# WELL-POSEDNESS OF DISTRIBUTION DEPENDENT SDES WITH SINGULAR DRIFTS

MICHAEL RÖCKNER AND XICHENG ZHANG

Abstract. 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 $\mu_{X_t}$ stands for the distribution of $X_t$. In this paper for non-degenerate $\sigma$, we show the strong well-posedness of the above SDE under some integrability assumptions in the spatial variable and Lipschitz continuity in $\mu$ about $b$ and $\sigma$. In particular, we extend the results of Krylov-Röckner [15] to the distribution dependent case.

**Keywords**: Distribution dependent SDEs, McKean-Vlasov system, Zvonkin's transformation, Singular drifts, Superposition principle

**AMS 2010 Mathematics Subject Classification:** Primary: 60H10, 35K55.

## 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):

$$\mathrm{d}X_t = b_t(X_t, \mu_{X_t})\mathrm{d}t + \sigma_t(X_t, \mu_{X_t})\mathrm{d}W_t, \tag{1.1}$$

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 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbf{P})$, 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 $\mu_{X_t}$ satisfies the following non-linear Fokker-Planck equation (abbreviated as FPE) in the distributional sense:

$$\partial_t \mu_{X_t} = (\mathscr{L}_t^{\sigma^X})^* \mu_{X_t} + \mathrm{div}(b_t^X \mu_{X_t}), \tag{1.2}$$

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

$$\mathscr{L}_t^{\sigma^X} f(x) := \frac{1}{2} \sum_{i,j,k=1}^d (\sigma_t^{ik}\sigma_t^{jk})(x, \mu_{X_t})\partial_i\partial_j f(x). \tag{1.3}$$

We note that if

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

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".

then DDSDE (1.1) 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 [14, 20, 24, 4, 6], in particular, [5] and references therein).

Up to now, there are numerous papers devoted to the study of this type of nonlinear FPEs and DDSDE (1.1). In [12], Funaki showed the existence of martingale solutions for (1.1) 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 (1.2) is also obtained. More recently, under some one-side Lipschitz assumptions, Wang [28] showed the strong well-posedness and some functional inequalities to DDSDE (1.1). In [9], Hammersley, Sitsa and Szpruch proved the existence of weak solutions to SDE (1.1) 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 (1.1) with drifts that have some discontinuities. When the diffusion matrix is uniformly non-degenerate and $b, \sigma$ 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 $\sigma$ does not depend on $\mu$ and is Lipschitz continuous in $x$ and $b$ is Lipschitz continuous with respect to $\mu$ with Lipschitz constant linearly depending on $x$. 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 $\sigma$ 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 (1.2) 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 (1.1) from [19] and [18]. In [1], a technique is developed to prove weak existence of solutions to (1.1) by first solving (1.2) which works also for coefficients whose dependence on $\mu_{X_t}$ is of "Nemytskii-type", i.e., are not continuous in $\mu_{X_t}$ in the weak topology.

In this work we are interested in extending Krylov-Röckner's result [15] 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:

$$\mathrm{d}X_t = \left(\int_{\mathbb{R}^d} b_t(X_t, y)\mu_{X_t}(\mathrm{d}y)\right)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \tag{1.4}$$

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

**(H$^b$)** $|b_t(x, y)| \leqslant h_t(x - y)$ for some $h \in L^q_{loc}(\mathbb{R}_+; \widetilde{L}^p(\mathbb{R}^d))$, 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 (2.2) below.

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

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

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

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

We note that the above DDSDE is not covered by Huang and Wang's recent results [10] since $\mu \mapsto \int_{\mathbb{R}^d} b_t(x,y)\mu(\mathrm{d}y)$ is not weakly continuous. In fact, if we let

$$B_t(x,\mu) := \int_{\mathbb{R}^d} b_t(x,y)\mu(\mathrm{d}y),\ \mu\in\mathcal{P}(\mathbb{R}^d), \tag{1.5}$$

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

$$\|\!|B_t(\cdot,\mu) - B_t(\cdot,\mu')|\!\|_p \leqslant \|\!|h_t|\!\|_p \|\mu-\mu'\|_{TV}, \tag{1.6}$$

where $\|\cdot\|_{TV}$ is the total variation distance, and $\|\!|\cdot|\!\|_p$ is defined by (2.2) below.

Throughout this paper we assume $d \geqslant 2$. One of the main results of this paper is stated as follows (but see also section 4 for corresponding results when the diffusion matrix $\sigma$ 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* $\beta$*-order moment, there is a unique strong solution to SDE* (1.4)*. Moreover, the following assertions hold:*

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

$$\partial_t\mu_t = \Delta\mu_t + \mathrm{div}\left(\mu_t(b_t(x,\cdot))\mu_t\right),\quad \lim_{t\downarrow 0}\mu_t(\mathrm{d}y) = \mathbf{P}\circ X_0^{-1}(\mathrm{d}y) \tag{1.7}$$

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

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

(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* $\gamma_0, c_0 \geqslant 1$ *such that for all* $t\in(0,T]$ *and* $y\in\mathbb{R}^d$*,*

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

*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* $\mathrm{div}b = 0$*, 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* $\gamma_1, c_1\geqslant 1$ *such that for all* $t\in(0,T]$ *and* $y\in\mathbb{R}^d$*,*

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

**Example 1.2.** *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* $\kappa>0$*,*

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

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

**Remark 1.3.** *Here an open question is to show the following propagation of chaos (see* [24]*): Given $N \in \mathbb{N}$, let $X^{N,j}, j = 1, \cdots, N$ solve the following SDEs*

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

*where $W_\cdot^j, j = 1, \cdots, N$ are $N$-independent $d$-dimensional Brownian motion. Let $X$ be the unique solution of SDE* (1.4) *in Theorem 1.1. Is it possible to show that*

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

*Even for bounded measurable $b$, the above question seems to be still open.*

To show the existence of a solution to DDSDE (1.4), by the well-known result for bounded measurable drift $b$ obtained in [21] (see also [17]and [32]), for each $n \in \mathbb{N}$, there is a solution to the following distribution dependent SDE:

$$\mathrm{d}X_t^n = \left(\int_{\mathbb{R}^d} b_t^n(X_t^n, y)\mu_{X_t^n}(\mathrm{d}y)\right)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \quad X_0^n = X_0, \tag{1.8}$$

where $b_t^n(x,y) := (-n)\vee b_t(x,y)\wedge n$. 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 $f \in \widetilde{\mathbb{L}}_{q_1}^{p_1}(T)$,

$$\sup_n \mathbf{E}\left(\int_0^T f_t(X_t^n)\mathrm{d}t\right) \leqslant C_T \|\!| f |\!\|_{\widetilde{\mathbb{L}}_{q_1}^{p_1}(T)}. \tag{1.9}$$

By this estimate and Zvonkin's technique, we can further show the tightness of $X_\cdot^n$ 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 $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$,

$$\sup_n \mathbf{E}\left(\int_0^t f_s(X_s^n, \tilde{X}_s^n)\mathrm{d}s\right) \leqslant \|\!| f |\!\|_{\widetilde{\mathbb{L}}_{q_0}^{p_0}(T)},$$

where $\tilde{X}_\cdot^n$ 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 2.7 below). Moreover, concerning the uniqueness, under assumption (1.6), 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 (1.1) when the drift satisfies **(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 (1.1) in two cases: the coefficients $b$ and $\sigma$ 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 (1.2) and prove Theorem 1.1.

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

- For $\theta > 0$, $\mathcal{P}_\theta(\mathbb{R}^d) := \big\{\mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^\theta \mu(\mathrm{d}x) < \infty\big\}$.
- 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 $\mathbf{S}_{\rm toch}$ 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 $X \in \mathbf{S}_{\rm toch}$, define
$$b_t^X(x) := b_t(x, \mu_{X_t}), \quad \mu_{X_t} := \mathbf{P} \circ X_t^{-1}. \tag{1.10}$$
- For a signed measure $\mu$, we denote by $\|\mu\|_{TV} := \sup_{\|f\|_\infty \leqslant 1} |\mu(f)|$ the total variation of $\mu$.
- For $j = 1, 2$, we introduce the index set $\mathscr{I}_j$ as following:
$$\mathscr{I}_j := \Big\{(p, q) \in (1, \infty) : \tfrac{d}{p} + \tfrac{2}{p} < j\Big\}. \tag{1.11}$$
- For a matrix $\sigma$, we use $\|\sigma\|_{HS}$ to denote the Hilbert-Schmidt norm of $\sigma$.
- We use $A \lesssim B$ (resp. $\asymp$) to denote $A \leqslant CB$ (resp. $C^{-1}B \leqslant A \leqslant CB$) for some unimportant constant $C \geqslant 1$, 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

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

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

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

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

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

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

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

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

$$\chi^z_r(x) := \chi((x - z)/r). \tag{2.1}$$

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

$$\widetilde{H}^{\alpha,p} := \Big\{f \in H^{\alpha,p}_{loc}(\mathbb{R}^d), \|\!|f|\!\|_{\alpha,p} := \sup_z \|f\chi^z_r\|_{\alpha,p} < \infty\Big\}, \tag{2.2}$$

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

$$\|\!|f|\!\|_{\widetilde{\mathbb{H}}^{\alpha,p}_q(S,T)} := \sup_{z\in\mathbb{R}^d} \|\chi^z_r f\|_{\mathbb{H}^{\alpha,p}_q(S,T)} < \infty. \tag{2.3}$$

For simplicity we shall write

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

and

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

The following lemma list some easy properties of $\widetilde{\mathbb{H}}^{\alpha,p}_q$ (see [35] 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 $C=C(d,\alpha,r,r',p,q)\geqslant 1$ such that*

$$C^{-1}\sup_z\|f\chi^z_{r'}\|_{\mathbb{H}^{\alpha,p}_q(T)}\leqslant\sup_z\|f\chi^z_r\|_{\mathbb{H}^{\alpha,p}_q(T)}\leqslant C\sup_z\|f\chi^z_{r'}\|_{\mathbb{H}^{\alpha,p}_q(T)}. \tag{2.4}$$

*In other words, the definition of $\widetilde{\mathbb{H}}^{\alpha,p}_q$ 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 $C=C(d,\alpha,p,p')>0$,*

$$[\![f]\!]_{\widetilde{\mathbb{L}}^{p'}_q(T)}\leqslant C[\![f]\!]_{\widetilde{\mathbb{H}}^{\alpha,p}_q(T)}. \tag{2.5}$$

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

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

(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 $f\in\widetilde{\mathbb{H}}^{\alpha,p}_q$, it holds that $f_\varepsilon\in L^q_{loc}(\mathbb{R};C^\infty_b(\mathbb{R}^d))$ and for some $C=C(d,\alpha,p,q)>0$,*

$$[\![f_\varepsilon]\!]_{\widetilde{\mathbb{H}}^{\alpha,p}_q(T)}\leqslant C[\![f]\!]_{\widetilde{\mathbb{H}}^{\alpha,p}_q(T)},\ \forall\varepsilon\in(0,1), \tag{2.6}$$

*and for any $\varphi\in C^\infty_c(\mathbb{R}^d)$,*

$$\lim_{\varepsilon\to0}\|(f_\varepsilon-f)\varphi\|_{\mathbb{H}^{\alpha,p}_q(T)}=0. \tag{2.7}$$

(v) *For $r=p/(p-1)$ and $s=q/(q-1)$,*

$$[\![f]\!]_{\widetilde{\mathbb{L}}^p_q(T)}\asymp[\![f]\!]'_{\widetilde{\mathbb{L}}^p_q(T)}=\sup_{[\![g]\!]^*_{\widetilde{\mathbb{L}}^r_s(T)}\leqslant1}\left|\int_0^T\!\!\int_{\mathbb{R}^d}f_t(x)g_t(x)\mathrm{d}x\mathrm{d}t\right|, \tag{2.8}$$

$$\text{and }[\![g]\!]^*_{\widetilde{\mathbb{L}}^r_s(T)}=\sup_{[\![f]\!]'_{\widetilde{\mathbb{L}}^p_q(T)}\leqslant1}\left|\int_0^T\!\!\int_{\mathbb{R}^d}f_t(x)g_t(x)\mathrm{d}x\mathrm{d}t\right|, \tag{2.9}$$

*where $[\![f]\!]'_{\widetilde{\mathbb{L}}^p_q(T)}:=\sup_{z\in\mathbb{Z}^d}\|\mathbf{1}_{Q_z}f\|_{\mathbb{L}^p_q(T)}$ and $[\![g]\!]^*_{\widetilde{\mathbb{L}}^r_s(T)}:=\sum_{z\in\mathbb{Z}^d}\|\mathbf{1}_{Q_z}g\|_{\mathbb{L}^r_s(T)}$,*

$$Q_z:=\Pi^d_{i=1}(z_i,z_i+1],\quad z=(z_1,\cdots,z_d)\in\mathbb{Z}^d. \tag{2.10}$$

*Proof.* The first four conclusions can be found in [35, Proposition 4.1]. We only prove (v). The equivalence between $[\![f]\!]_{\widetilde{\mathbb{L}}^p_q(T)}$ and $[\![f]\!]'_{\widetilde{\mathbb{L}}^p_q(T)}$ is obvious by definition. Concerning the others, we note that by Hölder's inequality,

$$\begin{aligned}\int_0^T\!\!\int_{\mathbb{R}^d}f_t(x)g_t(x)\mathrm{d}x\mathrm{d}t&=\sum_{z\in\mathbb{Z}^d}\int_0^T\!\!\int_{\mathbb{R}^d}\mathbf{1}_{Q_z}(x)f_t(x)g_t(x)\mathrm{d}x\mathrm{d}t\\&\leqslant\sum_{z\in\mathbb{Z}^d}\|\mathbf{1}_{Q_z}f\|_{\mathbb{L}^p_q(T)}\|\mathbf{1}_{Q_z}g\|_{\mathbb{L}^r_s(T)}\leqslant[\![f]\!]'_{\widetilde{\mathbb{L}}^p_q(T)}[\![g]\!]^*_{\widetilde{\mathbb{L}}^r_s(T)}.\end{aligned} \tag{2.11}$$

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

$$\lim_{n\to\infty}\|\mathbf{1}_{Q_n}f\|_{\mathbb{L}^p_q(T)}=[\![f]\!]'_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{2.12}$$

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 $\|g\|^*_{\widetilde{\mathbb{L}}^r_s(T)} = 1$ and

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

which together with (2.11) and (2.12) yields (2.8). 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 $\|f\|'_{\widetilde{\mathbb{L}}^p_q(T)} = 1$ and

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

which together with (2.11) yields (2.9). □

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 (1.11)) and $T > 0$. Assume that $\sigma_t(x,\mu) = \sigma_t(x)$ and $b_t(x,\mu) = b_t(x)$ are independent of $\mu$, and satisfy that for some $c_0 \geqslant 1$, $\gamma \in (0,1]$ and for all $t \geqslant 0$, $x, y, \xi \in \mathbb{R}^d$,*

$$c_0^{-1}|\xi| \leqslant |\sigma_t(x)\xi| \leqslant c_0|\xi|, \quad \|\sigma_t(x) - \sigma_t(y)\|_{HS} \leqslant c_0|x-y|^\gamma, \tag{2.13}$$

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

$$\partial_t u + (\mathscr{L}_t^\sigma - \lambda)u + b\cdot\nabla u = f, \quad u(T,x) = 0. \tag{2.14}$$

*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 $\lambda \geqslant 1$,*

$$\lambda^{1-\frac{\alpha}{2}-\frac{1}{q}}\|u\|_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} + \|u\|_{\widetilde{\mathbb{H}}^{2,p}_q(T)} \leqslant C\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{2.15}$$

(ii) *Let $(\sigma', b', f')$ be another set of coefficients satisfying the same assumptions as $(\sigma, b, f)$ with the same parameters $(\gamma, c_0, \kappa_0)$. Let $u'$ be the solution of (2.14) corresponding to $(\sigma', b', f')$. For any $\alpha \in [0, 2-\frac{2}{q})$, there is a constant $c_2 = c_2(\alpha,\Theta) > 0$ such that for all $\lambda \geqslant 1$,*

$$\begin{aligned}\lambda^{1-\frac{\alpha}{2}-\frac{1}{q}}\|u-u'\|_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} &\leqslant c_2\|f-f'\|_{\widetilde{\mathbb{L}}^p_q(T)} \\ &\quad + c_2\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}\big(\|\sigma-\sigma'\|_{\mathbb{L}^\infty(T)} + \|b-b'\|_{\widetilde{\mathbb{L}}^p_q(T)}\big).\end{aligned} \tag{2.16}$$

*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

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

By (2.15) and Hölder's inequality we have

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

Estimate (2.16) now follows by Sobolev's embedding (2.5) due to $\frac{d}{p}+\frac{2}{q}<1$ and (2.15). □

**Remark 2.3.** *It should be noted that if $b$ is bounded measurable, then the assertions in Theorem 2.2 holds 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 2.4** (Stochastic Gronwall's inequality)**.** *Let $\xi(t)$ and $\eta(t)$ be two non-negative 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*

$$\xi(t)\leqslant\eta(t)+\int_0^t\xi(s)\mathrm{d}A_s+M_t,\ \forall t\geqslant 0. \tag{2.17}$$

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

$$\big[\mathbf{E}(\xi(\tau)^*)^q\big]^{1/q}\leqslant\Big(\tfrac{p}{p-q}\Big)^{1/q}\Big(\mathbf{E}\mathrm{e}^{pA_\tau/(1-p)}\Big)^{(1-p)/p}\mathbf{E}\big(\eta(\tau)^*\big), \tag{2.18}$$

*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 2.5.** (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 $\nabla f\in L^1_{loc}(\mathbb{R}^d)$ and Lebesgue-almost all $x,y\in\mathbb{R}^d$,*

$$|f(x)-f(y)|\leqslant C|x-y|(\mathcal{M}_R|\nabla f|(x)+\mathcal{M}_R|\nabla f|(y)+\|f\|_\infty), \tag{2.19}$$

*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 $f\in\widetilde{\mathbb{L}}^p_q(T)$,*

$$\|\mathcal{M}_R f\|_{\widetilde{\mathbb{L}}^p_q(T)}\leqslant C\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{2.20}$$

We introduce the following notion about Krylov's estimates.

**Definition 2.6.** *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 $\kappa$ if for any $f\in\widetilde{\mathbb{L}}^p_q(T)$,*

$$\mathbf{E}\left(\int_0^T f_t(X_t)\mathrm{d}t\right)\leqslant\kappa\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{2.21}$$

*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

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

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

**Lemma 2.7.** *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 $f_t(x,y) \in \widetilde{\mathbb{L}}^{p_1,p_2}_{q_0}(T)$,*

$$\mathbf{E}\left(\int_0^T f_t(X_t, Y_t)\mathrm{d}t\right) \leqslant \kappa_1\kappa_2 |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^{p_1,p_2}_{q_0}(T)}. \tag{2.22}$$

*Proof.* Let $Z^1 = X$ and $Z^2 = Y$. First of all, by Krylov's estimate (2.21), 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

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

By Proposition 2.1 (v), we further have

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

where $Q_z$ is defined by (2.10). Now by the independence of $X, Y$ and Hölder's inequality, we have

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

which gives (2.22). 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 2.8.** *Let $X^n, Y^n, X, Y \in \mathbf{S}_{\mathrm{toch}}$ be such that for each $t \geqslant 0$, $X^n_t$ converges to $X_t$ almost surely and $Y^n_t$ 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 $C_1 > 0$,*

$$\sup_n \sup_{t\in[0,T]} \mathbf{E}|X^n_t|^\beta \leqslant C_1. \tag{2.23}$$

*If for each $(t,x)$, $\mu \mapsto b_t(x,\mu)$ is continuous with respect to the weak convergence topology and for some $\gamma > 1$, $C_2 > 0$ and all $Z \in \mathbf{S}_{\mathrm{toch}}$,*

$$|\!|\!| b^Z |\!|\!|_{\mathbb{L}^{\gamma p}_{\gamma q}(T)} \leqslant C_2, \tag{2.24}$$

*where $b^Z$ is defined by (1.10), then*

$$\lim_{n\to\infty} \mathbf{E}\left(\int_0^T |b^{Y_n}_t(X^n_t) - b^Y_t(X_t)|\mathrm{d}t\right) = 0. \tag{2.25}$$

*Proof.* To prove (2.25), it suffices to show the following:

$$\lim_{n\to\infty}\mathbf{E}\left(\int_0^T\left|b_t^{Y^n}(X_t^n)-b_t^Y(X_t^n)\right|\mathrm{d}t\right)=0, \tag{2.26}$$

$$\lim_{n\to\infty}\mathbf{E}\left(\int_0^T\left|b_t^{Y}(X_t^n)-b_t^Y(X_t)\right|\mathrm{d}t\right)=0. \tag{2.27}$$

We first look at (2.26). Since $\mu_{Y_t^n}$ weakly converges to $\mu_{Y_t}$ for each $t\geqslant 0$, by the assumption we have

$$b_t^{Y_n}(x)\overset{n\to\infty}{\to} b_t^Y(x),\quad \forall (t,x)\in\mathbb{R}_+\times\mathbb{R}^d. \tag{2.28}$$

For fixed $R,M>0$, since $X^n\in\mathbf{K}_{T,\kappa}^{p,q}$ (see (2.21)), by the definitions we have

$$\begin{aligned}
&\mathbf{E}\left(\int_0^T\mathbf{1}_{B_R}(X_t^n)\left|b_t^{Y_n}(X_t^n)-b_t^Y(X_t^n)\right|\mathrm{d}t\right)\leqslant\kappa\|\mathbf{1}_{B_R}(b^{Y_n}-b^Y)\|_{\widetilde{\mathbb{L}}_q^p(T)}\\
&\lesssim\left\|\mathbf{1}_{B_R}(b^{Y_n}-b^Y)\mathbf{1}_{|b^{Y_n}-b^Y|\leqslant M}\right\|_{\mathbb{L}_q^p(T)}+\left\|\mathbf{1}_{B_R}(b^{Y_n}-b^Y)\mathbf{1}_{|b^{Y_n}-b^Y|>M}\right\|_{\mathbb{L}_q^p(T)}\\
&\leqslant\left\|\mathbf{1}_{B_R}(b^{Y_n}-b^Y)\mathbf{1}_{|b^{Y_n}-b^Y|\leqslant M}\right\|_{\mathbb{L}_q^p(T)}+\left\|\mathbf{1}_{B_R}|b^{Y_n}-b^Y|^\gamma\right\|_{\mathbb{L}_q^p(T)}/M^{\gamma-1}.
\end{aligned}$$

By the dominated convergence theorem and (2.28), the first term converges to zero as $n\to\infty$ for each $M>0$. By (2.24), the second term converges to zero uniformly in $n$ as $M\to\infty$. Thus, we obtain that for any $R>0$,

$$\lim_{n\to\infty}\mathbf{E}\left(\int_0^T\mathbf{1}_{B_R}(X_t^n)\left|b_t^{Y_n}(X_t^n)-b_t^Y(X_t^n)\right|\mathrm{d}t\right)=0. \tag{2.29}$$

On the other hand, by Hölder and Chebyshev's inequalities and (2.23), we have

$$\begin{aligned}
&\mathbf{E}\left(\int_0^T\mathbf{1}_{B_R^c}(X_t^n)\left|b_t^{Y_n}(X_t^n)-b_t^Y(X_t^n)\right|\mathrm{d}t\right)\\
&\leqslant\int_0^T\mathbf{P}(|X_t^n|>R)^{\frac{\gamma-1}{\gamma}}\left(\mathbf{E}\left|b_t^{Y_n}(X_t^n)-b_t^Y(X_t^n)\right|^\gamma\right)^{\frac{1}{\gamma}}\mathrm{d}t\\
&\leqslant\sup_{t\in[0,T]}\mathbf{P}(|X_t^n|>R)^{\frac{\gamma-1}{\gamma}}T^{\frac{\gamma-1}{\gamma}}\left(\int_0^T\mathbf{E}\left|b_t^{Y_n}(X_t^n)-b_t^Y(X_t^n)\right|^\gamma\mathrm{d}t\right)^{\frac{1}{\gamma}}\\
&\leqslant\left(\frac{C_1T}{R^\beta}\right)^{\frac{\gamma-1}{\gamma}}\kappa^{\frac{1}{\gamma}}\|b^{Y_n}-b^Y\|_{\mathbb{L}_{\gamma q}^{\gamma p}(T)}\overset{(2.24)}{\leqslant}\left(\frac{C_1T}{R^\beta}\right)^{\frac{\gamma-1}{\gamma}}\kappa^{\frac{1}{\gamma}}\cdot 2C_2.
\end{aligned}$$

Combining this with (2.29), we obtain (2.26).

Next we show (2.27). Let $b_t^{Y,\varepsilon}(x):=b_t^Y(\cdot)*\varrho_\varepsilon(x)$ be a mollifying approximation of $b^Y$. By Proposition 2.1 (iv) and (2.23), as above one can derive that

$$\lim_{\varepsilon\to 0}\sup_{n\in\mathbb{N}\cup\{\infty\}}\mathbf{E}\left(\int_0^T\left|b_t^{Y,\varepsilon}(X_t^n)-b_t^Y(X_t^n)\right|\mathrm{d}t\right)=0, \tag{2.30}$$

where we have used the convention $X^\infty:=X$. On the other hand, since by (2.21),

$$\sup_n\mathbf{E}\left(\int_0^T\left|b_t^{Y,\varepsilon}(X_t^n)-b_t^{Y,\varepsilon}(X_t)\right|^\gamma\mathrm{d}t\right)\leqslant C\|b^{Y,\varepsilon}\|^\gamma_{\widetilde{\mathbb{L}}_{\gamma q}^{\gamma p}(T)},$$

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

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

which together with (2.30) yields (2.27). □

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:

$$b_t(x,\mu) = \int_{\mathbb{R}^d} \bar{b}_t(x,y)\mu(\mathrm{d}y), \tag{2.31}$$

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 2.9.** *Let $X^n, Y^n, X, Y \in \mathbf{S}_{\mathrm{toch}}$ be such that for each $t \geqslant 0$, $X_t^n$ converges to $X_t$ almost surely and $Y_t^n$ 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 $X^n \in \mathbf{K}_{T,\kappa}^{p_1,q_1}$ and $Y^n \in \mathbf{K}_{T,\kappa}^{p_2,q_2}$ for each $n \in \mathbb{N}$, and that there is a constant $C_1 > 0$ such that*

$$\sup_n \sup_{t\in[0,T]} \mathbf{E}\left(|X_t^n|^\beta + |Y_t^n|^\beta\right) \leqslant C_1. \tag{2.32}$$

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

$$\lim_{n\to\infty} \mathbf{E}\left(\int_0^T |b_t^{Y^n}(X_t^n) - b_t^Y(X_t)|\mathrm{d}t\right) = 0. \tag{2.33}$$

*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 $(X_t^n, Y_t^n) \to (X_t, Y_t)$ a.e. as $n \to \infty$ for each $t$. Notice that by the assumptions and (2.22),

$$\sup_{n\in\mathbb{N}_\infty} \mathbf{E}\left(\int_0^T |\bar{b}_t(X_t^n, Y_t^n)|^\gamma \mathrm{d}t\right) \leqslant \kappa^2 \|\bar{b}\|_{\widetilde{\mathbb{L}}_{\gamma q_0}^{\gamma p_1, \gamma p_2}(T)}^\gamma < \infty. \tag{2.34}$$

Let $\bar{b}_t^\varepsilon(x,y) = \bar{b}_t * \varrho_\varepsilon(x,y)$ be a mollifying approximation of $\bar{b}$. As in the proof of (2.26), we have

$$\lim_{\varepsilon\to 0} \sup_{n\in\mathbb{N}_\infty} \mathbf{E}\left(\int_0^T |\bar{b}_t^\varepsilon(X_t^n, Y_t^n) - \bar{b}_t(X_t^n, Y_t^n)|\mathrm{d}t\right) = 0. \tag{2.35}$$

Thus, to prove (2.33), it suffices to show that for fixed $\varepsilon \in (0,1)$,

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

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

which follows by (2.34) 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 (1.1). Let $\mathbb{C}$ be the space of all continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^d$, which is endowed with the usual Borel $\sigma$-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 $t \geqslant 0$, let $\mathcal{B}_t(\mathbb{C}) = \sigma\{w_s : s \leqslant t\}$ 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* (1.1) *with initial distribution $\nu \in \mathcal{P}(\mathbb{R}^d)$ if $\mathbb{P} \circ w_0^{-1} = \nu$ and for any $f \in C^\infty(\mathbb{R}^d)$,*

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

*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*

$$M_t^f := f(w_t) - f(w_0) - \int_0^t (\mathscr{L}_s^{\sigma^{\mathbb{P}}} f)(w_s)\mathrm{d}s - \int_0^t (b_s^{\mathbb{P}} \cdot \nabla f)(w_s)\mathrm{d}s, \tag{3.1}$$

*is a continuous local $\mathcal{B}_t(\mathbb{C})$-martingale under $\mathbb{P}$. All the martingale solutions of DDSDE* (1.1) *with coefficients $\sigma, b$ and initial distribution $\nu$ 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 $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbf{P})$. We call*

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

*a weak solution of DDSDE* (1.1) *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*

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

*and*

$$X_t = X_0 + \int_0^t b_s(X_s, \mu_{X_s})\mathrm{d}s + \int_0^t \sigma_s(X_s, \mu_{X_s})\mathrm{d}W_s, \quad \mathbf{P} - a.s. \tag{3.2}$$

**Remark 3.3.** 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

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

to DDSDE (1.1) 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 $\sigma$ and $b$:

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

$$c_0^{-1}|\xi|\leqslant|\sigma_t(x,\mu)\xi|\leqslant c_0|\xi|,\ \|\sigma_t(x,\mu)-\sigma_t(x',\mu)\|_{HS}\leqslant c_0|x-x'|^\gamma. \quad (3.3)$$

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 $\kappa_0>0$,

$$\sup_{Z\in\mathbf{S}_{\rm toch}}\|\!|b^Z|\!\|_{\widetilde{\mathbb{L}}^p_q(T)}\leqslant\kappa_0<\infty. \quad (3.4)$$

(ii) $b$ has the form (2.31) with $\bar{b}$ satisfying **($\mathbf{H}^b$)**.

It should be noticed that under **($\mathbf{H}^b$)**, (3.4) holds. Indeed, by definition we have

$$\begin{aligned}\|\!|b^Z|\!\|^q_{\widetilde{\mathbb{L}}^p_q(T)}&=\sup_{z\in\mathbb{R}^d}\int_0^T\left\|\chi^z_r\int_{\mathbb{R}^d}\bar{b}_s(\cdot,y)\mu_{Z_s}(\mathrm{d}y)\right\|^q_p\mathrm{d}s\\&\leqslant\sup_{z\in\mathbb{R}^d}\int_0^T\left\|\chi^z_r\int_{\mathbb{R}^d}h_s(\cdot-y)\mu_{Z_s}(\mathrm{d}y)\right\|^q_p\mathrm{d}s\\&\leqslant\sup_{z\in\mathbb{R}^d}\int_0^T\int_{\mathbb{R}^d}\left\|\chi^{z-y}_r h_s\right\|^q_p\mu_{Z_s}(\mathrm{d}y)\mathrm{d}s\\&\leqslant\int_0^T\sup_{z\in\mathbb{R}^d}\|\chi^z_r h_s\|^q_p\,\mu_{Z_s}(\mathbb{R}^d)\mathrm{d}s=\int_0^T\|\!|h_s|\!\|^q_p\mathrm{d}s.\end{aligned}$$

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

**Lemma 3.4.** *Let $\beta>0$. Under **($\mathbf{H}^{\sigma,b}$)**, for any $\nu\in\mathcal{P}_\beta(\mathbb{R}^d)$ and $Z\in\mathbf{S}_{\rm toch}$, there is a unique weak solution $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geqslant0},\mathbf{P};X,W)$ to the following SDE:*

$$\mathrm{d}X_t=b^Z_t(X_t)\mathrm{d}t+\sigma^Z_t(X_t)\mathrm{d}W_t,\quad\mathbf{P}\circ X_0^{-1}=\nu.$$

*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*

$$\mathbf{E}\left(\sup_{t\in[0,T]}|X_t|^\beta\right)\leqslant C_1(\mathbf{E}|X_0|^\beta+1), \quad (3.5)$$

*and for any $\delta<T$,*

$$\mathbf{E}\left(\sup_{t\in[0,T-\delta]}|X_{t+\delta}-X_t|^\beta\right)\leqslant C_1\delta^{\beta/2}. \quad (3.6)$$

(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 $0\leqslant t_0<t_1\leqslant T$ and $f\in\widetilde{\mathbb{L}}^{p_1}_{q_1}(t_0,t_1)$,*

$$\mathbf{E}\left(\int_{t_0}^{t_1}f_s(X_s)\mathrm{d}s\Big|\mathscr{F}_{t_0}\right)\leqslant C_2\|\!|f|\!\|_{\widetilde{\mathbb{L}}^{p_1}_{q_1}(t_0,t_1)}. \quad (3.7)$$

*Proof.* The proof of this lemma is essentially contained in [33]. For the reader's convenience, we sketch the proofs below. We use Zvonkin's transformation to kill the drift $b^Z$. For $\lambda,T>0$, consider the following backward PDE:

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

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

$$\lambda^{1-\frac{\alpha}{2}-\frac{2}{q}} \|u\|_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} + \|u\|_{\widetilde{\mathbb{H}}^{2,p}_q(T)} \leqslant c_1 \|b^Z\|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{3.8}$$

In particular, since $\frac{d}{p} + \frac{2}{q} < 1$, by (2.5) we can choose $\lambda$ large enough so that

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

Now if we define

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

then it is easy to see that

$$|x-y|/2 \leqslant |\Phi_t(x) - \Phi_t(y)| \leqslant 2|x-y|, \tag{3.9}$$

and

$$\partial_t \Phi + \mathscr{L}^{\sigma^Z}_t \Phi + b^Z \cdot \nabla \Phi = \lambda u. \tag{3.10}$$

By the generalized Itô formula and (3.10), we have

$$\begin{aligned} Y_t := \Phi_t(X_t) &= \Phi_0(X_0) + \lambda \int_0^t u_s(X_s)\mathrm{d}s + \int_0^t (\sigma^Z_s \cdot \nabla \Phi_s)(X_s)\mathrm{d}W_s, \\ &= \Phi_0(X_0) + \int_0^t \tilde{b}_s(Y_s)\mathrm{d}s + \int_0^t \tilde{\sigma}_s(Y_s)\mathrm{d}W_s, \end{aligned} \tag{3.11}$$

where

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

Moreover, by (3.8), (3.9) and the Sobolev embedding (2.5), it is easy to see that for some $c_2 = c_2(\Theta, T) > 0$ and $\gamma_0 = \gamma_0(\gamma, p, q) \in (0,1)$,

$$c_2^{-1}|\xi| \leqslant |\tilde{\sigma}_t(x)\xi| \leqslant c_2|\xi|, \quad \|\tilde{\sigma}_t(x) - \tilde{\sigma}_t(y)\|_{HS} \leqslant c_2|x-y|^{\gamma_0}, \tag{3.12}$$

and

$$\|\tilde{b}\|_{\mathbb{L}^\infty(T)} + \|\nabla \tilde{b}\|_{\mathbb{L}^\infty(T)} \leqslant 4\lambda. \tag{3.13}$$

By well-known results, SDE (3.11) admits a unique weak solution (cf. [23]). Moreover, as in [33], one can check that $X_t := \Phi_t^{-1}(Y_t)$ solves the original SDE.

(i) Let $\beta > 0$. By (3.12) and (3.13), estimate (3.5) directly follows by BDG's inequality. We prove (3.6). Fix $\delta \in (0,T)$. Let $\tau$ be any stopping time less than $T - \delta$. By equation (3.11) and BDG's inequality, we have

$$\begin{aligned} \mathbf{E}|Y_{\tau+\delta} - Y_\tau|^\beta &\lesssim \mathbf{E}\left|\int_\tau^{\tau+\delta} \tilde{b}_s(X_s)\mathrm{d}s\right|^\beta + \mathbf{E}\left|\int_\tau^{\tau+\delta} \tilde{\sigma}_s(X_s)\mathrm{d}W_s\right|^\beta \\ &\lesssim \|\tilde{b}\|^\beta_{\mathbb{L}^\infty(T)}\delta^\beta + \|\tilde{\sigma}\|^\beta_{\mathbb{L}^\infty(T)}\delta^{\beta/2} \leqslant C\delta^{\beta/2}, \end{aligned}$$

which yields (3.6) by [34, Lemma 2.7] and (3.9).

(ii) It was proved in [31, Theorem 2.1] (see also [30, Theorem 5.7]) 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 $0 \leqslant t_0 < t_1 \leqslant T$ and $f \in \widetilde{\mathbb{L}}^{p_1}_{q_1}(t_0, t_1)$,

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

By a change of variable and (3.9) again, we obtain (3.7). □

**Remark 3.5.** *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 $(p_1, q_1) \in \mathscr{I}_2$,*

$$\mathbf{E}\exp\left(\lambda\int_0^T |f_s(X_s)|\mathrm{d}s\right) \leqslant C_3, \tag{3.14}$$

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

Now we can show the following weak existence result.

**Theorem 3.6.** *Let $\beta > 2$. Under $(\mathbf{H}^{\sigma,b})$, for any $\nu \in \mathcal{P}_\beta(\mathbb{R}^d)$, there exists a weak solution $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbf{P}; X, W)$ to DDSDE (1.1) with $\mathbf{P}\circ X_0^{-1} = \nu$.*

*Proof.* Let $X_t^0 \equiv X_0$. For $n \in \mathbb{N}$, consider the following approximating SDE:

$$X_t^n = X_0^n + \int_0^t b_s^n(X_s^n, \mu_{X_s^n})\mathrm{d}s + \int_0^t \sigma_s(X_s^n, \mu_{X_s^n})\mathrm{d}W_s, \tag{3.15}$$

where

$$b_s^n(x,\mu) := (-n)\vee b_s(x,\mu)\wedge n, \quad \bar{b}_s^n(x,y) := (-n)\vee \bar{b}_s(x,y)\wedge n.$$

Since $b^n$ is bounded measurable, by [21] or [32, Theorem 1.2], there is a weak solution

$$(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbf{P}; X^n, W)$$

to DDSDE (3.15) with $\mathbf{P}\circ(X_0^n)^{-1} = \nu$. Moreover, since

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

by Lemma 3.4, the following uniform estimates hold:

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

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

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

$$\sup_n \mathbf{E}\left(\sup_{t\in[0,T-\delta]}|X_{t+\delta}^n - X_t^n|^\beta\right) \leqslant C_1\delta^{\beta/2}.$$

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

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

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

$$(\tilde{X}^n, \tilde{W}^n) \to (\tilde{X}, \tilde{W}), \quad \tilde{\mathbf{P}} - a.s. \tag{3.16}$$

and

$$\tilde{\mathbf{P}}\circ(\tilde{X}^n,\tilde{W}^n)^{-1}=\mathbb{Q}^n=\mathbf{P}\circ(X^n,W)^{-1},\quad \tilde{\mathbf{P}}\circ(\tilde{X},\tilde{W})^{-1}=\mathbb{Q}. \tag{3.17}$$

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

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

In other words, $\tilde{W}^n$ is an $\tilde{\mathscr{F}}^n_t$-Brownian motion. Thus, by (3.15) and (3.17) we have

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

By (ii), (3.16), Lemmas 2.8, 2.9 and [13, Theorem 6.22, p383], 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 2.9 are satisfied in the case that $b$ takes the form (2.31) with $\bar{b}$ satisfying **(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 $|\bar{b}_t(x,y)|\leqslant h_t(x-y)$, where for some $(p,q)\in\mathscr{I}_1$, $h\in L^q_{loc}(\mathbb{R}_+;\widetilde{L}^p(\mathbb{R}^d))\subset\widetilde{\mathbb{L}}^p_q$. One can choose $\gamma>1$ so that $\frac{d\gamma}{p}+\frac{\gamma}{q}<1$. Now if we take $p_1=p_2=\frac{p}{\gamma}$, $q_0=\frac{q}{\gamma}$, $q_1=q_2=\frac{2q}{q+\gamma}$, then it is easy to see that $(p_1,q_1)\in\mathscr{I}_2$ and

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

Thus one can apply Lemma 2.9 to conclude that

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

Moreover, as in showing (2.35), we also have

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

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

About the existence of strong solutions, we have

**Corollary 3.7.** *Let $\beta>2$. Under* **(H$^{\sigma,b}$)**, *if for some $(p_1,q_1)\in\mathscr{I}_1$,*

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

*then for any initial random variable $X_0$ with finite $\beta$-order moment, there exists a strong solution to DDSDE* (1.1).

*Proof.* Let $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geqslant 0},\mathbf{P};X,W)$ be a weak solution of DDSDE (1.1). Define

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

Consider the following SDE:

$$\mathrm{d}Z_t=b^X_t(Z_t)\mathrm{d}t+\sigma^X_t(Z_t)\mathrm{d}W_t.$$

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 3.8.** *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 third variable $\mu$:

**$(\mathbf{A}_\theta^{\sigma,b})$** Let $(p,q),(p_1,q_1)\in\mathscr{I}_1$ and $\theta\geqslant 1$. It holds that

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

and there are $\ell\in L^q_{loc}(\mathbb{R}_+)$ and a constant $c_0\geqslant 1$ such that for any two random variables $X,Y$ with finite $\theta$-order moments,

$$\begin{aligned}&\|b_t(\cdot,\mu_X)-b_t(\cdot,\mu_Y)\|_p\leqslant \ell_t\|X-Y\|_\theta,\\ &\|\sigma_t(\cdot,\mu_X)-\sigma_t(\cdot,\mu_Y)\|_\infty\leqslant c_0\|X-Y\|_\theta,\end{aligned}\tag{4.1}$$

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

Notice that (4.1) is equivalent to that for all $\mu,\mu'\in\mathcal{P}_\theta(\mathbb{R}^d)$,

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

where $\mathcal{W}_\theta$ is the usual Wasserstein metric of $\theta$-order. For convenience, we would like to use (4.1) rather than introducing the Wasserstein metric.

**Remark 4.1.** *We note that in* [10], (4.1) *is assumed to hold for* $p=\infty$.

We first show the following strong uniqueness result.

**Theorem 4.2.** *Let $\theta\geqslant 1$ and $\beta>2\vee\theta$. Under $(\mathbf{H}^{\sigma,b})$ and $(\mathbf{A}_\theta^{\sigma,b})$, for any initial random variable $X_0$ with finite $\beta$-order moment, there is a unique strong solution to DDSDE* (1.1).

*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 $\gamma>1$,

$$\|\ell\|_{L^{\gamma q}(0,1)}+\sup_{Z\in\mathbf{S}_{\rm toch}}\|b^Z\|_{\mathbb{L}^{\gamma q}_{\gamma p}(1)}<\infty.\tag{4.2}$$

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 3.7. 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:

$$\partial_t u^X+(\mathscr{L}_t^{\sigma^X}-\lambda)u+b^X\cdot\nabla u^X+b^X=0,\ u^X(T,x)=0.\tag{4.3}$$

By Theorem 2.2, for $\lambda \geqslant 1$, 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 $\lambda \geqslant 1$ and $T \in [0,1]$,

$$\lambda^{1-\frac{\alpha}{2}-\frac{2}{q}} \|\!| u^X |\!\|_{\widetilde{\mathbb{H}}^{\alpha,p}_\infty(T)} + \|\!| u^X |\!\|_{\widetilde{\mathbb{H}}^{2,p}_q(T)} \leqslant c_1 \|\!| b^X |\!\|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{4.4}$$

In particular, since $\frac{d}{p} + \frac{2}{q} < 1$, by (2.5), we can choose $\lambda$ large enough so that

$$\|u^X\|_{\mathbb{L}^\infty(T)} + \|\nabla u^X\|_{\mathbb{L}^\infty(T)} \leqslant 1/2, \quad \forall T \in [0,1]. \tag{4.5}$$

Below we shall fix such a $\lambda$ and define

$$\Phi^X_t(x) := x + u^X_t(x).$$

It is easy to see that

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

(ii) By the generalized Itô formula, we have

$$\tilde{X}_t := \Phi^X_t(X_t) = \Phi^X_0(X_0) + \lambda \int_0^t u^X_s(X_s)\mathrm{d}s + \int_0^t \tilde{\sigma}^X_s(X_s)\mathrm{d}W_s, \tag{4.6}$$

where

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

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

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

Noting that by (4.5),

$$|x-y| \leqslant 2|\Phi^X_t(x) - \Phi^X_t(y)| \leqslant 2|\Phi^X_t(x) - \Phi^Y_t(y)| + 2\|u^X - u^Y\|_{\mathbb{L}^\infty(T)}$$

and

$$|\Phi^X_t(x) - \Phi^Y_t(y)| \leqslant 2|x-y| + \|u^X - u^Y\|_{\mathbb{L}^\infty(T)},$$

we have

$$|\xi_t| \leqslant 2|\tilde{\xi}_t| + 2\|u^X - u^Y\|_{\mathbb{L}^\infty(T)}, \quad |\tilde{\xi}_t| \leqslant 2|\xi_t| + \|u^X - u^Y\|_{\mathbb{L}^\infty(T)}. \tag{4.7}$$

By (4.6) and again Itô's formula, we have for any $\beta \geqslant 1$,

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

Since by (4.5),

$$|u^X_t(x) - u^Y_t(y)| \leqslant |x-y| + \|u^X - u^Y\|_{\mathbb{L}^\infty(T)},$$

by Young's inequality we obtain

$$I_2 \lesssim \int_0^t |\tilde{\xi}_s|^\beta \mathrm{d}s + \lambda \int_0^t |u^X_s(X_s) - u^Y_s(Y_s)|^\beta \mathrm{d}s$$

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

Let

$$g^X_s(x) := |\nabla^2 u^X_s(x)| + |\nabla\sigma^X_s(x)| + \|\nabla u^X\|_{\mathbb{L}^\infty(T)} + \|\sigma^X\|_{\mathbb{L}^\infty(T)}.$$

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

$$\begin{aligned}
&|\tilde{\sigma}^X_s(x) - \tilde{\sigma}^Y_s(y)|\\
&\leqslant \|\sigma^Y\|_{\mathbb{L}^\infty(T)}|\nabla\Phi^X_s(x) - \nabla\Phi^Y_s(y)| + |\sigma^X_s(x) - \sigma^Y_s(y)|\cdot\|\nabla\Phi^X\|_{\mathbb{L}^\infty(T)}\\
&\leqslant \|\sigma^Y\|_{\mathbb{L}^\infty(T)}\Big(|\nabla u^X_s(x) - \nabla u^X_s(y)| + |\nabla u^X_s(y) - \nabla u^Y(s,y)|\Big)\\
&\quad + \Big(|\sigma^X_s(x) - \sigma^X_s(y)| + |\sigma^X_s(y) - \sigma^Y_s(y)|\Big)\cdot\|\nabla\Phi^X\|_{\mathbb{L}^\infty(T)}\\
&\overset{(2.19)}{\lesssim} |x-y|\Big(\mathcal{M}_1 g^X_s(x) + \mathcal{M}_1 g^X_s(y)\Big) + \|\nabla u^X - \nabla u^Y\|_{\mathbb{L}^\infty(T)} + \|\sigma^X_s - \sigma^Y_s\|_\infty.
\end{aligned}$$

Hence,

$$\begin{aligned}
I_4 + I_5 \lesssim \int_0^t \Big(|\xi_s|^\beta + |\tilde{\xi}_s|^\beta\Big)\Big(\mathcal{M}g^X_s(X_s) + \mathcal{M}g^X_s(Y_s)\Big)^2\mathrm{d}s\\
+ T\|\nabla u^X - \nabla u^Y\|^\beta_{\mathbb{L}^\infty(T)} + \int_0^t \|\sigma^X_s - \sigma^Y_s\|^\beta_\infty\mathrm{d}s.
\end{aligned}$$

Combining the above calculations and noting that $|\tilde{\xi}_0| \leqslant \|u^X_0 - u^Y_0\|_\infty$, we obtain

$$\begin{aligned}
|\tilde{\xi}_t|^\beta &\lesssim \|u^X - u^Y\|^\beta_{\mathbb{H}^{1,\infty}_\infty(T)} + \int_0^t \Big(|\tilde{\xi}_s|^\beta + |\xi_s|^\beta + \|\xi_s\|^\beta_\theta\Big)\mathrm{d}s\\
&\quad + \int_0^t \Big(|\xi_s|^\beta + |\tilde{\xi}_s|^\beta\Big)\Big(\mathcal{M}_1 g^X_s(X_s) + \mathcal{M}_1 g^X_s(Y_s)\Big)^2\mathrm{d}s + M_t,
\end{aligned} \tag{4.8}$$

where $M_t$ is a continuous local martingale.

(iii) Now we define

$$A_t := t + \int_0^t \Big(\mathcal{M}_1 g^X_s(X_s) + \mathcal{M}_1 g^X_s(Y_s)\Big)^2\mathrm{d}s.$$

By (4.8) and (4.7), we obtain that for all $t \in [0,T]$,

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

Note that by the assumption and (2.20),

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

and

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

Since $(\frac{p}{2},\frac{q}{2}), (\frac{p_1}{2},\frac{q_1}{2}) \in \mathscr{I}_2$, by Khasminskii's estimate (3.14), we have

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

Thus we can use the stochastic Gronwall inequality (2.18) to derive that

$$\sup_{s\in[0,T]}\|\xi_s\|^\beta_\theta = \left(\sup_{s\in[0,T]}\mathbf{E}|\xi_s|^\theta\right)^{\beta/\theta} \lesssim \|u^X - u^Y\|^\beta_{\mathbb{H}^{1,\infty}_\infty(T)} + \int_0^T \|\xi_s\|^\beta_\theta\mathrm{d}s. \tag{4.9}$$

Noticing that by (4.1),

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

and

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

we have by (2.16),

$$\begin{aligned}
&\|u^X - u^Y\|_{\mathbb{H}^{1,\infty}_\infty(T)}\\
&\lesssim \|\!|b^X - b^Y|\!\|_{\widetilde{\mathbb{L}}^p_q(T)} + \|\!|b^X|\!\|_{\widetilde{\mathbb{L}}^p_q(T)} \left(\|\sigma^X - \sigma^Y\|_{\mathbb{L}^\infty(T)} + \|\!|b^X - b^Y|\!\|_{\widetilde{\mathbb{L}}^p_q(T)}\right)\\
&\lesssim \left(\|\ell\|_{L^q(0,T)} + \|\!|b^X|\!\|_{\widetilde{\mathbb{L}}^p_q(T)}\right) \sup_{t\in[0,T]} \|\xi_t\|_\theta \overset{(4.2)}{\lesssim} T^{\frac{\gamma-1}{\gamma q}} \sup_{t\in[0,T]} \|\xi_t\|_\theta.
\end{aligned}$$

Substituting this into (4.9), we obtain

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

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 (2.31) does not satisfy (4.1). 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 $\theta \geqslant 1$, let

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

We assume

**$(\widetilde{\mathbf{A}}^{\sigma,b}_\theta)$** Let $(p,q),(p_1,q_1) \in \mathscr{I}_1$ and $\theta \geqslant 1$ and $\sigma_t(x,\mu) = \sigma_t(x)$. It holds that

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

and there is an $\ell \in L^q_{loc}(\mathbb{R}_+)$ such that for all $\mu,\mu' \in \mathcal{P}(\mathbb{R}^d)$ and $t \geqslant 0$,

$$\|\!|b(t,\cdot,\mu) - b(t,\cdot,\mu')|\!\|_p \leqslant \ell_t \|\phi_\theta \cdot (\mu - \mu')\|_{TV}. \tag{4.10}$$

It should be noted that [27, Theorem 6.15] implies,

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

**Theorem 4.3.** *Let $\theta \geqslant 1$ and $\beta > 2\theta$. Under $(\mathbf{H}^{\sigma,b})$ and $(\widetilde{\mathbf{A}}^{\sigma,b}_\theta)$, for any initial random variable $X_0$ with finite $\beta$-order moment, there is a unique weak solution to DDSDE* (1.1)*, which is also a unique strong solution.*

*Proof.* We use the Girsanov transform in the same way asin [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 3.7), without loss of generality, let $X^{(i)}, i = 1,2$ be two solutions of SDE (1.1) defined on the same probability space $(\Omega,\mathscr{F},\mathbf{P})$ and with the same Brownian motion and starting point $\xi$. That is,

$$\mathrm{d}X_t^{(i)} = \sigma_t(X_t^{(i)})\mathrm{d}W_t + b_t(X_t^{(i)},\mu_t^{(i)})\mathrm{d}t, \quad X_0^{(i)} = \xi, \tag{4.11}$$

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 (2.13) under our assumptions, it is well known that there is a unique weak solution to SDE

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

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

$$\sup_{t\in[0,T]} \mathbf{E}|Z_t|^\beta \leqslant C\Big(\mathbf{E}|\xi|^\beta + 1\Big). \tag{4.12}$$

Define

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

and

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

Since $\|\tilde{b}^{(i)}\|_{\widetilde{\mathbb{L}}_q^p(T)} \leqslant \|b^{X^{(i)}}\|_{\widetilde{\mathbb{L}}_q^p(T)} < \infty$ for some $(p, q) \in \mathscr{I}_1$, by Khasminskii's estimate (3.14), we have

$$\mathbf{E}\exp\left\{\gamma\int_0^T |\tilde{b}_s^{(i)}(Z_s)|^2\mathrm{d}s\right\} \leqslant C_{T,\gamma}, \quad \forall\gamma > 0, \tag{4.13}$$

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

$$\mathbf{E}(\mathscr{E}_T^{(i)})^\gamma \leqslant C_{T,\gamma} < \infty. \tag{4.14}$$

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

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

Since the above SDE admits a unique strong solution (see also (4.11)), we have

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

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

$$\begin{aligned}
\|\phi_\theta \cdot (\mu_T^{(1)} - \mu_T^{(2)})\|_{TV} &= \|\phi_\theta \cdot ((\mathscr{E}_T^{(1)}\mathbf{P}) \circ Z_T^{-1} - (\mathscr{E}_T^{(2)}\mathbf{P}) \circ Z_T^{-1})\|_{TV} \\
&\leqslant \mathbf{E}\Big(\phi_\theta(Z_T)|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}|\Big) \leqslant \|\phi_\theta(Z_T)\|_{\delta/(\delta-1)}\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta \\
&= \|1 + |Z_T|^\theta\|_{\beta/\theta}\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta \overset{(4.12)}{\leqslant} C\|\mathscr{E}_T^{(1)} - \mathscr{E}_T^{(2)}\|_\delta.
\end{aligned} \tag{4.15}$$

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

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

where $M$ is a continuous local martingale. Since $\delta<2$, by the stochastic Gronwall inequality (2.18) and (4.13), 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 (3.7), we further have

$$\begin{aligned}\|\mathscr{E}_T^{(1)}-\mathscr{E}_T^{(2)}\|_\delta^2&\overset{(4.14)}{\lesssim}\left(\int_0^T\mathbf{E}|\tilde b_t^{(1)}(Z_t)-\tilde b_t^{(2)}(Z_t)|^{2\gamma}\mathrm{d}t\right)^{\frac{1}{\gamma}}\\&\lesssim\|\!|\tilde b^{(1)}-\tilde b^{(2)}|^{2\gamma}\|\!|^{1/\gamma}_{\widetilde{\mathbb{L}}^{p/(2\gamma)}_{q/(2\gamma)}(T)}=\|\!|\tilde b^{(1)}-\tilde b^{(2)}\|\!|^2_{\widetilde{\mathbb{L}}^p_q(T)}\\&\lesssim\left(\int_0^T\|\!|b_t(\cdot,\mu_t^{(1)})-b_t(\cdot,\mu_t^{(2)})\|\!|_p^q\mathrm{d}t\right)^{\frac{2}{q}}\\&\overset{(4.10)}{\lesssim}\left(\int_0^T\ell_t^q\|\phi_\theta\cdot(\mu_t^{(1)}-\mu_t^{(2)})\|_{TV}^q\mathrm{d}t\right)^{\frac{2}{q}},\end{aligned}$$

which together with (4.15) yields

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

By Gronwall's inequality, we obtain

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

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 (1.1) and nonlinear Fokker-Planck equation (1.2), which was first proved in [1, 2], and is based on a result for linear Fokker-Planck equations due to Trevisan [25] (see also [11] 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 $T>0$,*

$$\int_0^T\int_{\mathbb{R}^d}\Big(|(\sigma_t^{ik}\sigma_t^{jk})(x,\mu_t)|+|b_t(x,\mu_t)|\Big)\mu_t(\mathrm{d}x)\mathrm{d}t<\infty.\tag{5.1}$$

*Then $\mu_t$ solves the nonlinear Fokker-Planck equation (1.2) in the distributional sense if and only if there exists a martingale solution $\mathbb{P}\in\mathscr{M}_\nu^{\sigma,b}$ to DDSDE (1.1) 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 martingale $\mu_t:=\mu_{X_t},t\geqslant0$, satisfying (5.1), then there is at most one solution to (1.2) satisfying (5.1).*

*Proof.* If $\mathbb{P} \in \mathscr{M}_\nu^{\sigma,b}$ and $\mu_t = \mathbb{P} \circ w_t^{-1}$, then by (5.1) and Itô's formula, it is easy to see that $\mu_t$ solves (1.2). Now we assume $\mu_t$ solves (1.2). Consider the following linear Fokker-Planck equation:

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

where $b_t^\mu(x) := b_t(x, \mu_t)$ and $\sigma_t^\mu(x) := \sigma_t(x, \mu_t)$. Since $\mu_t$ 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 wellposedness result about the nonlinear Fokker-Planck equations.

**Theorem 5.2.** *In the situations of Theorems 4.2 and 4.3, there is a unique continuous curve $\mu_t$ solving the nonlinear Fokker-Planck equation* (1.2).

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 (1.7) are consequences of Theorem 4.3 and Theorem 5.1. We now aim to show the existence and smoothness of the density $\rho_t^X(y)$. Let $\mu_t$ be the solution of the Fokker-Planck equation (1.7). We consider the following SDE:

$$\mathrm{d}X_t = b_t^\mu(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t,\ X_0 = \xi, \tag{5.2}$$

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 [8, Theorems 1.1 and 1.3]), which is continuous in $(s,x;t,y)$ on $\{(s,x;t,y) : 0 \leqslant s < t < \infty, x, y \in \mathbb{R}^d\}$ and satisfies the following two-sided estimate: For any $T > 0$, there are constants $c_0, \gamma_0 > 1$ such that for all $0 \leqslant s < t \leqslant T$ and $x, y \in \mathbb{R}^d$

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

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

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

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

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

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

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

Strong uniqueness of SDE (5.2) ensures that $\rho_t^X(y)\mathrm{d}y = \mu_t(\mathrm{d}y)$. The desired estimates now follow from the above estimates. □

**Acknowledgement:** The authors thank Dr. Xing Huang for pointing out an error in the earlier version.

## References


[1] Barbu V. and Röckner M.: Probabilistic representation for solutions to nonlinear Fokker-Planck equations. *SIAM J. Math. Anal.* **50** (2018), no. 4, 4246-4260.

[2] Barbu V. and Röckner M.: From nonlinear Fokker-Planck equations to solutions of distribution dependent SDE. arXiv:1808.10706.

[3] Bogachev V.I., Krylov N.V. Röckner M. and Shaposhinikov S.V.: *Fokker-Planck-Kolmogorov equations.* AMS, 2015.

[4] Carmona R. and Delarue F.: Probabilistic analysis of mean-field games. *SIAM J. Control and Optimization*, 51(4):2705-2734, 2013.

[5] Carmona R. and Delarue F.: *Probabilistic theory of mean field games with applications. II. Mean field games with common noise and master equations.* Probability Theory and Stochastic Modeling, 84. Springer, 2018.

[6] Carrillo J.A., Gvalani R.S., Pavliotis G.A. and Schlichting A.: Long-time behavior and phase transitions for the McKean-Vlasov equation on the torus. arXiv: 1806.01719v2.

[7] Chiang T.: McKean-Vlasov equations with discontinuous coefficients. *Soochow J. Math.*, 20(4):507-526, 1994.

[8] Chen Z.Q., Hu E., Xie L. and Zhang X.: Heat kernels for non-symmetric diffusion operators with jumps. *J. Differential Equations* 263 (2017) 6576-6634.

[9] Hammersley W., Siska D. and Szpruch L.: Mckean-Vlasov SDEs under measure dependent Lyapunov conditions. arXiv:1802.03974v2.

[10] Huang X. and Wang F.Y.: Distribution dependent SDEs with singular coefficients. arXiv:1805.01682v1.

[11] Figalli A.: Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients. *J. Funct. Anal.* 254 (2008), no. 1, 109-153.

[12] Funaki T.: A certain class of diffusion processes associalted with nonlinear parabolic equations. *Prob. Theory and Relat. Fields*, 67(3):331-348,1984.

[13] Jacod J. and Shiryaev A. N.: *Limit theorems for stochastic processes.* Springer-Verlag, 2002.

[14] Kac M.: Foundations of Kinetic Theory. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics, pages 171197, Berkeley, Calif., 1956. University of California Press.

[15] Krylov N.V. and Röckner M.: Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Relat. Fields*, **131** (2005), 154-196.

[16] Kurtz T.G.: Weak and strong solutions of general stochastic models. *Electron. Common. Probab.*, 19(58):1-16, 2014.

[17] Li J. and Min H.: Weak solutions of mean-field stochastic differential equations and application to zero-sum stochastic differential games. *SIAM Journal on Control and Optimization*, 54(3):1826-1858, 2016.

[18] Manita O.A., Romanov M.S., Shaposhnikov S.V.: On uniqueness of solutions to nonlinear Fokker-Planck-Kolmogorov equations. *Nonlinear Anal.* 128 (2015), 199-226.

[19] Manita, O. A.; Shaposhnikov, S. V.: Nonlinear parabolic equations for measures. (Russian) Algebra i Analiz 25 (2013), no. 1, 64–93; translation in St. Petersburg Math. J. 25 (2014), no. 1, 4362 .

[20] McKean H. P.: A class of Markov processes associated with nonlinear parabolic equations. *Proc Nat. Acad Sci USA*, 56(6):1907-1911, 1966.

[21] Mishura Y.S. and Veretennikov A.Y.: Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations, arXiv:1603.02212v4.

[22] Scheutzow M.: A stochastic Gronwall's lemma. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, **16**, No. 2 (2013) 1350019 (4 pages).

[23] Stroock D.W. and Varadhan S.S.: *Multidimensional diffusion processes.* Springer-Verlag, Berlin, 1979.

[24] Sznitman A.S.: *Topics in propagation of chaos.* In École d'Été de Prob. de Saint-Flour XIX-1989, Vol. 1464, *Lect. Notes in Math.*, pages 165-251. Springer-Verlag, 1991.

[25] Trevisan D.: Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients. *Electron. J. Probab.***21**, (2016), Paper No. 22, 41 pp.

[26] Vlasov A. A.: The vibrational properties of an electron gas. *Soviet Physics Uspekhi*, 10(6):721, 1968.

[27] Villani C.: *Optimal transport: old and new*. Springer-Verlag, 2009.
[28] Wang F.Y.: Distribution dependent SDEs for Landau type equations. *Stoch. Procc. Appl.*, 128, no. 2, 595-621(2018).
[29] Xia P., Xie L., Zhang X. and Zhao G.: $L^q(L^p)$-theory of stochastic differential equations. Preprint.
[30] Xie L. and Zhang X.: Ergodicity of stochastic differential equations with jumps and singular coefficients. To appear in *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques*, arXiv:1705.07402.
[31] Zhang X.: Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients. *Electron. J. Probab.* **16** (2011), 1096-1116.
[32] Zhang X.: A discretized version of Krylov's estimate and its applications. arXiv:1909.09976.
[33] Zhang X. and Zhao G.: Heat kernel and ergodicity of SDEs with distributional drifts. arXiv.1710.10537.
[34] Zhang X. and Zhao G.: Singular Brownian Diffusion Processes. *Communications in Mathematics and Statistics,* pp.1-49, 2018.
[35] Zhang X. and Zhao, G.: Stochastic Lagrangian path for Leray solutions of 3D Navier-Stokes equations, arXiv: 1904.04387.
[36] Zvonkin, A.K.: A transformation of the phase space of a diffusion process that removes the drift. *Mat. Sbornik.* **93** (135) (1974), 129-149.

Michael Röckner: Fakultät für Mathematik, Universität Bielefeld, 33615, Bielefeld, Germany, Email: roeckner@math.uni-bielefeld.de

Xicheng Zhang: School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R.China, Email: XichengZhang@gmail.com