.1214/ECP.v19-2833
[16] Lacker, D. (2018). On a strong form of propagation of chaos for McKean–Vlasov equations. *Electron. Commun. Probab.* **23** Paper No. 45, 11. MR3841406 https://doi.org/10.1214/18-ECP150
[17] Li, J. and Min, H. (2016). Weak solutions of mean-field stochastic differential equations and application to zero-sum stochastic differential games. *SIAM J. Control Optim*. **54** 1826–1858. MR3517577 https://doi.org/10.1137/15M1015583
[18] Manita, O.A., Romanov, M.S. and Shaposhnikov, S.V. (2015). On uniqueness of solutions to nonlinear Fokker–Planck–Kolmogorov equations. *Nonlinear Anal*. **128** 199–226. MR3399526 https://doi.org/10.1016/j.na.2015.08.008
[19] Manita, O.A. and Shaposhnikov, S.V. (2013). Nonlinear parabolic equations for measures. *Algebra i Analiz* **25** 64–93. Translation in St. Petersburg Math. J. **25** (2014) 43–62. MR3113428 https://doi.org/10.1090/s1061-0022-2013-01279-9
[20] McKean, H.P. Jr. (1966). A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci. USA* **56** 1907–1911. MR0221595 https://doi.org/10.1073/pnas.56.6.1907
[21] Mishura, Y.S. and Veretennikov, A.Y. (2016). Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. Preprint. Available at arXiv:1603.02212v4.
[22] Scheutzow, M. (2013). A stochastic Gronwall lemma. *Infin. Dimens. Anal. Quantum Probab. Relat. Top*. **16** 1350019, 4. MR3078830 https://doi.org/10.1142/S0219025713500197
[23] Stroock, D.W. and Varadhan, S.R.S. (1979). *Multidimensional Diffusion Processes*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **233**. Berlin: Springer. MR0532498
[24] Sznitman, A.-S. (1991). Topics in propagation of chaos. In *École D'Été de Probabilités de Saint-Flour XIX*—1989. *Lecture Notes in Math*. **1464** 165–251. Berlin: Springer. MR1108185 https://doi.org/10.1007/BFb0085169
[25] Trevisan, D. (2016). Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electron. J. Probab.* **21** Paper No. 22, 41. MR3485364 https://doi.org/10.1214/16-EJP4453
[26] Villani, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Berlin: Springer. MR2459454 https://doi.org/10.1007/978-3-540-71050-9
[27] Vlasov, A.A. (1968). The vibrational properties of an electron gas. *Sov. Phys., Usp*. **10** 721.
[28] Wang, F.-Y. (2018). Distribution dependent SDEs for Landau type equations. *Stochastic Process. Appl*. **128** 595–621. MR3739509 https://doi.org/10.1016/j.spa.2017.05.006

[29] Xia, P., Xie, L., Zhang, X. and Zhao, G. (2020). $L^q(L^p)$-theory of stochastic differential equations. *Stochastic Process*. *Appl*. **130** 5188–5211. MR4108486 https://doi.org/10.1016/j.spa.2020.03.004
[30] Xie, L. and Zhang, X. (2020). Ergodicity of stochastic differential equations with jumps and singular coefficients. *Ann*. *Inst*. *Henri Poincaré Probab. Stat*. **56** 175–229. MR4058986 https://doi.org/10.1214/19-AIHP959
[31] Zhang, X. (2019). A discretized version of Krylov's estimate and its applications. *Electron*. *J*. *Probab*. **24** Paper No. 131, 17. MR4040991 https://doi.org/10.1214/19-ejp390
[32] Zhang, X. and Zhao, G. (2017). Heat kernel and ergodicity of SDEs with distributional drifts. Preprint. Available at arXiv:1710.10537.
[33] Zhang, X. and Zhao, G. (2018). Singular Brownian diffusion processes. *Commun*. *Math*. *Stat*. **6** 533–581. MR3877717 https://doi.org/10.1007/s40304-018-0164-7
[34] Zhang, X. and Zhao, G. (2019). Stochastic Lagrangian path for Leray solutions of 3D Navier–Stokes equations. *Commun. Math. Phys.* To appear. Available at arXiv:1904.04387.
[35] Zvonkin, A.K. (1974). A transformation of the phase space of a diffusion process that will remove the drift. *Mat*. *Sb*. (*N.S.*) **93(135)** 129–149, 152. MR0336813

Received October 2019 and revised April 2020