Distribution Dependent Stochastic Differential Equations*

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

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

[b)] Department of Mathematics, City University of HongKong, HongKong, China

xinghuang@tju.edu.cn, rppzoe@gmail.com, wangfy@tju.edu.cn

June 17, 2021

Abstract

Due to their intrinsic link with nonlinear Fokker-Planck equations and many other applications, distribution dependent stochastic differential equations (DDSDEs for short) have been intensively investigated. In this paper we summarize some recent progresses in the study of DDSDEs, which include the correspondence of weak solutions and nonlinear Fokker-Planck equations, the well-posedness, regularity estimates, exponential ergodicity, long time large deviations, and comparison theorems.

AMS subject Classification: 60B05, 60B10.
Keywords: DDSDE, nonlinear Fokker-Planck equation, Bismut formula, Wasserstein distance, gradient estimate.

1 Introduction

To characterize nonlinear PDEs in Vlasov's kinetic theory, Kac [27, 28] proposed the "propagation of chaos" of mean field particle systems, which stimulated McKean [33] to study nonlinear Fokker-Planck equations using stochastic differential equations with distribution dependent drifts, see [45] for a theory on mean field particle systems and applications.

In general, a nonlinear Fokker-Planck equation can be characterized by the following distribution dependent stochastic differential equations (DDSDEs for short):

$$\mathrm{d}X_t = b(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X_t, \mathscr{L}_{X_t})\mathrm{d}W_t, \tag{1.1}$$

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

where W_t is an m-dimensional Brownian motion on a complete filtration probability space $(\Omega, \{\mathscr{F}_t\}_{t\ge 0}, \mathbb{P})$, $\mathscr{L}_\xi$ is the distribution (i.e. the law) of a random variable ξ,

$$b = (b_i)_{1\le i\le d} : [0,\infty) \times \mathbb{R}^d \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}^d,$$
$$\sigma = (\sigma_{ij})_{1\le i\le d, 1\le j\le m} : [0,\infty) \times \mathbb{R}^d \times \mathscr{P}(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^m$$

are measurable, and $\mathscr{P}(\mathbb{R}^d)$ is the space of probability measures on $\mathbb{R}^d$ equipped with the weak topology. Due to the pioneering work [33] of McKean, the DDSDE (1.1) is also called McKean-Vlasov SDE or mean field SDE.

Definition 1.1. Let $s \ge 0$.

(1) A continuous adapted process $(X_{s,t})_{t\ge s}$ is called a solution of (1.1) from time s, if

$$\int_s^t \mathbb{E}\big[|b(r, X_{s,r}, \mathscr{L}_{X_{s,r}})| + \|\sigma(r, X_{s,r}, \mathscr{L}_{X_{s,r}})\|^2\big]\mathrm{d}r < \infty, \quad t \ge s,$$

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

$$X_{s,t} = X_{s,s} + \int_s^t b(r, X_{s,r}, \mathscr{L}_{X_{s,r}})\mathrm{d}r + \int_s^t \sigma(r, X_{s,r}, \mathscr{L}_{X_{s,r}})\mathrm{d}W_r, \quad t \ge s.$$

When $s = 0$ we simply denote $X_t = X_{0,t}$.

(2) A couple $(\tilde{X}_{s,t}, \tilde{W}_t)_{t\ge s}$ is called a weak solution of (1.1) from time s, if $\tilde{W}_t$ is the m-dimensional Brownian motion on a complete filtration probability space $(\tilde{\Omega}, \{\tilde{\mathscr{F}}_t\}_{t\ge 0}, \tilde{\mathbb{P}})$ such that $(\tilde{X}_{s,t})_{t\ge s}$ is a solution of (1.1) from time s for $(\tilde{W}_t, \tilde{\mathbb{P}})$ replacing $(W_t, \mathbb{P})$. (1.1) is called weakly unique for an initial distribution $\nu \in \mathscr{P}(\mathbb{R}^d)$, if all weak solutions with distribution ν at time s are equal in law.

(3) Let $\hat{\mathscr{P}}(\mathbb{R}^d)$ be a subspace of $\mathscr{P}(\mathbb{R}^d)$. (1.1) is said to have strong (respectively, weak) well-posedness for initial distributions in $\hat{\mathscr{P}}(\mathbb{R}^d)$, if for any $\mathscr{F}_s$-measurable $X_{s,s}$ with $\mathscr{L}_{X_{s,s}} \in \hat{\mathscr{P}}(\mathbb{R}^d)$ (respectively, any initial distribution $\nu \in \hat{\mathscr{P}}(\mathbb{R}^d)$ at time s), it has a unique strong (respectively, weak) solution. We call the equation well-posed if it is both strongly and weakly well-posed.

According to Yamada-Watanabe principle, for classical SDEs the strong well-posedness implies the weak one. But this does not apply to DDSDEs, see Theorem 3.2 below for a modified Yamada-Watanabe principle.

In this paper, we summarize the following recent progress in the study of the DDSDE (1.1): the correspondence between the weak solution of (1.1) and the associated nonlinear Fokker-Planck equation (Section 2), criteria on the well-posedness (i.e. existence and uniqueness of solutions) (Section 3), regularity of distributions (Section 4), exponential ergodicity (Section 5), long time large deviations (Section 6), and comparison theorems (Section 7). Corresponding results for general models of path-distribution dependent SDEs/SPDEs can be found in [2, 19, 37].

2 Weak solution and nonlinear Fokker-Planck equation

In this part, we first introduce the "superposition principle" which provides a correspondence between the weak solution of (1.1) and the solution of the associated nonlinear Fokker-Planck equation on $\mathscr{P}(\mathbb{R}^d)$, then present some typical examples.

2.1 Superposition principle

Consider the following nonlinear Fokker-Planck equation on $\mathscr{P}(\mathbb{R}^d)$:

$$\partial_t \mu_t = L^*_{t,\mu_t}\mu_t, \tag{2.1}$$

where for any $(t,\mu) \in [0,\infty) \times \mathscr{P}(\mathbb{R}^d)$, the Kolmogorov operator $L_{t,\mu}$ on $\mathbb{R}^d$ is given by

$$L_{t,\mu} := \frac{1}{2}\sum_{i,j=1}^d (\sigma\sigma^*)_{ij}(t,\cdot,\mu)\partial_i\partial_j + \sum_{i=1}^d b_i(t,\cdot,\mu)\partial_i,$$

for σ^* being the transposition of σ.

Definition 2.1. For $s \geq 0$, $\mu_\cdot \in C([s,\infty);\mathscr{P}(\mathbb{R}^d))$ is called a solution of (2.1) from time s, if

$$\int_s^t \mathrm{d}r \int_{\mathbb{R}^d} \big\{\|\sigma(r,x,\mu_r)\|^2 + |b(r,x,\mu_r)|\big\}\mu_r(\mathrm{d}x) < \infty, \quad t > s,$$

and for any $f \in C_0^\infty(\mathbb{R}^d)$,

$$\mu_t(f) := \int_{\mathbb{R}^d} f \mathrm{d}\mu_t = \mu_s(f) + \int_s^t \mu_r(L_{r,\mu_r}f)\mathrm{d}r, \quad t \geq s. \tag{2.2}$$

Now, assume that $(\tilde{X}_t, \tilde{W}_t)_{t\geq s}$ is a weak solution of (1.1) from time s under a complete filtration probability space $(\tilde{\Omega}, \{\tilde{\mathscr{F}}_t\}_{t\geq s}, \tilde{\mathbb{P}})$, and let $\mu_t = \mathscr{L}_{\tilde{X}_t|\tilde{\mathbb{P}}} := \tilde{\mathbb{P}} \circ (\tilde{X}_t)^{-1}$ be the distribution of $\tilde{X}_t$ under the probability $\tilde{\mathbb{P}}$. By Itô's formula we have

$$\mathrm{d}f(\tilde{X}_t) = \big\{L_{t,\mu_t}f(\tilde{X}_t)\big\}\mathrm{d}t + \langle \nabla f(\tilde{X}_t), \sigma(t,\tilde{X}_t,\mu_t)\mathrm{d}\tilde{W}_t\rangle.$$

Integrating both sides over $[s,t]$ and taking expectations, we obtain (2.2) so that $\mu_\cdot$ solves (2.1) by definition. Indeed, the following result due to [5, 6] also ensures the converse, i.e. a solution of (2.1) gives a weak solution of (1.1), see Section 2 of [6] (and [5]).

Theorem 2.1 ([5, 6])**.** *Let $(s,\zeta) \in [0,\infty) \times \mathscr{P}(\mathbb{R}^d)$. Then the DDSDE* (1.1) *has a weak solution $(\tilde{X}_t, \tilde{W}_t)_{t\geq s}$ starting from s with $\mathscr{L}_{\tilde{X}_s|\tilde{\mathbb{P}}} = \zeta$, if and only if* (2.1) *has a solution $(\mu_t)_{t\geq s}$ starting from s with $\mu_s = \zeta$. In this case $\mu_t = \mathscr{L}_{\tilde{X}_t|\tilde{\mathbb{P}}}$, $t \geq s$.*

2.2 Some examples

In this part, we introduce some typical nonlinear PDES and state their corresponding DDSDEs.

Example 2.1 (Landau type equations). Consider the following nonlinear PDE for probability density functions $(f_t)_{t\geq 0}$ on $\mathbb{R}^d$:

$$\partial_t f_t = \frac{1}{2}\text{div}\bigg\{\int_{\mathbb{R}^d} a(\cdot - z)\big(f_t(z)\nabla f_t - f_t \nabla f_t(z)\big)\text{d}z\bigg\}, \tag{2.3}$$

where $a : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ has weak derivatives. For the real-world model of homogenous Landau equation we have $d = 3$ and

$$a(x) = |x|^{2+\gamma}\Big(I - \frac{x \otimes x}{|x|^2}\Big), \quad x \in \mathbb{R}^3$$

for some constant $\gamma \in [-3, 1]$. In this case (2.3) is a limit version of Boltzmann equation (for thermodynamic system) when all collisions become grazing. To characterize this equation using SDE, let $m = d$, $b = \frac{1}{2}\text{div}a$ and $\sigma = \sqrt{a}$. Consider the DDSDE

$$\text{d}X_t = (b * \mathscr{L}_{X_t})(X_t)\text{d}t + (\sigma * \mathscr{L}_{X_t})(X_t)\text{d}B_t, \tag{2.4}$$

where

$$(f * \mu)(x) := \int_{\mathbb{R}^d} f(x - z)\mu(\text{d}z).$$

Then the distribution density $f_t(x) := \frac{\mathscr{L}_{X_t}(\text{d}x)}{\text{d}x}$ solves the Landau type equation (2.3).

Example 2.2 (Porous media equation). Consider the following nonlinear PDE for probability density functions on $\mathbb{R}^d$:

$$\partial_t f_t = \Delta f_t^3. \tag{2.5}$$

Then for any solution to the DDSDE (1.1) with coefficients

$$b = 0, \quad \sigma(x, \mu) = \sqrt{2}\frac{\text{d}\mu}{\text{d}x}(x) I_{d\times d},$$

the probability density function solves the porous media equation (2.5).

Example 2.3 (Granular media equation). Consider the following nonlinear PDE for probability density functions on $\mathbb{R}^d$:

$$\partial_t f_t = \Delta f_t + \text{div}\big\{f_t \nabla V + f_t \nabla(W * f_t)\big\}. \tag{2.6}$$

Then the associated DDSDE (1.1) has coefficients

$$b(x, \mu) = -\nabla V(x) - \nabla(W * \mu)(x), \quad \sigma(x, \mu) = \sqrt{2} I_{d\times d},$$

where

$$(W * \mu)(x) := \int_{\mathbb{R}^d} W(x - y)\mu(\text{d}y).$$

3 Well-posedness

We first introduce a fixed-point argument in distribution and a modified Yamada-Watanabe principle, then present results on the existence and uniqueness for monotone and singular coefficients respectively.

3.1 Fixed-point in distribution and Yamada-Watanabe principle

Let $\hat{\mathscr{P}}(\mathbb{R}^d)$ be a subspace of $\mathscr{P}(\mathbb{R}^d)$, and let $\hat{\rho}$ be a complete metric on $\hat{\mathscr{P}}(\mathbb{R}^d)$ inducing the Borel sigma algebra of the weak topology. Typical examples include

$$\hat{\mathscr{P}}(\mathbb{R}^d) = \mathscr{P}_p(\mathbb{R}^d) := \big\{\mu \in \mathscr{P}(\mathbb{R}^d) : \mu(|\cdot|^p) < \infty\big\}$$

for $p > 0$, with L^p-Wasserstein distance

$$\mathbb{W}_p(\mu,\nu) := \inf_{\pi\in\mathscr{C}(\mu,\nu)} \bigg(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p \pi(\mathrm{d}x,\mathrm{d}y)\bigg)^{\frac{1}{p\vee 1}}, \quad \mu,\nu\in\mathscr{P}_p(\mathbb{R}^d).$$

When $p = 0$ this reduces to the total variation norm

$$\|\mu-\nu\|_{TV} := 2 \sup_{A\in\mathscr{B}(\mathbb{R}^d)} |\mu(A)-\nu(A)|.$$

For any $T > s \geq 0$ and $\nu \in \hat{\mathscr{P}}(\mathbb{R}^d)$, consider the path space over $\hat{\mathscr{P}}(\mathbb{R}^d)$

$$\hat{\mathscr{C}}^\nu_{s,T} := \big\{\mu_\cdot \in C([s,T];\hat{\mathscr{P}}(\mathbb{R}^d)) : \mu_s = \nu\big\},$$

which is then complete under the metric

$$\hat{\rho}_{s,T}(\mu_\cdot,\nu_\cdot) := \sup_{t\in[s,T]} \hat{\rho}(\mu_t,\nu_t).$$

Theorem 3.1. *Let $T > s \geq 0$, and let X_s be an $\mathscr{F}_s$-measurable random variable with $\nu := \mathscr{L}_{X_s} \in \hat{\mathscr{P}}(\mathbb{R}^d)$. Assume that for any $\mu \in \hat{\mathscr{C}}^\nu_{s,T}$, the classical SDE*

$$\mathrm{d}X^\mu_t = b(t,X^\mu_t,\mu_t)\mathrm{d}t + \sigma(t,X^\mu_t,\mu_t)\mathrm{d}W_t, \quad t\in[s,T], X^\mu_s = X_s \tag{3.1}$$

has a unique solution, and the map

$$\mu \in \hat{\mathscr{C}}^\nu_{s,T} \mapsto \Phi_{s,T}\mu := (\mathscr{L}_{X^\mu_t})_{t\in[s,T]} \in \hat{\mathscr{C}}^\nu_{s,T}$$

is contractive. Then the DDSDE (1.1) *has well-posedness for initial distributions in* $\hat{\mathscr{P}}(\mathbb{R}^d)$.

Proof. By the fixed-point theorem, the map $\Phi_{s,T}$ has a unique fixed point μ in $\hat{\mathscr{C}}^\mu_{s,T}$, so that by the definition of $\Phi_{s,T}$ we have $\mathscr{L}_{X^\mu_t} = \mu_t, t\in[s,T]$, i.e. in this case $(X^\mu_t)_{t\in[s,T]}$ is a solution of (1.1) from time s starting at X_s. If (1.1) has another solution $(\hat{X}_t)_{t\in[s,T]}$ with $\mathscr{L}_{\hat{X}_\cdot} \in \hat{C}^\mu_{s,T}$, then $\mu := \mathscr{L}_{\hat{X}_\cdot}$ is a fixed point of $\Phi_{s,T}$ so that $\mathscr{L}_{X_t} = \mathscr{L}_{\hat{X}_t} =: \mu_t, t\in[s,T]$. Therefore, by the uniqueness of (3.1) we have $\mathscr{L}_{X_t} = \mathscr{L}_{\hat{X}_t} = X^\mu_t$, which implies the uniqueness of (1.1) with $\mathscr{L}_{X_\cdot} \in \hat{\mathscr{C}}^\mu_{s,T}$. Since the strong well-posedness of (3.1) implies the weak one, the same argument leads to the weak well-posedness of the DDSDE (1.1) starting from ν at time s. □

The Yamada-Watanabe principle [55] (see [30] for a general version) is a fundamental tool in the study of well-posedness for SDEs with singular coefficients. In the present distribution dependent setting, the original statement does not apply, but we have the following modified version due to [21, Lemma 3.4].

Theorem 3.2 ([21])**.** *Let $T > s \geq 0$, and let X_s be an $\mathscr{F}_s$-measurable random variable with $\nu := \mathscr{L}_{X_s} \in \hat{\mathscr{P}}(\mathbb{R}^d)$. Assume that for any $\mu \in \hat{\mathscr{C}}^{\nu}_{s,T}$, the classical SDE* (3.1) *has a unique solution with initial value X_s at time s. If* (1.1) *for $t \in [s, T]$ has a weak solution with initial distribution ν at time s, and has pathwise uniqueness with initial value X_s at times s, then it has well-posedness for initial distributions in $\hat{\mathscr{P}}(\mathbb{R}^d)$.*

3.2 The monotone case

(H_3^1) For every $t \geq 0$, b_t is continuous on $\mathbb{R}^d \times \mathscr{P}_\theta(\mathbb{R}^d)$, b is bounded on bounded sets in $[0,\infty) \times \mathbb{R}^d \times \mathscr{P}_\theta(\mathbb{R}^d)$. Moreover, there exists $K \in L^1_{loc}([0,\infty);(0,\infty))$ such that

$$\begin{aligned}&\|\sigma(t,x,\mu)-\sigma(t,y,\nu)\|^2 \leq K(t)\big\{|x-y|^2+\mathbb{W}_\theta(\mu,\nu)^2\big\},\\ &\langle b(t,x,\mu)-b(t,y,\nu),x-y\rangle \leq K(t)\big\{|x-y|^2+\mathbb{W}_\theta(\mu,\nu)|x-y|\big\},\\ &|b(t,0,\delta_0)|+\|\sigma(t,0,\delta_0)\|_{HS}^2 \leq K(t),\quad t\geq 0, x,y\in\mathbb{R}^d,\mu,\nu\in\mathscr{P}_\theta(\mathbb{R}^d),\end{aligned}$$

where δ_0 is the Dirac measure at $0 \in \mathbb{R}^d$.

Under this monotone condition we have the following result essentially due to [51], where a stronger growth condition on $|b(t,0,\mu)|$ is assumed. See also [17] for the well-posedness under integrated Lyapunov conditions which may cover more examples.

Theorem 3.3 ([51])**.** *Assume (H_3^1) for some $\theta \in [1,\infty)$, and let $\sigma(t,x,\mu)$ does not depend on μ when $\theta < 2$.*

(1) *The DDSDE* (1.1) *has well-posedness for initial distributions in $\mathscr{P}_\theta(\mathbb{R}^d)$. Moreover, for any $p \geq \theta$ and $s \geq 0$, $\mathbb{E}|X_{s,s}|^p < \infty$ implies*

$$\mathbb{E}\sup_{t\in[s,T]}|X_{s,t}|^p<\infty,\quad T\geq t\geq s\geq 0.$$

(2) *There exists increasing $\psi : [0,\infty) \to [0,\infty)$ such that for any two solutions $X_{s,t}$ and $Y_{s,t}$ of* (1.1) *with $\mathscr{L}_{X_{s,s}}, \mathscr{L}_{Y_{s,s}} \in \mathscr{P}_\theta(\mathbb{R}^d)$,*

$$\mathbb{E}|X_{s,t}-Y_{s,t}|^\theta \leq \big(\mathbb{E}|X_{s,s}-Y_{s,s}|^\theta\big)\mathrm{e}^{\int_s^t \psi(r)\mathrm{d}r},\quad t\geq s\geq 0. \tag{3.2}$$

Consequently,

$$\lim_{\mathbb{E}|X_{s,s}-Y_{s,s}|^\theta\to 0}\mathbb{P}\Big(\sup_{r\in[s,t]}|X_{s,r}-Y_{s,r}|\geq\varepsilon\Big)=0,\quad t>s\geq 0,\varepsilon>0; \tag{3.3}$$

and

$$\mathbb{W}_\theta(P^*_{s,t}\mu_0,P^*_{s,t}\nu_0)^\theta \leq \mathbb{W}_2(\mu_0,\nu_0)^\theta\mathrm{e}^{\int_s^t \psi(r)\mathrm{d}r},\quad t\geq s\geq 0. \tag{3.4}$$

Proof. We briefly explain the proof of Theorem 3.3(1), while (2) can be easily proven by using Itô's formula. For any $T > s \geq 0$, $\nu \in \mathscr{P}_\theta(\mathbb{R}^d)$ and $\mu \in \hat{\mathscr{C}}^\nu_{s,T}$, (H_3^1) implies that (3.1) for $t \in [s, T]$ is well-posed with initial distribution ν at s. Moreover, by Itô's formula, and (H_3^1) with $\sigma(t, x, \mu)$ not depending on μ when $\theta < 2$, we find a large enough constant $\lambda > 0$ such that $\Phi_{s,T}$ is contractive on $\hat{\mathscr{C}}^\nu_{s,T}$ under the complete metric

$$\hat{\rho}_{s,T}(\mu, \tilde{\mu}) := \sup_{t\in[s,T]} \mathrm{e}^{-\lambda(t-s)} \mathbb{W}_\theta(\mu_t, \tilde{\mu}_t), \quad \mu, \tilde{\mu} \in \hat{C}^\nu_{s,T}.$$

Then the well-posedness follows from Theorem 3.1. □

3.3 The singular case

In this part, we consider the existence and uniqueness of (1.1) with singular drift and non-degenerate noise. We first introduce some results derived in [58, 43, 57] for distribution dependent drifts satisfying local integrability conditions in time and space but bounded in distribution, in [24] for the case with locally integrable drifts having linear growth in distribution, and in [22] for drifts with an integrable term and a Lipchitz term. These three situations are mutually incomparable.

3.3.1 Integrability in time-space and boundedness in distribution

When the noise is possibly degenerate, the strong/weak well-posedness will be discussed in the next section under a monotone condition.

We will consider weak solutions having finite ϕ-moment, for ϕ in the following class:

$$\mathbf{\Phi} := \big\{\phi \in C^\infty([0,\infty); [1,\infty)) : 0 \leq \phi' \leq c\phi \text{ for some constant } c > 0\big\}.$$

Let

$$\mathscr{P}_\phi(\mathbb{R}^d) := \big\{\mu \in \mathscr{P}(\mathbb{R}^d) : \|\mu\|_\phi := \mu(\phi(|\cdot|)) < \infty\big\},$$

which is equipped with the ϕ-total variation norm

$$\|\mu - \nu\|_{\phi,TV} := \sup_{|f|\leq\phi(|\cdot|)} \big|\mu(f) - \nu(f)\big|, \quad \mu, \nu \in \mathscr{P}(\mathbb{R}^d).$$

We denote $\|\cdot\|_{\phi,TV}$ by $\|\cdot\|_{\theta,TV}$ when $\phi = 1 + |\cdot|^\theta$ for some $\theta \geq 0$. For fixed $T > 0$, let

$$\mathscr{C}_{T,\phi} = C([0,T]; \mathscr{P}_\phi(\mathbb{R}^d)) := \Big\{\mu : [0,T] \to \mathscr{P}_\phi(\mathbb{R}^d), \lim_{t\to s} \|\mu_t - \mu_s\|_{\phi,TV} = 0,\ s \in [0,T]\Big\},$$

which is a complete space under the metric

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

For any $\mu \in \mathscr{C}_{T,\phi}$, denote

$$b^\mu(t,x) := b(t,x,\mu_t), \quad \sigma^\mu(t,x) := \sigma(t,x,\mu_t), \quad a^\mu(t,x) := \frac{1}{2}\big\{\sigma^\mu(\sigma^\mu)^*\big\}(t,x), \quad (t,x) \in [0,T] \times \mathbb{R}^d.$$

Definition 3.1 (Linear Functional Derivative)**.** Let $\phi \in \mathbf{\Phi}$. A function $f : \mathscr{P}_\phi(\mathbb{R}^d) \to \mathbb{R}$ is said to have linear functional derivative $D^F f : \mathscr{P}_\phi(\mathbb{R}^d) \to \mathbb{R}$, if it is measurable, and

(i) $D^F f$ is measurable with $\int_{\mathbb{R}^d} D^F f(\mu) \mathrm{d}\mu = 0$;

(ii) For any compact $K \subset \mathscr{P}_\phi(\mathbb{R}^d)$, $\sup_{\mu \in K} |D^F f(\mu)| \le k\phi(|\cdot|)$ holds for some constant $k > 0$;

(iii) For any $\mu, \nu \in \mathscr{P}_\phi(\mathbb{R}^d)$,

$$\lim_{s \downarrow 0} \frac{f((1-s)\mu + s\nu) - f(\mu)}{s} = \int_{\mathbb{R}^d} D^F f(\mu)(y)(\nu - \mu)(\mathrm{d}y).$$

By taking $\nu = \delta_y$, we see that if f has linear functional derivative, then the convex extrinsic derivative

$$\tilde{D}^E f(\mu)(y) := \lim_{s \downarrow 0} \frac{f((1-s)\mu + s\delta_y) - f(\mu)}{s} = D^F f(\mu)(y) - \int_{\mathbb{R}^d} D^F f(\mu)(y) \mathrm{d}\mu$$

exists. See [39] for links of more derivatives in measure. For $i = 1, 2$, let

$$\mathscr{I}_i = \Big\{(p, q) \in (1, \infty) \times (1, \infty) : \frac{d}{p} + \frac{2}{q} < i\Big\}. \tag{3.5}$$

Definition 3.2. For any $p \ge 1$, let $\tilde{L}_p$ be the space of all measurable functions g on $\mathbb{R}^d$ such that

$$\|g\|_{\tilde{L}_p} := \sup_{z \in \mathbb{R}^d} \bigg(\int_{\mathbb{R}^d} |g(x)|^p 1_{\{|x-z| \le 1\}} \mathrm{d}x \bigg)^{\frac{1}{p}} < \infty.$$

Moreover, for any $p, q \ge 1$, let $\tilde{L}_p^q(T)$ be the space of measurable functions f on $[0, T] \times \mathbb{R}^d$ such that

$$\|f\|_{\tilde{L}_p^q(T)} := \sup_{z \in \mathbb{R}^d} \bigg(\int_0^T \bigg(\int_{\mathbb{R}^d} |f(t, x)|^p 1_{\{|x-z| \le 1\}} \mathrm{d}x \bigg)^{\frac{q}{p}} \mathrm{d}t \bigg)^{\frac{1}{q}} < \infty.$$

It is clear that

$$\|f\|_{\tilde{L}_p^q(T)} \le \|f\|_{L^q([0,T]; \tilde{L}_p)} := \bigg(\int_0^T \|f(t, \cdot)\|_{\tilde{L}_p}^q \bigg)^{\frac{1}{q}}.$$

The following result is due to [58, Theorems 3.5 and 3.9], see also [44] for a special case where $\phi(r) = r^2$ and $b(t, x, \cdot)$ is bounded and Lipschitz continuous in the total variation norm uniformly in (t, x).

Theorem 3.4 ([58])**.** *Let $\sigma\sigma^*$ be invertible, $\phi \in \mathbf{\Phi}$, and $p, q \in (1, \infty)$ with $\varepsilon := 1 - \frac{d}{p} - \frac{2}{q} > 0$.*

(1) *If there exist constants $\alpha \in (0, 1), N > 1,$ and $r > \frac{2}{\varepsilon}$ such that for any $\mu \in \mathscr{C}_{T,\phi}$,*

$$\sup_{t \in [0,T], x \ne y} \frac{\|a^\mu(t, x) - a^\mu(t, y)\|}{|x - y|^\alpha} + \sup_{(t,x) \in [0,T] \times \mathbb{R}^d} \big\{\|a^\mu\| + \|(a^\mu)^{-1}\|\big\}(t, x) + \|b^\mu\|_{\tilde{L}_p^q(T)} \le N,$$

and that

$$\lim_{\rho_{T,\phi}(\nu,\mu)\to 0}\bigg\{\int_0^T \|a^\mu(t,\cdot)-a^\nu(t,\cdot)\|_\infty^r \mathrm{d}t + \|b^\mu - b^\nu\|_{\tilde L_p^q(T)}\bigg\} = 0,$$

then (1.1) *has a weak solution for* $t\in[0,T]$ *and any initial distribution in* $\mathscr{P}_\phi(\mathbb{R}^d)$.

(2) *In addition to conditions in* (1), *if for any* $(t,x)\in[0,T]\times\mathbb{R}^d$, $\sigma(t,x,\cdot)$ *has linear functional derivative on* $\mathscr{P}_\phi(\mathbb{R}^d)$, *and there exist constants* $\beta\in(0,1), C>0$ *and some* $K\in L^q([0,T];(0,\infty))$ *such that*

$$\begin{aligned}&\sup_{(t,\mu)\in[0,T]\times\mathscr{P}_\phi(\mathbb{R}^d)}\big\|D^F\sigma(t,x,\cdot)(\mu)(y)-D^F\sigma(t,x',\cdot)(\mu)(y')\big\|\\&\le C(|x-x'|+|y-y'|)^\beta,\qquad x,x',y,y'\in\mathbb{R}^d,\\&\|b(t,\cdot,\mu)-b(t,\cdot,\nu)\|_{\tilde L_p}\le K(t)\|\mu-\nu\|_{\phi,TV},\quad t\in[0,T],\mu,\nu\in\mathscr{P}_\phi(\mathbb{R}^d),\end{aligned}$$

then (1.1) *is has weak well-posedness for* $t\in[0,T]$ *and initial distribution in* $\mathscr{P}_\phi(\mathbb{R}^d)$.

When $\sigma=\sqrt{2}I_{d\times d}$ and

$$|b(t,x,\mu)|\le\int_{\mathbb{R}^d}h_t(x-y)\mu(\mathrm{d}y)$$

holds for some $(p,q)\in\mathscr{I}_1$ and $h\ge 0$ with $\|h\|_{L^q([0,T];\tilde L_p)}<\infty$, the well-posedness for (1.1) is proved in [43, Theorem 1.1]. In general, [43] presents the following result.

Theorem 3.5 ([43])**.** *Assume that for each* t,x, $b(t,x,\cdot)$ *and* $\sigma(t,x,\cdot)$ *are weakly continuous, and there exist* $c_0>1$ *and* $\gamma\in(0,1]$ *such that for all* $t\ge 0,x,y,\xi\in\mathbb{R}^d$ *and* $\mu\in\mathscr{P}(\mathbb{R}^d)$,

$$c_0^{-1}|\xi|\le|\sigma(t,x,\mu)\xi|\le c_0|\xi|,\quad |\sigma(t,x,\mu)-\sigma(t,y,\mu)|\le|x-y|^\gamma.$$

Moreover, under the weak topology of $\mathscr{P}(\mathbb{R}^d)$,

$$\sup_{\mu\in C([0,T];\mathscr{P}(\mathbb{R}^d))}\|b^\mu\|_{\tilde L_p^q(T)}<\infty$$

holds for some $(p,q)\in\mathscr{I}_1$. *Then for any* $\beta>2$ *and* $\nu\in\mathscr{P}_\beta(\mathbb{R}^d)$, *there exists a weak solution to* (1.1) *with initial distribution* ν. *If in addition,* $\sigma(t,x,\mu)$ *does not depend on* μ, $|\nabla\sigma|\in\tilde L_{p_1}^{q_1}(T)$ *and*

$$\|b(t,\cdot,\mu)-b(t,\cdot,\nu)\|_{\tilde L_p}\le\ell_t\|\mu-\nu\|_{\theta,TV},\quad \mu,\nu\in\mathscr{P}_\theta$$

for some $\ell\in L^q([0,T])$, $\theta\ge 1$ *and* $(p_1,q_1)\in\mathscr{I}_1$, *then for any* $\beta>2\theta$, (1.1) *has well-posedness from time* 0 *for initial distributions in* $\mathscr{P}_\beta(\mathbb{R}^d)$.

The following weak existence for (1.1) with supercritical drift is due to [57].

Theorem 3.6 ([57])**.** *Let* $\sigma=\sqrt{2}I_{d\times d}$, $b(t,x,\mu)=\int_{\mathbb{R}^d}K(t,x,y)\mu(\mathrm{d}y)$ *for some measurable function* K *on* $[0,T]\times\mathbb{R}^d\times\mathbb{R}^d$ *such that* $\mathrm{div}K(t,\cdot,y)\le 0$ *and*

$$K(t,x,y)\le h_t(x,y)$$

holds for some $(p,q)\in\mathscr{I}_2$ *and* $h\ge 0$ *with* $\|h\|_{L^q([0,T];\tilde L_p)}<\infty$. *Then for any* $\beta\in[0,2/(\frac{d}{p}+\frac{2}{q}))$ *and* $\nu\in\mathscr{P}_\beta(\mathbb{R}^d)$, (1.1) *has a weak solution with initial distribution* ν.

3.3.2 Integrability in time-space with linear growth in distribution

Comparing with above results, besides the singularity in x in the following we also allow $b(t,x,\mu)$ to have a linear growth in μ.

(H_3^2) Let $\theta \geq 1$. There exists a constant $K > 0$ such that for any $t \in [0,T], x,y \in \mathbb{R}^d$ and $\mu,\nu \in \mathscr{P}_\theta$,

$$\begin{aligned}&\|\sigma(t,x,\mu)\|^2 \vee \|(\sigma\sigma^*)^{-1}(t,x,\mu)\| \leq K,\\ &\|\sigma(t,x,\mu)-\sigma(t,y,\nu)\| \leq K\big(|x-y| + \mathbb{W}_\theta(\mu,\nu)\big),\\ &\|\{\sigma(t,x,\mu)-\sigma(t,y,\mu)\} - \{\sigma(t,x,\nu)-\sigma(t,y,\nu)\}\| \leq K|x-y|\mathbb{W}_\theta(\mu,\nu).\end{aligned}$$

Moreover, there exists nonnegative $f \in \tilde{L}_p^q(T)$ for some $(p,q) \in \mathscr{I}_1$ such that

$$\begin{aligned}&|b(t,x,\mu)| \leq (1+\|\mu\|_\theta) f_t(x),\\ &|b(t,x,\mu)-b(t,x,\nu)| \leq f_t(x)\|\mu-\nu\|_{\theta,TV}, \quad t\in[0,T], x\in\mathbb{R}^d, \mu,\nu\in\mathscr{P}_\theta.\end{aligned}$$

Theorem 3.7 ([24]). *Assume (H_3^2). Then (1.1) is well-posed for initial distributions in $\mathscr{P}_{\theta+} := \cap_{m>\theta}\mathscr{P}_m$, and the solution satisfies $\mathscr{L}_{X_\cdot} \in C([0,T];\mathscr{P}_\theta)$, the space of continuous maps from $[0,T]$ to $\mathscr{P}_\theta$ under the metric $\mathbb{W}_\theta$. Moreover,*

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X_t|^\theta\Big] < \infty. \tag{3.6}$$

3.3.3 Drifts with time-space integrable and Lipschitz terms

In this part we allow the drift to include a Lipschitz continuous term in x, but the price we have to pay is that the singular term is in $L_p^q(T)$ rather than $\tilde{L}_p^q(T)$ and the diffusion does not depend on distribution.

For any $p,q \geq 1$, let $L_p^q(T)$ be the space of measurable functions f on $[0,T]\times\mathbb{R}^d$ such that

$$\|f\|_{L_p^q(T)} := \left(\int_0^T \left(\int_{\mathbb{R}^d} |f(t,x)|^p \mathrm{d}x\right)^{\frac{q}{p}} \mathrm{d}t\right)^{\frac{1}{q}} < \infty.$$

(H_3^{3a}) $\sigma(t,x,\mu) = \sigma(t,x)$ does not depend on μ and is uniformly continuous in $x \in \mathbb{R}^d$ uniformly in $t \in [0,T]$; the weak gradient $\nabla\sigma(t,\cdot)$ exists for a.e. $t \in [0,T]$ satisfying $|\nabla\sigma|^2 \in L_p^q(T)$ for some $(p,q) \in \mathscr{I}_1$; and there exists a constant $K_1 \geq 1$ such that

$$K_1^{-1} I_{d\times d} \leq (\sigma\sigma^*)(t,x) \leq K_1 I_{d\times d}, \quad (t,x) \in [0,T]\times\mathbb{R}^d. \tag{3.7}$$

(H_3^{3b}) $b = \bar{b} + \hat{b}$, where $\bar{b}$ and $\hat{b}$ satisfy

$$\begin{aligned}&|\hat{b}(t,x,\gamma) - \hat{b}(t,y,\tilde{\gamma})| + |\bar{b}(t,x,\gamma) - \bar{b}(t,x,\tilde{\gamma})|\\ &\leq K_2(\|\gamma-\tilde{\gamma}\|_{TV} + \mathbb{W}_\theta(\gamma,\tilde{\gamma}) + |x-y|), \quad t\in[0,T], x,y\in\mathbb{R}^d, \gamma,\tilde{\gamma}\in\mathscr{P}_\theta(\mathbb{R}^d)\end{aligned} \tag{3.8}$$

for some constants $\theta, K_2 \geq 1$, and for (p,q) in (H_3^{3a}), it holds that

$$\sup_{t\in[0,T],\gamma\in\mathscr{P}_\theta(\mathbb{R}^d)} |\hat{b}(t,0,\gamma)| + \sup_{\mu\in C([0,T];\mathscr{P}_\theta(\mathbb{R}^d))} \||\bar{b}^\mu|^2\|_{L_p^q(T)} < \infty. \tag{3.9}$$

(H_3^{3c}) For any $\mu \in \mathscr{B}([0,T];\mathscr{P}(\mathbb{R}^d))$, the class of measurable maps from $[0,T]$ to $\mathscr{P}(\mathbb{R}^d)$, $|b^\mu|^2 \in L^q_{p,loc}(T)$ for (p,q) in (H_3^{3a}). Moreover, there exists an increasing function $\Gamma : [0,\infty) \to (0,\infty)$ satisfying $\int_1^\infty \frac{1}{\Gamma(x)}\mathrm{d}x = \infty$ such that

$$\langle b(t,x,\delta_0),x\rangle \le \Gamma(|x|^2),\quad t\in[0,T], x\in\mathbb{R}^d. \tag{3.10}$$

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

$$|b(t,x,\gamma)-b(t,x,\tilde\gamma)| \le K_3\|\gamma-\tilde\gamma\|_{TV},\quad t\in[0,T], x\in\mathbb{R}^d, \gamma,\tilde\gamma\in\mathscr{P}(\mathbb{R}^d). \tag{3.11}$$

Theorem 3.8 ([22]). *Assume* (H_3^{3a}).

(1) *If* (H_3^{3c}) *holds, then* (1.1) *is well-posed for initial initial distributions in* $\mathscr{P}_\theta(\mathbb{R}^d)$. *Moreover,*

$$\|P_t^*\mu_0 - P_t^*\nu_0\|_{TV}^2 \le 2\mathrm{e}^{\frac{K_1K_3^2t}{2}}\|\mu_0-\nu_0\|_{TV}^2,\quad t\in[0,T], \mu_0,\nu_0\in\mathscr{P}(\mathbb{R}^d). \tag{3.12}$$

(2) *Let* (H_3^{3b}) *hold. Then* (1.1) *is well-posed for initial distributions in* $\mathscr{P}_\theta(\mathbb{R}^d)$. *Moreover, for any* $m\in(\frac{\theta}{2},\infty)\cap[1,\infty)$, *there exists a constant* $c>0$ *such that*

$$\begin{aligned}&\|P_t^*\mu_0 - P_t^*\nu_0\|_{TV} + \mathbb{W}_\theta(P_t^*\mu_0,P_t^*\nu_0)\\ &\le c\big\{\|\mu_0-\nu_0\|_{TV} + \mathbb{W}_{2m}(\mu_0,\nu_0)\big\},\quad t\in[0,T], \mu_0,\nu_0\in\mathscr{P}_\theta(\mathbb{R}^d).\end{aligned} \tag{3.13}$$

4 Regularity estimates

In this section, we introduce some results on the regularity of distributions for the DDSDE (1.1). We first establish the log-Harnack inequality, which implies the "gradient estimate" and entropy estimate, then establish the Bismut formula for the Lions derivative of the distribution, and finally study the derivative estimate on the distribution. In the first two cases the noise does not depend on the distribution, while the last part applies also to distribution dependent noise.

4.1 Log-Harnack inequality

The dimension-free Harnack inequality was founded in [47] for diffusion semigroups on Riemannian manifolds, and as a weaker version the log-Harnack inequality was introduced in [42, 49] for (reflecting) diffusion processes and SDEs. See the monograph [50] for the study of these type inequalities and applications. In this part, we introduce the log-Harnack inequality established in [51] and [41] for DDSDEs with non-degenerate and degenerate noise respectively. We will only consider distribution, independent noise, since the log-Harnack inequality is not yet available for DDSDEs with distribution dependent noise.

4.1.1 The non-degenerate case

Consider the following special version of (1.1):

$$\mathrm{d}X_t = b(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \tag{4.1}$$

where b and σ satisfy the following assumption.

(H_4^1) $\sigma(t,x)$ is invertible and Lipschitzian in x locally uniformly in $t \ge 0$, and there exist increasing functions $\kappa_0, \kappa_1, \kappa_2, \lambda : [0,\infty) \to (0,\infty)$ such that for any $t \in [0,T], x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$, we have

$$\|\sigma(t,\cdot)^{-1}\|_\infty \le \lambda(t), \quad |b(t,0,\mu)|^2 + \|\sigma(t,x)\|^2 \le \kappa_0(t)(1 + |x|^2 + \mu(|\cdot|^2)), \tag{4.2}$$

$$\begin{aligned} &2\langle b(t,x,\mu) - b(t,y,\nu), x - y\rangle + \|\sigma(t,x) - \sigma(t,y)\|_{HS}^2 \\ &\le \kappa_1(t)|x-y|^2 + \kappa_2(t)|x-y|\mathbb{W}_2(\mu,\nu). \end{aligned} \tag{4.3}$$

Obviously, (H_4^1) implies assumptions (H_3^1) for $\theta = 2$, so that Theorem 3.3 ensures the well-posedness of (4.1) with initial distributions in $\mathscr{P}_2(\mathbb{R}^d)$. For any $f \in \mathscr{B}_b(\mathbb{R}^d)$, consider

$$P_{s,t}f(\mu) := \mathbb{E}^\mu f(X_{s,t}) = \int_{\mathbb{R}^d} f(y)(P_{s,t}^*\mu)(\mathrm{d}y), \quad \mu \in \mathscr{P}_2(\mathbb{R}^d), t \ge s \ge 0,$$

where $\mathbb{E}^\mu$ is the expectation taking for the solution $(X_{s,t})_{t\ge s}$ of (4.1) with $\mathscr{L}_{X_{s,s}} = \mu$, recall that in this case we denote $P_{s,t}^*\mu = \mathscr{L}_{X_{s,t}}$. Let

$$\phi(s,t) = \lambda(t)^2\bigg(\frac{\kappa_1(t)}{1 - \mathrm{e}^{-\kappa_1(t)(t-s)}} + \frac{t\kappa_2(t)^2 \exp[2(t-s)(\kappa_1(t) + \kappa_2(t))]}{2}\bigg), \quad 0 \le s < t.$$

Theorem 4.1 ([51]). *Assume (H_4^1) and let $t > s \ge 0$. Then for any $\mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^d)$,*

$$(P_{s,t}\log f)(\nu_0) \le \log(P_{s,t}f)(\mu_0) + \phi(s,t)\mathbb{W}_2(\mu_0,\nu_0)^2, \quad f \in \mathscr{B}_b^+(\mathbb{R}^d).$$

Consequently, the following assertions hold:

(1) *For any $\mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^d)$,*

$$\|P_{s,t}^*\mu_0 - P_{s,t}^*\nu_0\|_{TV} \le \sqrt{2\phi(s,t)}\mathbb{W}_2(\mu_0,\nu_0).$$

(2) *For any $\mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, $P_{s,t}^*\mu_0$ and $P_{s,t}^*\nu_0$ are equivalent and the Radon-Nykodim derivative satisfies the entropy estimate*

$$\mathrm{Ent}(P_{s,t}^*\nu_0|P_{s,t}^*\mu_0) := \int_{\mathbb{R}^d}\bigg\{\log\frac{\mathrm{d}P_{s,t}^*\nu_0}{\mathrm{d}P_{s,t}^*\mu_0}\bigg\}\mathrm{d}P_{s,t}^*\nu_0 \le \phi(s,t)\mathbb{W}_2(\mu_0,\nu_0)^2.$$

Idea of Proof. We only consider $s = 0$. According to the method of coupling by change of measures summarized in [50, Section 1.1], the main steps of the proof include:

(S1) Let $(X_t)_{t\geq 0}$ solve (4.1) with $\mathscr{L}_{X_0} = \mu_0$. By the uniqueness we have $\mu_t := \mathbb{P}_t^*\mu_0 = \mathscr{L}_{X_t}$, and the equation (4.1) reduces to

$$\mathrm{d}X_t = b_t(X_t, \mu_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t. \tag{4.4}$$

(S2) Construct a process $(Y_t)_{t\in[0,T]}$ such that for a weighted probability measure $\mathbb{Q} := R_T\mathbb{P}$,

$$X_T = Y_T \ \mathbb{Q}\text{-a.s.}, \quad \text{and } \mathscr{L}_{Y_T}|_{\mathbb{Q}} = P_T^*\nu_0 =: \nu_T. \tag{4.5}$$

Obviously, (S1) and (S2) imply

$$(P_Tf)(\mu_0) = \mathbb{E}[f(X_T)] \text{ and } (P_Tf)(\nu_0) = \mathbb{E}_{\mathbb{Q}}[f(Y_T)] = \mathbb{E}[R_Tf(X_T)], \quad f \in \mathscr{B}_b(\mathbb{R}^d). \tag{4.6}$$

Combining this with Young's inequality, we obtain the log-Harnack inequality:

$$\begin{aligned}(P_T\log f)(\nu_0) &\leq \mathbb{E}[R_T\log R_T] + \log\mathbb{E}[f(X_T)] \\ &= \log(P_Tf)(\mu_0) + \mathbb{E}[R_T\log R_T], \quad f \in \mathscr{B}_b^+(\mathbb{R}^d).\end{aligned} \tag{4.7}$$

□

4.1.2 The degenerate case

Consider the following distribution dependent stochastic Hamiltonian system for $(X_t, Y_t) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$:

$$\begin{cases}\mathrm{d}X_t = \big(AX_t + BY_t\big)\mathrm{d}t, \\ \mathrm{d}Y_t = Z(t, (X_t, Y_t), \mathscr{L}_{(X_t,Y_t)})\mathrm{d}t + \sigma_t\mathrm{d}W_t,\end{cases} \tag{4.8}$$

where A is a $d_1\times d_1$-matrix, B is a $d_1\times d_2$-matrix, σ is a $d_2\times d_2$-matrix, W_t is the d_2-dimensional Brownian motion on a complete filtration probability space $(\Omega, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, and

$$Z : [0,\infty) \times \mathbb{R}^{d_1+d_2} \times \mathscr{P}_2(\mathbb{R}^{d_1+d_2}) \to \mathbb{R}^{d_2}, \quad \sigma : [0,\infty) \to \mathbb{R}^{d_2} \otimes \mathbb{R}^{d_2}$$

are measurable. We assume

(H_4^2) $\sigma(t)$ is invertible, there exists a locally bounded function $K : [0,\infty) \to [0,\infty)$ such that

$$\|\sigma(t)^{-1}\| \leq K(t), \quad |Z(t,x,\mu) - Z(t,y,\nu)| \leq K(t)\big\{|x-y| + \mathbb{W}_2(\mu,\nu)\big\}$$

holds for all $t \geq 0, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$ and $x, y \in \mathbb{R}^{d_1+d_2}$, and A, B satisfy the following Kalman's rank condition for some $k \geq 1$:

$$\mathrm{Rank}[A^0B, \cdots, A^{k-1}B] = d_1, \quad A^0 := I_{d_1\times d_1}.$$

Obviously, this assumption implies (H_3^1), so that (4.8) has a unique solution (X_t, Y_t) for any initial value (X_0, Y_0) with $\mu := \mathscr{L}_{(X_0,Y_0)} \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$. Let $P_t^*\mu := \mathscr{L}_{(X_t,Y_t)}$ and

$$(P_t f)(\mu) := \int_{\mathbb{R}^{d_1+d_2}} f \mathrm{d}P_t^*\mu, \quad t \geq 0, f \in \mathscr{B}_b(\mathbb{R}^{d_1+d_2}).$$

By [51, Theorem 3.1], the Lipschitz continuity of Z implies

$$\mathbb{W}_2(P_t^*\mu, P_t^*\nu) \leq \mathrm{e}^{Kt}\mathbb{W}_2(\mu,\nu), \quad t \geq 0, \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}) \tag{4.9}$$

for some constant $K > 0$. The following result is due to [41, Section 5.1].

Theorem 4.2 ([41])**.** *Assume (H_4^2). Then there exists an increasing function $C : [0,\infty) \to (0,\infty)$ such that for any $T > 0$,*

$$(P_T \log f)(\nu) \leq \log(P_T f)(\mu) + \frac{C(T)}{T^{4k-1}\wedge 1}\mathbb{W}_2(\mu,\nu)^2, \quad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}), f \in \mathscr{B}_b^+(\mathbb{R}^{d_1+d_2}).$$

Consequently,

$$\mathrm{Ent}(P_T^*\nu | P_T^*\mu) \leq \frac{C(T)}{T^{4k-1}\wedge 1}\mathbb{W}_2(\mu,\nu)^2, \quad T > 0, \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}). \tag{4.10}$$

4.2 Bismut formula for the Lions derivative of $P_t f$

We first introduce the intrinsic and Lions derivatives for functionals of measures, then present the Bismut formula for the Lions derivative of $P_t f$ for non-degenerate and degenerate DDSDEs respectively. The main results are taken from [38], see also [2] for extensions to distribution-path dependent SDEs.

4.2.1 Intrinsic and Lions derivatives

Definition 4.1. Let $f : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$.

(1) If for any $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$,

$$D_\phi^I f(\mu) := \lim_{\varepsilon \downarrow 0} \frac{f(\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1}) - f(\mu)}{\varepsilon} \in \mathbb{R}$$

exists, and is a bounded linear functional in ϕ, we call f intrinsic differentiable at μ. In this case, there exists a unique $D^I f(\mu) \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$ such that

$$\langle D^I f(\mu), \phi\rangle_{L^2(\mu)} = D_\phi^I f(\mu), \quad \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu).$$

We call $D^I f(\mu)$ the intrinsic derivative of f at μ. If f is intrinsic differentiable at all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, we call it intrinsic differentiable on $\mathscr{P}_2(\mathbb{R}^d)$ and denote

$$\|D^I f(\mu)\| := \|D^I f(\mu)\|_{L^2(\mu)} = \left(\int_{\mathbb{R}^d} |D^I f(\mu)|^2 \mathrm{d}\mu\right)^{\frac{1}{2}}.$$

(2) If f is intrinsic differentiable and for any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$,

$$\lim_{\|\phi\|_{L^2(\mu)} \to 0} \frac{f(\mu \circ (\mathrm{Id} + \phi)^{-1}) - f(\mu) - D^I_\phi f(\mu)}{\|\phi\|_{L^2(\mu)}} = 0,$$

we call f L-differentiable on $\mathscr{P}_2(\mathbb{R}^d)$. In this case, $D^I f(\mu)$ is also denoted by $D^L f(\mu)$, and is called the L-derivative of f at μ.

Intrinsic derivative was first introduced in [1] in the configuration space over a Riemannian manifold, while the L-derivative appeared in the Lecture notes [9] for the study of mean field games and is also called Lions derivative in references.

Note that the derivative $D^I f(\mu) \in L^2(\mathbb{R}^d \to \mathbb{R}^d; \mu)$ is μ-a.e. defined. In applications, we take its continuous version if exists. The following classes of L-differentiable functions are often used in analysis:

(a) $f \in C^1(\mathscr{P}_2(\mathbb{R}^d))$: if f is L-differentiable such that for every $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, there exists a μ-version $D^L f(\mu)(\cdot)$ such that $D^L f(\mu)(x)$ is jointly continuous in $(x, \mu) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$.

(b) $f \in C^1_b(\mathscr{P}_2(\mathbb{R}^d))$: if $f \in C^1(\mathscr{P}_2(\mathbb{R}^d))$ and $D^L f(\mu)(x)$ is bounded.

(c) $f \in C^2(\mathscr{P}_2(\mathbb{R}^d))$: if $f \in C^1(\mathscr{P}_2(\mathbb{R}^d))$ and $Df(\mu)(x)$ is L-differentiable in μ and differentiable in $x \in \mathbb{R}^d$, such that $\nabla\{D^L f(\mu)\}(x)$ and

$$(D^L)^2 f(\mu)(x, y) := \big(\{D^L[D^L f(\mu)(x)]_i(y)\}_j\big)_{1 \le i,j \le d} \in \mathbb{R}^d \otimes \mathbb{R}^d$$

are jointly continuous in $(\mu, x, y) \in \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d$.

(d) $f \in C^2_b(\mathscr{P}_2(\mathbb{R}^d))$: if $f \in C^2(\mathscr{P}_2(\mathbb{R}^d))$ and all derivatives $D^L f(\mu)(x), (D^L)^2 f(\mu)(x, y)$ and $\nabla(D^L f(\mu))(x)$ are bounded.

(e) $f \in C^{1,1}(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$: if f is a continuous function on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$ such that $f(\cdot, \mu) \in C^1(\mathbb{R}^d)$ for $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $f(x, \cdot) \in C^1(\mathscr{P}_2(\mathbb{R}^d))$ for $x \in \mathbb{R}^d$, and

$$\nabla f(x, \mu), \quad D^L f(x, \mu)(y)$$

are jointly continuous in $(x, \mu, y) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$. If moreover these derivatives are bounded, we denote $f \in C^{1,1}_b(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$.

(f) $f \in C^{2,2}(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$, if f is a continuous function on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$ such that $f(\cdot, \mu) \in C^2(\mathbb{R}^d)$ for $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $f(x, \cdot) \in C^2(\mathscr{P}_2(\mathbb{R}^d))$ for $x \in \mathbb{R}^d$,

$$(D^L \nabla f)(x, \mu)(y) := \big(\{D^L[\partial_{x_i} f(x, \mu)]\}_j\big)_{1 \le i,j \le d} \in \mathbb{R}^d \otimes \mathbb{R}^d$$

exists, and all derivatives

$$\begin{aligned}&\nabla f(x, \mu),\ \nabla^2 f(x, \mu),\ D^L f(x, \mu)(y),\ (D^L \nabla f)(x, \mu)(y)\\ &\nabla\{D^L f(x, \mu)(\cdot)\}(y),\ (D^L)^2 f(x, \mu)(y, z)\end{aligned}$$

are jointly continuous in $(x, \mu, y, z) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d$. If moreover these derivatives are bounded, we denote $f \in C^{2,2}_b(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$.

Consider $f \in \mathscr{F}C_b^2(\mathscr{P}_2(\mathbb{R}^d))$, i.e.

$$f(\mu) = g(\mu(h_1), \cdots, \mu(h_n)), \ n \geq 1, g \in C^2(\mathbb{R}^n), h_i \in C_b^2(\mathbb{R}^d).$$

Then it is easy to see that $f \in C_b^2(\mathscr{P}_2(\mathbb{R}^d))$ with

$$\begin{aligned}
D^L f(\mu)(y) &= \sum_{i=1}^d (\partial_i g)(\mu(h_1), \cdots, \mu(h_n)) \nabla h_i(y), \\
\nabla\{D^L f(\mu)\}(y) &= \sum_{i=1}^d (\partial_i g)(\mu(h_1), \cdots, \mu(h_n)) \nabla^2 h_i(y), \\
(D^L)^2 f(\mu)(y, z) &= \sum_{i,j=1}^d (\partial_i \partial_j g)(\mu(h_1), \cdots, \mu(h_n)) \{\nabla h_i(y)\} \otimes \{\nabla h_j(z)\}.
\end{aligned}$$

4.2.2 Bismut formula for non-degenerate DDSDEs

Consider the DDSDE (4.1) with coefficients satisfying the following assumption.

(H_4^3) In addition to (H_4^1), $b_t, \sigma_t \in C^{1,1}(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$ such that

$$\begin{aligned}
&\max\Big\{\|\nabla b_t(\cdot, \mu)(x)\|, \|D^L b_t(x, \cdot)(\mu)\|, \ \frac{1}{2}\|\nabla \sigma_t(\cdot, \mu)(x)\|^2, \frac{1}{2}\|D^L \sigma_t(x, \cdot)(\mu)\|^2\Big\} \\
&\leq K(t), \quad t \geq 0, x \in \mathbb{R}^d, \mu \in \mathscr{P}_2(\mathbb{R}^d)
\end{aligned}$$

holds for some continuous function $K : [0, \infty) \to [0, \infty)$.

By Theorem 3.3, for any initial value $X_0 \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$, (4.1) has a unique solution $(X_t)_{t \geq 0}$. Let $P_t^* \mu = \mathscr{L}_{X_t}$ for $\mathscr{L}_{X_0} = \mu$, and consider the L-derivative of the functionals in μ:

$$P_T f(\mu) := \mathbb{E}^\mu f(X_t) = \int_{\mathbb{R}^d} f(y)(P_T^* \mu)(\mathrm{d}y), \quad T > 0, f \in \mathscr{B}_b(\mathbb{R}^d).$$

Given $\phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu)$, the following linear SDE has a unique solution v_t^ϕ on $\mathbb{R}^d$:

$$\begin{aligned}
\mathrm{d}v_t^\phi = &\Big\{\nabla_{v_t^\phi} b(t, \cdot, \mathscr{L}_{X_t})(X_t) + \big(\mathbb{E}\langle D^L b(t, y, \cdot)(\mathscr{L}_{X_t})(X_t), v_t^\phi\rangle\big)\big|_{y = X_t}\Big\}\mathrm{d}t \\
&+ \big\{\nabla_{v_t^\phi} \sigma(t, \cdot)(X_t)\big\}\mathrm{d}W_t, \quad v_0^\phi = \phi(X_0), \ t \geq 0.
\end{aligned} \tag{4.11}$$

The following result is taken from [39, Theorem 2.1 and Corollary 2.2].

Theorem 4.3 ([39])**.** *Assume (H_4^3). Then for any $f \in \mathscr{B}_b(\mathbb{R}^d), \mu \in \mathscr{P}_2(\mathbb{R}^d)$ and $T > 0$, $P_T f$ is L-differentiable at μ such that for any $g \in C^1([0, T])$ with $g_0 = 0$ and $g_T = 1$,*

$$D_\phi^L (P_T f)(\mu) = \mathbb{E}\Big[f(X_T) \int_0^T \langle g_t' \sigma_t(X_t, \mathscr{L}_{X_t})^{-1} v_t^\phi, \ \mathrm{d}W_t\rangle\Big], \ \phi \in L^2(\mathbb{R}^d \to \mathbb{R}^d, \mu),$$

where X_t solves (4.1) for $\mathscr{L}_{X_0} = \mu$. Moreover, the limit

$$D^L_\phi P_T^*\mu := \lim_{\varepsilon\downarrow 0} \frac{P_T^*\mu\circ(\mathrm{Id}+\varepsilon\phi)^{-1} - P_T^*\mu}{\varepsilon} = \psi P_T^*\mu \tag{4.12}$$

exists in the total variational norm, where ψ is the unique element in $L^2(\mathbb{R}^d\to\mathbb{R}, P_T^*\mu)$ such that $\psi(X_T) = \mathbb{E}\big(\int_0^T \langle g_t'\sigma_t(X_t,\mathscr{L}_{X_t})^{-1}v_t^\phi,\ \mathrm{d}W_t\rangle\big|X_T\big)$, and $(\psi P_T^*\mu)(A) := \int_A \psi \mathrm{d}P_T^*\mu,\ A\in\mathscr{B}(\mathbb{R}^d)$. Consequently, for any $T>0, f\in\mathscr{B}_b(\mathbb{R}^d)$ and $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^d)$,

$$\|D^L(P_Tf)(\mu)\|^2 \le \frac{(P_Tf^2)(\mu) - (P_tf(\mu))^2}{\int_0^T \lambda_t^{-2}\mathrm{e}^{-8K(t)t}\mathrm{d}t},$$

$$\|P_T^*\mu - P_T^*\nu\|_{TV}^2 := 4\sup_{A\in\mathscr{B}(\mathbb{R}^d)} |(P_T^*\mu)(A) - (P_T^*\nu)(A)|^2 \le \frac{4\mathbb{W}_2(\mu,\nu)^2}{\int_0^T \lambda_t^{-2}\mathrm{e}^{-8K(t)t}\mathrm{d}t}.$$

4.2.3 Bismut formula for degenerate DDSDEs

Consider the following distribution dependent stochastic Hamiltonian system for $X_t = (X_t^{(1)}, X_t^{(2)})$ on $\mathbb{R}^{d_1+d_2} = \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$:

$$\begin{cases} \mathrm{d}X_t^{(1)} = b_t^{(1)}(X_t)\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = b_t^{(2)}(X_t,\mathscr{L}_{X_t})\mathrm{d}t + \sigma_t\mathrm{d}W_t, \end{cases} \tag{4.13}$$

where $(W_t)_{t\ge 0}$ is a d_2-dimensional Brownian motion as before, and for each $t\ge 0$, σ_t is an invertible $d_2\times d_2$-matrix,

$$b_t = (b_t^{(1)}, b_t^{(2)}) : \mathbb{R}^{d_1+d_2}\times\mathscr{P}_2(\mathbb{R}^{d_1+d_2}) \to \mathbb{R}^{d_1+d_2}$$

is measurable with $b_t^{(1)}(x,\mu) = b_t^{(1)}(x)$ independent of the distribution μ. Let $\nabla = (\nabla^{(1)},\nabla^{(2)})$ be the gradient operator on $\mathbb{R}^{d_1+d_2} = \mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$, where $\nabla^{(i)}$ is the gradient in the i-th component, $i=1,2$. Let $\nabla^2 = \nabla\nabla$ denote the Hessian operator on $\mathbb{R}^{d_1+d_2}$. We assume

(H_4^4) For every $t\ge 0$, $b_t^{(1)}\in C_b^2(\mathbb{R}^{d_1+d_2}\to\mathbb{R}^{d_1})$, $b_t^{(2)}\in C^{1,1}(\mathbb{R}^{d_1+d_2}\times\mathscr{P}_2(\mathbb{R}^{d_1+d_2})\to\mathbb{R}^{d_2})$, and there exists an increasing function $K:[0,\infty)\to[0,\infty)$ such that

$$\|\nabla b_t(\cdot,\mu)(x)\| + \|D^Lb_t^{(2)}(x,\cdot)(\mu)\| + \|\nabla^2 b_t^{(1)}(x)\| \le K(t),\quad t\ge 0, (x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d).$$

There exist $B\in\mathscr{B}_b([0,T]\to\mathbb{R}^{d_1}\otimes\mathbb{R}^{d_2})$, an increasing function $\theta\in C([0,T];\mathbb{R}^1)$ with $\theta_t>0$ for $t\in(0,T]$, and $\varepsilon\in(0,1)$ such that

$$\langle(\nabla^{(2)}b_t^{(1)} - B_t)B_t^*a, a\rangle \ge -\varepsilon|B_t^*a|^2,\quad a\in\mathbb{R}^{d_1},$$

$$\int_0^t s(T-s)K_{T,s}B_sB_s^*K_{T,s}^*\mathrm{d}s \ge \theta_t I_{d_1\times d_1},\quad t\in(0,T],$$

where for any $s\ge 0$, $\{K_{t,s}\}_{t\ge s}$ is the unique solution of the following linear random ODE on $\mathbb{R}^{d_1}\otimes\mathbb{R}^{d_1}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}K_{t,s} = (\nabla^{(1)}b_t^{(1)})(X_t)K_{t,s},\quad t\ge s, K_{s,s} = I_{d_1\times d_1}.$$

Example 4.1. Let

$$b_t^{(1)}(x) = Ax^{(1)} + Bx^{(2)}, \quad x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^{d_1+d_2}$$

for some $d_1 \times d_1$-matrix A and $d_1 \times d_2$-matrix B. If the Kalman's rank condition

$$\mathrm{Rank}[B, AB, \cdots, A^k B] = d_1$$

holds for some $k \geq 1$, then (H_4^4) is satisfied with $\theta_t = c_T t$ for some constant $c_T > 0$.

According to the proof of [52, Theorem 1.1], (H_4^4) implies that the matrices

$$Q_t := \int_0^t s(T-s) K_{T,s} \nabla^{(2)} b_s^{(1)}(X_s) B_s^* K_{T,s}^* \mathrm{d}s, \quad t \in (0,T]$$

are invertible with

$$\|Q_t^{-1}\| \leq \frac{1}{(1-\varepsilon)\theta_t}, \quad t \in (0,T]. \tag{4.14}$$

For $(X_t)_{t\in[0,T]}$ solving (4.13) with $\mathscr{L}_{X_0} = \mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$ and $\phi = (\phi^{(1)}, \phi^{(2)}) \in L^2(\mathbb{R}^{d_1+d_2} \to \mathbb{R}^{d_1+d_2}, \mu)$, let

$$\begin{aligned}
\alpha_t^{(2)} &= \frac{T-t}{T}\phi^{(2)}(X_0) - \frac{t(T-t)B_t^* K_{T,t}^*}{\int_0^T \theta_s^2 \mathrm{d}s} \int_t^T \theta_s^2 Q_s^{-1} K_{T,0} \phi^{(1)}(X_0) \mathrm{d}s \\
&\quad - t(T-t) B_t^* K_{T,t}^* Q_T^{-1} \int_0^T \frac{T-s}{T} K_{T,s} \nabla^{(2)}_{\phi^{(2)}(X_0)} b_s^{(1)}(X_s) \mathrm{d}s, \\
\alpha_t^{(1)} &= K_{t,0}\phi^{(1)}(X_0) + \int_0^t K_{t,s} \nabla^{(2)}_{\alpha_s^{(2)}} b_s^{(1)}(X_s(x))\, \mathrm{d}s, \quad t \in [0,T],
\end{aligned}$$

and define

$$\begin{aligned}
h_t^\alpha := \int_0^t \sigma_s^{-1} \Big\{ &\big(\mathbb{E}\langle D^L b_s^{(2)}(y,\cdot)(\mathscr{L}_{X_s})(X_s), \alpha_s\rangle\big)\big|_{y=X_s} \\
&+ \nabla_{\alpha_s} b_s^{(2)}(\cdot, \mathscr{L}_{X_s})(X_s) - (\alpha_s^{(2)})' \Big\} \mathrm{d}s, \quad t \in [0,T].
\end{aligned} \tag{4.15}$$

Let $(D^*, \mathscr{D}(D^*))$ be the Malliavin divergence operator associated with the Brownian motion $(W_t)_{t\in[0,T]}$. The following result is due to [39, Theorem 2.3].

Theorem 4.4 ([39])**.** *Assume(H_4^4). Then $h^\alpha \in \mathscr{D}(D^*)$ with $\mathbb{E}|D^*(h^\alpha)|^p < \infty$ for all $p \in [1,\infty)$. Moreover, for any $f \in \mathscr{B}_b(\mathbb{R}^{d_1+d_2})$ and $T>0$, $P_T f$ is L-differentiable such that*

$$D_\phi^L (P_T f)(\mu) = \mathbb{E}\big[f(X_T)\, D^*(h^\alpha)\big]$$

holds for $\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}), \phi \in L^2(\mathbb{R}^{d_1+d_2} \to \mathbb{R}^{d_1+d_2}, \mu)$ and h^α in (4.15). Consequently:

(1) *The formula (4.12) holds for the unique* $\psi \in L^2(\mathbb{R}^{d_1+d_2} \to \mathbb{R}, P_T^*\mu)$ *such that* $\psi(X_T) = \mathbb{E}(D^*(h^\alpha)|X_T)$.

(2) *There exists a constant* $c \geq 0$ *such that for any* $T > 0$,

$$\|D^L(P_T f)(\mu)\| \leq c\sqrt{P_T|f|^2(\mu) - (P_T f)^2(\mu)}\frac{\sqrt{T}(T^2 + \theta_T)}{\int_0^T \theta_s^2 \mathrm{d}s}, \quad f \in \mathscr{B}_b(\mathbb{R}^{d_1+d_2}),$$

$$\|P_T^*\mu - P_T^*\nu\|_{TV} \leq c\mathbb{W}_2(\mu,\nu)\frac{\sqrt{T}(T^2 + \theta_T)}{\int_0^T \theta_s^2 \mathrm{d}s}, \quad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}).$$

4.3 Lions derivative estimates on $P_t f$

In this part we estimate $D^L P_T f$ for DDSDE with σ also depending on μ, which thus extends the corresponding derivative estimate presented in Theorem 4.3.

Consider the DDSDE (1.1) with coefficients satisfying the following assumption which, by Theorem 3.3, implies the well-posedness.

(H_4^5) For any $t \geq 0$, $b_t, \sigma_t \in C^{1,1}(\mathbb{R}^d \times \mathscr{P}_2)$, and there exists an increasing function $K : [0,\infty) \to [1,\infty)$ such that for any $t \geq 0, x, y \in \mathbb{R}^d$ and $\mu \in \mathscr{P}_2(\mathbb{R}^d)$,

$$K_t^{-1} I_{d\times d} \leq (\sigma_t\sigma_t^*)(x,\mu) \leq K_t I_{d\times d},$$

$$\begin{aligned}|b_t(x,\mu)| &+ \|\nabla b_t(\cdot,\mu)(x)\| + \|D^L\{b_t(x,\cdot)\}(\mu)\| \\ &+ \|\nabla\{\sigma_t(\cdot,\mu)\}(x)\|^2 + \|D^L\{\sigma_t(x,\cdot)\}(\mu)\|^2 \leq K_t,\end{aligned}$$

$$\begin{aligned}&\|D^L\{b_t(x,\cdot)\}(\mu) - D^L\{b_t(y,\cdot)\}(\mu)\| + \|D^L\{\sigma_t(x,\cdot)\}(\mu) - D^L\{\sigma_t(y,\cdot)\}(\mu)\| \\ &\leq K_t|x-y|.\end{aligned}$$

Let $P_{s,t}f(\mu) := \mathbb{E}[f(X_{s,t})]$ for $f \in \mathscr{B}_b(\mathbb{R}^d)$ and $(X_{s,t})_{t\geq s\geq 0}$ solving (1.1) with $\mathscr{L}_{X_{s,s}} = \mu \in \mathscr{P}_2(\mathbb{R}^d)$. The following result is due to [23, Theorem 1.1].

Theorem 4.5 ([23])**.** *Assume* (H_4^5). *Then for any* $t > s \geq 0$ *and* $f \in \mathscr{B}_b(\mathbb{R}^d)$, $P_{s,t}f$ *is L-differentiable, and there exists an increasing function* $C : [0,\infty) \to (0,\infty)$ *such that*

$$\|D^L P_t f(\mu)\| \leq \frac{C_t\|f\|_\infty}{\sqrt{t-s}}, \quad t > s, f \in \mathscr{B}_b(\mathbb{R}^d).$$

Consequently, for any $t > 0$ *and* $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$,

$$\|P_{s,t}^*\mu - P_{s,t}^*\nu\|_{TV} := 2\sup_{\|f\|_\infty\leq 1}|P_{s,t}f(\mu) - P_{s,t}f(\nu)| \leq \frac{2C_t}{\sqrt{t-s}}\mathbb{W}_2(\mu,\nu).$$

5 Exponential ergodicity in entropy

The convergence in entropy for stochastic systems is an important topic in both probability theory and mathematical physics, and has been well studied for Markov processes by using the log-Sobolev inequality, see for instance [7] and references therein. However, the existing results derived in the literature do not apply to DDSDEs. In 2003, Carrillo, McCann and Villani [11] proved the exponential convergence in a mean field entropy of the following granular media equation for probability density functions $(\rho_t)_{t\geq 0}$ on $\mathbb{R}^d$:

$$\partial_t \rho_t = \Delta \rho_t + \mathrm{div}\big\{\rho_t \nabla (V + W * \rho_t)\big\}, \tag{5.1}$$

where the internal potential $V \in C^2(\mathbb{R}^d)$ satisfies $\mathrm{Hess}_V \geq \lambda I_{d\times d}$ for a constant $\lambda > 0$ and the $d\times d$-unit matrix $I_{d\times d}$, and the interaction potential $W \in C^2(\mathbb{R}^d)$ satisfies $W(-x) = W(x)$ and $\mathrm{Hess}_W \geq -\delta I_{d\times d}$ for some constant $\delta \in [0, \lambda/2)$. Recall that we write $M \geq \lambda I_d$ for a constant λ and a $d\times d$-matrix M, if $\langle Mv, v\rangle \geq \lambda |v|^2$ holds for any $v \in \mathbb{R}^d$. To introduce the mean field entropy, let $\mu_V(\mathrm{d}x) := \frac{\mathrm{e}^{-V(x)}\mathrm{d}x}{\int_{\mathbb{R}^d} \mathrm{e}^{-V(x)}\mathrm{d}x}$, recall the classical relative entropy

$$\mathrm{Ent}(\nu|\mu) := \begin{cases} \mu(\rho \log \rho), & \text{if } \nu = \rho\mu, \\ \infty, & \text{otherwise} \end{cases}$$

for $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$, and consider the free energy functional

$$E^{V,W}(\mu) := \mathrm{Ent}(\mu|\mu_V) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x-y)\mu(\mathrm{d}x)\mu(\mathrm{d}y), \quad \mu \in \mathscr{P}(\mathbb{R}^d),$$

where we set $E^{V,W}(\mu) = \infty$ if either $\mathrm{Ent}(\mu|\mu_V) = \infty$ or the integral term is not well defined. Then the associated mean field entropy $\mathrm{Ent}^{V,W}$ is defined by

$$\mathrm{Ent}^{V,W}(\mu) := E^{V,W}(\mu) - \inf_{\nu\in\mathscr{P}} E^{V,W}(\nu), \quad \mu \in \mathscr{P}(\mathbb{R}^d). \tag{5.2}$$

According to [11], for V and W satisfying the above mentioned conditions, $E^{V,W}$ has a unique minimizer μ_∞, and $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ for probability density ρ_t solving (5.1) converges to μ_∞ exponentially in the mean field entropy:

$$\mathrm{Ent}^{V,W}(\mu_t) \leq \mathrm{e}^{-(\lambda-2\delta)t}\mathrm{Ent}^{V,W}(\mu_0), \quad t \geq 0.$$

Recently, this result was generalized in [15] by establishing the uniform log-Sobolev inequality for the associated mean field particle systems, such that $\mathrm{Ent}^{V,W}(\mu_t)$ decays exponentially for a class of non-convex $V \in C^2(\mathbb{R}^d)$ and $W \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$, where $W(x,y) = W(y,x)$ and $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ for ρ_t solving the nonlinear PDE

$$\partial_t \rho_t = \Delta \rho_t + \mathrm{div}\big\{\rho_t \nabla (V + W \circledast \rho_t)\big\}, \tag{5.3}$$

where

$$W \circledast \rho_t := \int_{\mathbb{R}^d} W(\cdot, y)\rho_t(y)\mathrm{d}y. \tag{5.4}$$

In this case, $\mathrm{Ent}^{V,W}$ is defined in (5.2) for the free energy functional

$$E^{V,W}(\mu) := \mathrm{Ent}(\mu|\mu_V) + \frac{1}{2}\int_{\mathbb{R}^d\times\mathbb{R}^d} W(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y), \quad \mu \in \mathscr{P}(\mathbb{R}^d).$$

To study (5.3) using probability methods, we consider the following DDSDE with initial distribution μ_0:

$$\mathrm{d}X_t = \sqrt{2}\mathrm{d}B_t - \nabla\big\{V + W \circledast \mathscr{L}_{X_t}\big\}(X_t)\mathrm{d}t, \tag{5.5}$$

where B_t is the d-dimensional Brownian motion, $\mathscr{L}_{X_t}$ is the distribution of X_t, and

$$(W \circledast \mu)(x) := \int_{\mathbb{R}^d} W(x,y)\mu(\mathrm{d}y), \quad x \in \mathbb{R}^d, \mu \in \mathscr{P}(\mathbb{R}^d) \tag{5.6}$$

provided the integral exists. Let $\rho_t(x) = \frac{(\mathscr{L}_{X_t})(\mathrm{d}x)}{\mathrm{d}x}$, $t \ge 0$. By Itô's formula and the integration by parts formula, we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^d}(\rho_t f)(x)\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(X_t)] = \mathbb{E}\big[\big(\Delta - \nabla V - \nabla\{W \circledast \rho_t\}\big)f(X_t)\big]\\
&= \int_{\mathbb{R}^d}\rho_t(x)\big\{\Delta f - \langle \nabla V + \nabla\{W \circledast \rho_t\}, \nabla f\rangle\big\}(x)\mathrm{d}x\\
&= \int_{\mathbb{R}^d} f(x)\{\Delta\rho_t + \mathrm{div}[\rho_t\nabla V + \rho_t\nabla(W \circledast \rho_t)]\big\}(x)\mathrm{d}x, \quad t \ge 0,\ f \in C_0^\infty(\mathbb{R}^d).
\end{aligned}$$

Therefore, ρ_t solves (5.3). On the other hand, by this fact and the uniqueness of (5.3) and (5.5), if ρ_t solves (5.3) with $\mu_0(\mathrm{d}x) := \rho_0(x)\mathrm{d}x$, then $\rho_t(x)\mathrm{d}x = \mathscr{L}_{X_t}(\mathrm{d}x)$ for X_t solving (5.5) with $\mathscr{L}_{X_0} = \mu_0$.

To extend the study of [11, 15], we investigate the exponential convergence in entropy for the following DDSDE on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b(X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \tag{5.7}$$

where W_t is the m-dimensional Brownian motion on a complete filtration probability space $(\Omega, \{\mathscr{F}_t\}_{t\ge 0}, \mathbb{P})$,

$$\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m, \quad b : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$$

are measurable.

Unlike in [11, 15] where the mean field particle systems are used to estimate the mean field entropy, we use the log-Harnack inequality introduced in [49, 42] and the Talagrand inequality developed in [46, 7, 35]. Since the log-Harnack inequality is not yet available when σ depends on the distribution, in (5.7) we only consider distribution-free σ.

In the following subsections, we first present a criterion on the exponential convergence for DDSDEs by using the log-Harnack and Talagrand inequalities, then prove the exponential convergence for granular media type equations which generalizes the framework of [15], and finally consider exponential convergence for (5.7) with non-degenerate and degenerate noises respectively.

5.1 A criterion with application to Granular media type equations

In general, we consider the following DDSDE:

$$\mathrm{d}X_t = \sigma(X_t, \mathscr{L}_{X_t})\mathrm{d}W_t + b(X_t, \mathscr{L}_{X_t})\mathrm{d}t, \tag{5.8}$$

where W_t is the m-dimensional Brownian motion and

$$\sigma : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^m, \quad b : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$$

are measurable. We assume that this SDE is strongly and weakly well-posed for square integrable initial values. It is in particular the case if b is continuous on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$ and there exists a constant $K > 0$ such that

$$\begin{aligned} &\langle b(x,\mu) - b(y,\nu), x-y\rangle^+ + \|\sigma(x,\mu) - \sigma(y,\nu)\|^2 \le K\{|x-y|^2 + \mathbb{W}_2(\mu,\nu)^2\}, \\ &|b(0,\mu)| \le K\Big(1 + \sqrt{\mu(|\cdot|^2)}\Big), \quad x, y \in \mathbb{R}^d, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d), \end{aligned} \tag{5.9}$$

see for instance [51]. See also [24, 58] and references therein for the well-posedness of DDSDEs with singular coefficients. For any $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, let $P_t^*\mu = \mathscr{L}_{X_t}$ for the solution X_t with initial distribution $\mathscr{L}_{X_0} = \mu$. Let

$$P_t f(\mu) = \mathbb{E}[f(X_t)] = \int_{\mathbb{R}^d} f \mathrm{d}P_t^*\mu, \quad t \ge 0, f \in \mathscr{B}_b(\mathbb{R}^d).$$

We have the following equivalence on the exponential convergence of $P_t^*\mu$ in Ent and $\mathbb{W}_2$.

Theorem 5.1 ([41]). *Assume that P_t^* has a unique invariant probability measure $\mu_\infty \in \mathscr{P}_2(\mathbb{R}^d)$ such that for some constants $t_0, c_0, C > 0$ we have the log-Harnack inequality*

$$P_{t_0}(\log f)(\nu) \le \log P_{t_0} f(\mu) + c_0 \mathbb{W}_2(\mu,\nu)^2, \quad \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d), f \in \mathscr{B}_b^+(\mathbb{R}^d) \tag{5.10}$$

and the Talagrand inequality

$$\mathbb{W}_2(\mu, \mu_\infty)^2 \le C\mathrm{Ent}(\mu|\mu_\infty), \quad \mu \in \mathscr{P}_2(\mathbb{R}^d). \tag{5.11}$$

(1) *If there exist constants $c_1, \lambda, t_1 \ge 0$ such that*

$$\mathbb{W}_2(P_t^*\mu, \mu_\infty)^2 \le c_1 \mathrm{e}^{-\lambda t} \mathbb{W}_2(\mu, \mu_\infty)^2, \quad t \ge t_1, \mu \in \mathscr{P}_2(\mathbb{R}^d), \tag{5.12}$$

then

$$\begin{aligned} &\max\big\{c_0^{-1}\mathrm{Ent}(P_t^*\mu|\mu_\infty), \mathbb{W}_2(P_t^*\mu, \mu_\infty)^2\big\} \\ &\le c_1 \mathrm{e}^{-\lambda(t-t_0)} \min\big\{\mathbb{W}_2(\mu,\mu_\infty)^2, C\mathrm{Ent}(\mu|\mu_\infty)\big\}, \quad t \ge t_0 + t_1, \mu \in \mathscr{P}_2(\mathbb{R}^d). \end{aligned} \tag{5.13}$$

(2) *If for some constants $\lambda, c_2, t_2 > 0$*

$$\mathrm{Ent}(P_t^*\mu|\mu_\infty) \le c_2 \mathrm{e}^{-\lambda t} \mathrm{Ent}(\mu|\mu_\infty), \quad t \ge t_2, \mu \in \mathscr{P}_2(\mathbb{R}^d), \tag{5.14}$$

then

$$\begin{aligned} &\max\big\{\mathrm{Ent}(P_t^*\mu, \mu_\infty), C^{-1}\mathbb{W}_2(P_t^*\mu, \mu_\infty)^2\big\} \\ &\le c_2 \mathrm{e}^{-\lambda(t-t_0)} \min\big\{c_0\mathbb{W}_2(\mu,\mu_\infty)^2, \mathrm{Ent}(\mu|\mu_\infty)\big\}, \quad t \ge t_0 + t_2, \mu \in \mathscr{P}_2(\mathbb{R}^d). \end{aligned} \tag{5.15}$$

When $\sigma\sigma^*$ is invertible and does not depend on the distribution, the log-Harnack inequality (5.10) has been established in [51]. The Talagrand inequality was first found in [46] for μ_∞ being the Gaussian measure, and extended in [7] to μ_∞ satisfying the log-Sobolev inequality

$$\mu_\infty(f^2 \log f^2) \le C\mu_\infty(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^d), \mu_\infty(f^2) = 1, \tag{5.16}$$

see [35] for an earlier result under a curvature condition, and see [48] for further extensions.

To illustrate this result, we consider the granular media type equation for probability density functions $(\rho_t)_{t\ge 0}$ on $\mathbb{R}^d$:

$$\partial_t \rho_t = \operatorname{div}\big\{a\nabla\rho_t + \rho_t a\nabla(V + W \circledast \rho_t)\big\}, \tag{5.17}$$

where $W \circledast \rho_t$ is in (5.4), and the functions

$$a : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d, \quad V : \mathbb{R}^d \to \mathbb{R}, \quad W : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

satisfy the following assumptions.

(H_5^1) $a := (a_{ij})_{1\le i,j\le d} \in C_b^2(\mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d)$, and $a \ge \lambda_a I_{d\times d}$ for some constant $\lambda_a > 0$.

(H_5^2) $V \in C^2(\mathbb{R}^d), W \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$ with $W(x,y) = W(y,x)$, and there exist constants $\kappa_0 \in \mathbb{R}$ and $\kappa_1, \kappa_2, \kappa_0' > 0$ such that

$$\operatorname{Hess}_V \ge \kappa_0 I_{d\times d}, \quad \kappa_0' I_{2d\times 2d} \ge \operatorname{Hess}_W \ge \kappa_0 I_{2d\times 2d}, \tag{5.18}$$

$$\langle x, \nabla V(x)\rangle \ge \kappa_1 |x|^2 - \kappa_2, \quad x \in \mathbb{R}^d. \tag{5.19}$$

Moreover, for any $\lambda > 0$,

$$\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{e}^{-V(x)-V(y)-\lambda W(x,y)}\mathrm{d}x\mathrm{d}y < \infty. \tag{5.20}$$

(H_5^3) There exists a function $b_0 \in L^1_{loc}([0,\infty))$ with

$$r_0 := \frac{\|\operatorname{Hess}_W\|_\infty}{4} \int_0^\infty \mathrm{e}^{\frac{1}{4}\int_0^t b_0(s)\mathrm{d}s}\mathrm{d}t < 1$$

such that for any $x, y, z \in \mathbb{R}^d$,

$$\big\langle y - x, \nabla V(x) - \nabla V(y) + \nabla W(\cdot,z)(x) - \nabla W(\cdot,z)(y)\big\rangle \le |x-y| b_0(|x-y|).$$

For any $N \ge 2$, consider the Hamiltonian for the system of N particles:

$$H_N(x_1, \cdots, x_N) = \sum_{i=1}^N V(x_i) + \frac{1}{N-1}\sum_{1\le i<j\le N} W(x_i, x_j),$$

and the corresponding finite-dimensional Gibbs measure

$$\mu^{(N)}(\mathrm{d}x_1, \cdots, \mathrm{d}x_N) = \frac{1}{Z_N} \mathrm{e}^{-H_N(x_1,\cdots,x_N)} \mathrm{d}x_1 \cdots \mathrm{d}x_N,$$

where $Z_N := \int_{\mathbb{R}^{dN}} \mathrm{e}^{-H_N(x)}\mathrm{d}x < \infty$ due to (5.20) in (H_2). For any $1 \le i \le N$, the conditional marginal of $\mu^{(N)}$ given $z \in \mathbb{R}^{d(N-1)}$ is given by

$$\begin{aligned} &\mu_z^{(N)}(\mathrm{d}x) := \frac{1}{Z_N(z)} \mathrm{e}^{-H_N(x|z)}\mathrm{d}x, \quad Z_N(z) := \int_{\mathbb{R}^d} \mathrm{e}^{-H_N(x|z)}\mathrm{d}x, \\ &H_N(x|z) := V(x) - \log \int_{\mathbb{R}^{d(N-1)}} \mathrm{e}^{-\sum_{i=1}^{N-1}\{V(z_i) + \frac{1}{N-1} W(x,z_i)\}} \mathrm{d}z_1 \cdots \mathrm{d}z_{N-1}. \end{aligned}$$

We have the following result.

Theorem 5.2 ([41])**.** *Assume (H_5^1)-(H_5^3). If there is a constant $\beta > 0$ such that the uniform log-Sobolev inequality*

$$(5.21) \qquad \mu_z^{(N)}(f^2 \log f^2) \le \frac{1}{\beta} \mu_z^{(N)}(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^d), \mu_z^{(N)}(f^2) = 1, N \ge 2, z \in \mathbb{R}^{d(N-1)}$$

holds, then there exists a unique $\mu_\infty \in \mathscr{P}_2(\mathbb{R}^d)$ and a constant $c > 0$ such that

$$(5.22) \quad \mathbb{W}_2(\mu_t, \mu_\infty)^2 + \mathrm{Ent}(\mu_t|\mu_\infty) \le c \mathrm{e}^{-\lambda_a \beta (1-r_0)^2 t} \min\left\{\mathbb{W}_2(\mu_0, \mu_\infty)^2 + \mathrm{Ent}(\mu_0|\mu_\infty)\right\}, \quad t \ge 1$$

holds for any probability density functions $(\rho_t)_{t\ge0}$ solving (5.17)*, where $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x, t \ge 0$.*

This result allows V and W to be non-convex. For instance, let $V = V_1 + V_2 \in C^2(\mathbb{R}^d)$ such that $\|V_1\|_\infty \wedge \|\nabla V_1\|_\infty < \infty$, $\mathrm{Hess}_{V_2} \ge \lambda I_{d\times d}$ for some $\lambda > 0$, and $W \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$ with $\|W\|_\infty \wedge \|\nabla W\|_\infty < \infty$. Then the uniform log-Sobolev inequality (5.21) holds for some constant $\beta > 0$.

5.2 The non-degenerate case

In this part, we make the following assumptions:

(H_5^4) b is continuous on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$ and there exists a constant $K > 0$ such that (5.9) holds.

(H_5^5) $\sigma\sigma^*$ is invertible with $\lambda := \|(\sigma\sigma^*)^{-1}\|_\infty < \infty$, and there exist constants $K_2 > K_1 \ge 0$ such that for any $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$,

$$\|\sigma(x) - \sigma(y)\|_{HS}^2 + 2\langle b(x,\mu) - b(y,\nu), x - y\rangle \le K_1 \mathbb{W}_2(\mu,\nu)^2 - K_2|x-y|^2.$$

According to Theorem 3.3, if (H_3^1) holds, then for any initial value $X_0 \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$, (5.7) has a unique solution which satisfies

$$\mathbb{E}\Big[\sup_{t\in[0,T]} |X_t|^2\Big] < \infty, \quad T \in (0,\infty).$$

Let $P_t^*\mu = \mathscr{L}_{X_t}$ for the solution X_t with $\mathscr{L}_{X_0} = \mu$. We have the following result.

Theorem 5.3 ([41]). *Assume* (H_5^4) *and* (H_5^5). *Then* P_t^* *has a unique invariant probability measure* μ_∞ *such that*

$$\max\left\{\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2, \mathrm{Ent}(P_t^*\mu|\mu_\infty)\right\} \le \frac{c_1}{t\wedge 1}\mathrm{e}^{-(K_2-K_1)t}\mathbb{W}_2(\mu,\mu_\infty)^2, \quad t>0, \mu\in\mathscr{P}_2(\mathbb{R}^d) \tag{5.23}$$

holds for some constant $c_1 > 0$. *If moreover* $\sigma \in C_b^2(\mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m)$, *then there exists a constant* $c_2 > 0$ *such that for any* $\mu \in \mathscr{P}_2(\mathbb{R}^d), t \ge 1$,

$$\max\left\{\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2, \mathrm{Ent}(P_t^*\mu|\mu_\infty)\right\} \le c_2\mathrm{e}^{-(K_2-K_1)t}\min\left\{\mathbb{W}_2(\mu,\mu_\infty)^2, \mathrm{Ent}(\mu|\mu_\infty)\right\}. \tag{5.24}$$

To illustrate this result, we consider the granular media equation (5.3), for which we take

$$\sigma = \sqrt{2}I_{d\times d}, \qquad b(x,\mu) = -\nabla\left\{V + W \circledast \mu\right\}(x), \quad (x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d). \tag{5.25}$$

The following example is not included by Theorem 5.2 since the function W may be non-symmetric.

Example 5.1 (Granular media equation). Consider (5.3) with $V \in C^2(\mathbb{R}^d)$ and $W \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$ satisfying

$$\mathrm{Hess}_V \ge \lambda I_{d\times d}, \quad \mathrm{Hess}_W \ge \delta_1 I_{2d\times 2d}, \quad \|\mathrm{Hess}_W\| \le \delta_2 \tag{5.26}$$

for some constants $\lambda_1, \delta_2 > 0$ and $\delta_1 \in \mathbb{R}$. If $\lambda + \delta_1 - \delta_2 > 0$, then there exists a unique $\mu_\infty \in \mathscr{P}_2(\mathbb{R}^d)$ and a constant $c > 0$ such that for any probability density functions $(\rho_t)_{t\ge0}$ solving (5.3), $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ satisfies

$$\begin{aligned}&\max\left\{\mathbb{W}_2(\mu_t,\mu_\infty)^2, \mathrm{Ent}(\mu_t|\mu_\infty)\right\}\\ &\le c\mathrm{e}^{-(\lambda+\delta_1-\delta_2)t}\min\left\{\mathbb{W}_2(\mu_0,\mu_\infty)^2, \mathrm{Ent}(\mu_0|\mu_\infty)\right\}, \quad t\ge 1.\end{aligned} \tag{5.27}$$

Proof. Let σ and b be in (5.25). Then (5.26) implies (H_3^1) and

$$\langle b(x,\mu) - b(y,\nu), x-y\rangle \le -(\lambda_1+\delta_1)|x-y|^2 + \delta_2|x-y|\mathbb{W}_1(\mu,\nu),$$

where we have used the formula

$$\mathbb{W}_1(\mu,\nu) = \sup\{\mu(f)-\nu(f): \ \|\nabla f\|_\infty \le 1\}.$$

So, by taking $\alpha = \frac{\delta_2}{2}$ and noting that $\mathbb{W}_1 \le \mathbb{W}_2$, we obtain

$$\begin{aligned}\langle b(x,\mu) - b(y,\nu), x-y\rangle &\le -\big(\lambda+\delta_1-\alpha\big)|x-y|^2 + \frac{\delta_2^2}{4\alpha}\mathbb{W}_1(\mu,\nu)^2\\ &\le -\Big(\lambda+\delta_1-\frac{\delta_2}{2}\Big)|x-y|^2 + \frac{\delta_2}{2}\mathbb{W}_2(\mu,\nu)^2, \quad x,y\in\mathbb{R}^d, \mu,\nu\in\mathscr{P}_2(\mathbb{R}^d).\end{aligned}$$

Therefore, if (5.26) holds for $\lambda + \delta_1 - \delta_2 > 0$, Theorem 5.3 implies that P_t^* has a unique invariant probability measure $\mu_\infty \in \mathscr{P}_2(\mathbb{R}^d)$, such that (5.27) holds for $\mu_0 \in \mathscr{P}_2(\mathbb{R}^d)$. When $\mu_0 \notin \mathscr{P}_2(\mathbb{R}^d)$, we have $\mathbb{W}_2(\mu_0,\mu_\infty)^2 = \infty$ since $\mu_\infty \in \mathscr{P}_2(\mathbb{R}^d)$. Combining this with the Talagrand inequality

$$\mathbb{W}_2(\mu_0,\mu_\infty)^2 \le C\mathrm{Ent}(\mu_0|\mu_\infty)$$

for some constant $C > 0$, see the proof of Theorem 5.3, we have $\mathrm{Ent}(\mu_0|\mu_\infty) = \infty$ for $\mu_0 \notin \mathscr{P}_2(\mathbb{R}^d)$, so that (5.27) holds for all $\mu_0 \in \mathscr{P}(\mathbb{R}^d)$. □

5.3 The degenerate case

When $\mathbb{R}^k$ with some $k \in \mathbb{N}$ is considered, to emphasize the space we use $\mathscr{P}(\mathbb{R}^k)$ ($\mathscr{P}_2(\mathbb{R}^k)$) to denote the class of probability measures (with finite second moment) on $\mathbb{R}^k$. Consider the following McKean-Vlasov stochastic Hamiltonian system for $(X_t, Y_t) \in \mathbb{R}^{d_1+d_2} := \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$:

$$\begin{cases} \mathrm{d}X_t = BY_t \mathrm{d}t, \\ \mathrm{d}Y_t = \sqrt{2}\mathrm{d}W_t - \Big\{B^*\nabla V(\cdot, \mathscr{L}_{(X_t,Y_t)})(X_t) + \beta B^*(BB^*)^{-1}X_t + Y_t\Big\}\mathrm{d}t, \end{cases} \tag{5.28}$$

where $\beta > 0$ is a constant, B is a $d_1 \times d_2$-matrix such that BB^* is invertible, and

$$V : \mathbb{R}^{d_1} \times \mathscr{P}_2(\mathbb{R}^{d_1+d_2}) \to \mathbb{R}^{d_2}$$

is measurable. Let

$$\begin{aligned} &\psi_B((x,y),(\bar{x},\bar{y})) := \sqrt{|x-\bar{x}|^2 + |B(y-\bar{y})|^2}, \quad (x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{d_1+d_2}, \\ &\mathbb{W}_2^{\psi_B}(\mu,\nu) := \inf_{\pi \in \mathscr{C}(\mu,\nu)} \bigg\{ \int_{\mathbb{R}^{d_1+d_2}\times\mathbb{R}^{d_1+d_2}} {\psi_B}^2 \mathrm{d}\pi \bigg\}^{\frac{1}{2}}, \quad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}). \end{aligned}$$

We assume

(H_5^6) $V(x,\mu)$ is differentiable in x such that $\nabla V(\cdot,\mu)(x)$ is Lipschitz continuous in $(x,\mu) \in \mathbb{R}^{d_1} \times \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$. Moreover, there exist constants $\theta_1, \theta_2 \in \mathbb{R}$ with

$$\theta_1 + \theta_2 < \beta, \tag{5.29}$$

such that for any $(x,y),(x',y') \in \mathbb{R}^{d_1+d_2}$ and $\mu,\mu' \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$,

$$\begin{aligned} &\big\langle BB^*\{\nabla V(\cdot,\mu)(x) - \nabla V(\cdot,\mu')(x')\}, x - x' + (1+\beta)B(y-y')\big\rangle \\ &\geq -\theta_1\psi_B((x,y),(x',y'))^2 - \theta_2\mathbb{W}_2^{\psi_B}(\mu,\mu')^2. \end{aligned} \tag{5.30}$$

Obviously, (H_5^6) implies (H_3^1) for $d = m = d_1 + d_2$, $\sigma = \mathrm{diag}\{0, \sqrt{2}I_{d_2\times d_2}\}$, and

$$b((x,y),\mu) = \big(By, -B^*\nabla V(\cdot,\mu)(x) - \beta B^*(BB^*)^{-1}x - y\big).$$

So, according to [51], (5.28) is well-posed for any initial value in $L^2(\Omega \to \mathbb{R}^{d_1+d_2}, \mathscr{F}_0, \mathbb{P})$. Let $P_t^*\mu = \mathscr{L}_{(X_t,Y_t)}$ for the solution with initial distribution $\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$.

Theorem 5.4 ([41])**.** *Assume (H_5^6). Then P_t^* has a unique invariant probability measure μ_∞ such that for any $t > 0$ and $\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$,*

$$\max\big\{\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2, \mathrm{Ent}(P_t^*\mu|\mu_\infty)\big\} \leq \frac{c\mathrm{e}^{-2\kappa t}}{(1\wedge t)^3}\min\big\{\mathrm{Ent}(\mu|\mu_\infty), \mathbb{W}_2(\mu,\mu_\infty)^2\big\} \tag{5.31}$$

holds for some constant $c > 0$ and

$$\kappa := \frac{2(\beta - \theta_1 - \theta_2)}{2 + 2\beta + \beta^2 + \sqrt{\beta^4 + 4}} > 0. \tag{5.32}$$

Example 5.2 (Degenerate granular media equation). Let $m \in \mathbb{N}$ and $W \in C^2(\mathbb{R}^m \times \mathbb{R}^{2m})$. Consider the following PDE for probability density functions $(\rho_t)_{t\geq 0}$ on $\mathbb{R}^{2m}$:

$$\partial_t \rho_t(x,y) = \Delta_y \rho_t(x,y) - \langle \nabla_x \rho_t(x,y), y\rangle + \langle \nabla_y \rho_t(x,y), \nabla_x (W \circledast \rho_t)(x) + \beta x + y\rangle, \tag{5.33}$$

where $\beta > 0$ is a constant, $\Delta_y, \nabla_x, \nabla_y$ stand for the Laplacian in y and the gradient operators in x, y respectively, and

$$(W \circledast \rho_t)(x) := \int_{\mathbb{R}^{2m}} W(x,z)\rho_t(z)\mathrm{d}z, \quad x \in \mathbb{R}^m.$$

If there exists a constant $\theta \in \big(0, \frac{2\beta}{1+3\sqrt{2+2\beta+\beta^2}}\big)$ such that

$$|\nabla W(\cdot, z)(x) - \nabla W(\cdot,\bar z)(\bar x)| \leq \theta\big(|x-\bar x| + |z - \bar z|\big), \quad x, \bar x \in \mathbb{R}^m, z, \bar z \in \mathbb{R}^{2m}, \tag{5.34}$$

then there exists a unique probability measure $\mu_\infty \in \mathscr{P}_2(\mathbb{R}^{2m})$ and a constant $c > 0$ such that for any probability density functions $(\rho_t)_{t\geq 0}$ solving (5.33), $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ satisfies

$$\max\big\{\mathbb{W}_2(\mu_t, \mu_\infty)^2, \mathrm{Ent}(\mu_t|\mu_\infty)\big\} \leq c\mathrm{e}^{-\kappa t} \min\big\{\mathbb{W}_2(\mu_0,\mu_\infty)^2, \mathrm{Ent}(\mu_0|\mu_\infty)\big\}, \quad t \geq 1 \tag{5.35}$$

holds for $\kappa = \frac{2\beta - \theta\big(1+3\sqrt{2+2\beta+\beta^2}\big)}{2+2\beta+\beta^2+\sqrt{\beta^4+4}} > 0$.

Proof. Let $d_1 = d_2 = m$ and (X_t, Y_t) solve (5.28) for

$$B := I_{m\times m}, \quad V(x,\mu) := \int_{\mathbb{R}^{2m}} W(x,z)\mu(\mathrm{d}z). \tag{5.36}$$

Let $\rho_t(z) = \frac{\mathscr{L}_{(X_t,Y_t)}(\mathrm{d}z)}{\mathrm{d}z}$. By Itô's formula and integration by parts formula, for any $f \in C_0^2(\mathbb{R}^{2m})$ we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}^{2m}} (\rho_t f)(z)\mathrm{d}z = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(X_t,Y_t)]\\
&= \int_{\mathbb{R}^{2m}} \rho_t(x,y)\big\{\Delta_y f(x,y) + \langle \nabla_x f(x,y), y\rangle - \langle \nabla_y f(x,y), \nabla_x V(x, \rho_t(z)\mathrm{d}z) + \beta x + y\rangle\big\}\mathrm{d}x\mathrm{d}y\\
&= \int_{\mathbb{R}^{2m}} f(x,y)\big\{\Delta_y \rho_t(x,y) - \langle \nabla_x \rho_t(x,y), y\rangle + \langle \nabla_y \rho_t(x,y), \nabla_x \mu_t(W(x,\cdot)) + \beta x + y\rangle\big\}\mathrm{d}x\mathrm{d}y.
\end{aligned}$$

Then ρ_t solves (5.33). On the other hand, by the uniqueness of of (5.28) and (5.33), for any solution ρ_t to (5.33) with $\mu_0(\mathrm{d}z) := \rho_0(z)\mathrm{d}z \in \mathscr{P}_2(\mathbb{R}^{2m})$ for $d = 2m$, $\rho_t(z)\mathrm{d}z = \mathscr{L}_{(X_t,Y_t)}(\mathrm{d}z)$ for the solution to (5.28) with initial distribution μ_0. So, as explained in the proof of Example 2.1, by Theorem 5.4 we only need to verify (H_5^6) for B, V in (5.36) and

$$\theta_1 = \theta\Big(\frac{1}{2} + \sqrt{2+2\beta+\beta^2}\Big), \quad \theta_2 = \frac{\theta}{2}\sqrt{2+2\beta+\beta^2}, \tag{5.37}$$

so that the desired assertion holds for

$$\kappa := \frac{2(\beta-\theta_1-\theta_2)}{2+2\beta+\beta^2+\sqrt{\beta^4+4}} = \frac{2\beta-\theta(1+3\sqrt{2+2\beta+\beta^2})}{2+2\beta+\beta^2+\sqrt{\beta^4+4}}.$$

By (5.34) and $V(x,\mu):=\mu(W(x,\cdot))$, for any constants $\alpha_1,\alpha_2,\alpha_3>0$ we have

$$\begin{aligned} I &:= \big\langle \nabla V(\cdot,\mu)(x) - \nabla V(\cdot,\bar\mu)(\bar x), x-\bar x+(1+\beta)(y-\bar y)\big\rangle \\ &= \int_{\mathbb{R}^{2m}} \big\langle \nabla W(\cdot,z)(x)-\nabla W(\cdot,z)(\bar x), x-\bar x+(1+\beta)(y-\bar y)\big\rangle \mu(\mathrm{d}z) \\ &\quad + \big\langle \mu(\nabla_{\bar x} W(\bar x,\cdot)) - \bar\mu(\nabla_{\bar x}W(\bar x,\cdot)), x-\bar x+(1+\beta)(y-\bar y)\big\rangle \\ &\ge -\theta\big\{|x-\bar x|+\mathbb{W}_1(\mu,\bar\mu)\big\}\cdot\big(|x-\bar x|+(1+\beta)|y-\bar y|\big) \\ &\ge -\theta(\alpha_2+\alpha_3)\mathbb{W}_2(\mu,\bar\mu)^2 - \theta\Big\{\Big(1+\alpha_1+\frac{1}{4\alpha_2}\Big)|x-\bar x|^2+(1+\beta)^2\Big(\frac{1}{4\alpha_1}+\frac{1}{4\alpha_3}\Big)|y-\bar y|^2\Big\}. \end{aligned}$$

Take

$$\alpha_1 = \frac{\sqrt{2+2\beta+\beta^2}-1}{2},\quad \alpha_2=\frac{1}{2\sqrt{2+2\beta+\beta^2}},\quad \alpha_3=\frac{(1+\beta)^2}{2\sqrt{2+2\beta+\beta^2}}.$$

We have

$$\begin{aligned} 1+\alpha_1+\frac{1}{4\alpha_2} &= \frac12+\sqrt{2+2\beta+\beta^2}, \\ (1+\beta)^2\Big(\frac{1}{4\alpha_1}+\frac{1}{4\alpha_3}\Big) &= \frac12+\sqrt{2+2\beta+\beta^2}, \\ \alpha_2+\alpha_3 &= \frac12\sqrt{2+2\beta+\beta^2}. \end{aligned}$$

Therefore,

$$I \ge -\frac{\theta}{2}\sqrt{2+2\beta+\beta^2}\,\mathbb{W}_2(\mu,\bar\mu)^2 - \theta\Big(\frac12+\sqrt{2+2\beta+\beta^2}\Big)|(x,y)-(\bar x,\bar y)|^2,$$

i.e. (H_5^6) holds for B and V in (5.36) where $B=I_{m\times m}$ implies that ψ_B is the Euclidean distance on $\mathbb{R}^{2m}$, and for θ_1,θ_2 in (5.37). □

6 Donsker-Varadhan large deviations

The LDP (large deviation principle) is a fundamental tool characterizing the asymptotic behaviour of probability measures $\{\mu_\varepsilon\}_{\varepsilon>0}$ on a topological space E, see [13] and references within. Recall that μ_ε for small $\varepsilon>0$ is said to satisfy the LDP with speed $\lambda(\varepsilon)\to+\infty$ (as $\varepsilon\to0$) and rate function $I:E\to[0,+\infty]$, if I has compact level sets (i.e. $\{I\le r\}$ is compact for $r\in\mathbb{R}^+$), and for any Borel subset A of E,

$$-\inf_{A^o} I \le \liminf_{\varepsilon\to0}\frac{1}{\lambda(\varepsilon)}\log\mu_\varepsilon(A) \le \limsup_{\varepsilon\to0}\frac{1}{\lambda(\varepsilon)}\log\mu_\varepsilon(A) \le -\inf_{\bar A} I,$$

where A^o and $\bar{A}$ stand for the interior and the closure of A in E respectively.

In this part, we consider the Donsker-Varadhan type long time LDP [12] for $\mu_\varepsilon := \mathscr{L}_{L_{\varepsilon^{-1}}}$, where

$$L_t := \frac{1}{t}\int_0^t \delta_{X(s)}\mathrm{d}s, \quad t > 0$$

is the empirical measure for a path-distribution dependent SPDE.

Let $(\mathbb{H}, \langle\cdot,\cdot\rangle, |\cdot|)$ be a separable Hilbert space. For a fixed constant $r_0 > 0$, a path $\xi \in \mathscr{C} := C([-r_0, 0]; \mathbb{H})$ stands for a sample of the history with time length r_0. Recall that $\mathscr{C}$ is a Banach space with the uniform norm

$$\|\xi\|_\infty := \sup_{\theta\in[-r_0,0]} |\xi(\theta)|, \quad \xi \in \mathscr{C}.$$

For any map $\xi(\cdot) : [-r_0, \infty) \to \mathbb{H}$ and any time $t \geq 0$, its segment $\xi_\cdot : [0, \infty) \to \mathscr{C}$ is defined by

$$\xi_t(\theta) := \xi(t + \theta), \quad \theta \in [-r_0, 0], t \geq 0.$$

Let $\mathscr{P}(\mathscr{C})$ denote the space of all probability measures on $\mathscr{C}$ equipped with the weak topology, and let $\mathscr{L}_\eta$ stand for the distribution of a random variable η. Consider the following path-distribution dependent SPDE on $\mathbb{H}$:

$$\mathrm{d}X(t) = \{AX(t) + b(X_t, \mathscr{L}_{X_t})\}\mathrm{d}t + \sigma(\mathscr{L}_{X_t})\mathrm{d}W(t), \quad t \geq 0, \tag{6.1}$$

where

- $(A, \mathscr{D}(A))$ is a negative definite self-adjoint operator on $\mathbb{H}$;
- $W(t)$ is the cylindrical Brownian motion on a separable Hilbert space $\tilde{\mathbb{H}}$; i.e.

  $$W(t) = \sum_{i=1}^{\infty} B_i(t)\tilde{e}_i, \quad t \geq 0$$

  for an orthonormal basis $\{\tilde{e}_i\}_{i\geq1}$ on $\tilde{\mathbb{H}}$ and a sequence of independent one-dimensional Brownian motions $\{B_i\}_{i\geq1}$ on a complete filtration probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq0}, \mathbb{P})$, where $\mathscr{F}_0$ is rich enough such that for any $\pi \in \mathscr{P}(\mathscr{C} \times \mathscr{C})$ there exists a $\mathscr{C} \times \mathscr{C}$-valued random variable ξ on $(\Omega, \mathscr{F}_0, \mathbb{P})$ such that $\mathscr{L}_\xi = \pi$.
- $b : \mathscr{C} \times \mathscr{P}(\mathscr{C}) \to \mathbb{H}, \quad \sigma : \mathscr{P}(\mathscr{C}) \to \mathbb{L}(\tilde{\mathbb{H}}; \mathbb{H})$ are measurable.

Let X_t^ν denote the mild segment solution with initial distribution $\nu \in \mathscr{P}(\mathscr{C})$, which is a continuous adapted process on $\mathscr{C}$. We study the long time LDP for the empirical measure

$$L_t^\nu := \frac{1}{t}\int_0^t \delta_{X_s^\nu}\mathrm{d}s, \quad t > 0.$$

Definition 6.1. Let $\mathscr{P}(\mathscr{C})$ be equipped with the weak topology, let $\mathscr{A} \subset \mathscr{P}(\mathscr{C})$, and let $J : \mathscr{P}(\mathscr{C}) \to [0, \infty]$ have compact level sets, i.e. $\{J \leq r\}$ is compact in $\mathscr{P}(\mathscr{C})$ for any $r > 0$.

(1) $\{L_t^\nu\}_{\nu\in\mathscr{A}}$ is said to satisfy the upper bound uniform LDP with rate function J, denoted by $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_u(J)$, if for any closed $A \subset \mathscr{P}(\mathscr{C})$,

$$\limsup_{t\to\infty} \frac{1}{t} \sup_{\nu\in\mathscr{A}} \log \mathbb{P}(L_t^\nu \in A) \le -\inf_A J.$$

(2) $\{L_t^\nu\}_{\nu\in\mathscr{A}}$ is said to satisfy the lower bound uniform LDP with rate function J, denoted by $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_l(J)$, if for any open $A \subset \mathscr{P}(\mathscr{C})$,

$$\liminf_{t\to\infty} \frac{1}{t} \inf_{\nu\in\mathscr{A}} \log \mathbb{P}(L_t^\nu \in A) \ge -\inf_A J.$$

(3) $\{L_t^\nu\}_{\nu\in\mathscr{A}}$ is said to satisfy the uniform LDP with rate function J, denoted by $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP(J)$, if $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_u(J)$ and $\{L_t^\nu\}_{\nu\in\mathscr{A}} \in LDP_l(J)$.

We investigate the long time LDP for (6.1) in the following three situations respectively:

1) $r_0 = 0$ and $\mathbb{H}$ is finite-dimensional;

2) $r_0 = 0$ and $\mathbb{H}$ is infinite-dimensional;

3) $r_0 > 0$ and σ is constant.

When $r_0 > 0$ and σ is non-constant, the Donsker-Varadhan LDP is still unknown.

To state establish the LDP, we recall the Feller property, the strong Feller property and the irreducibility for a (sub-) Markov operator P. Let $\mathscr{B}_b(\mathscr{C})$ (resp. $C_b(\mathscr{C})$) be the space of bounded measurable (resp. continuous) real functions on $\mathscr{C}$. Let P be a sub-Markov operator on $\mathscr{B}_b(\mathscr{C})$, i.e. it is a positivity-preserving linear operator with $P1 \le 1$. P is called strong Feller if $P\mathscr{B}_b(\mathscr{C}) \subset C_b(\mathscr{C})$, is called Feller if $PC_b(\mathscr{C}) \subset C_b(\mathscr{C})$, and is called μ-irreducible for some $\mu \in \mathscr{P}(\mathscr{C})$ if $\mu(1_A P 1_B) > 0$ holds for any $A, B \in \mathscr{B}(\mathscr{C})$ with $\mu(A)\mu(B) > 0$.

6.1 Distribution dependent SDE on $\mathbb{R}^d$

Let $r_0 = 0$, $\mathbb{H} = \mathbb{R}^d$ and $\tilde{\mathbb{H}} = \mathbb{R}^m$ for some $d, m \in \mathbb{N}$. In this case, we combine the linear term Ax with the drift term $b(x,\mu)$, so that (6.1) reduces to

$$\mathrm{d}X(t) = b(X(t), \mathscr{L}_{X(t)})\mathrm{d}t + \sigma(\mathscr{L}_{X(t)})\mathrm{d}W(t), \tag{6.2}$$

where $b : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^m$ and $W(t)$ is the m-dimensional Brownian motion. We assume

(H_6^1) b is continuous, σ is bounded and continuous such that

$$2\langle b(x,\mu) - b(y,\nu), x-y\rangle + \|\sigma(\mu) - \sigma(\nu)\|_{HS}^2 \le -\kappa_1|x-y|^2 + \kappa_2 \mathbb{W}_2(\mu,\nu)^2$$

holds for some constants $\kappa_1 > \kappa_2 \ge 0$ and all $x, y \in \mathbb{R}^d, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d)$.

Under (H_6^1), for any $X(0) \in L^2(\Omega \to \mathbb{R}^d, \mathscr{F}_0, \mathbb{P})$, the equation (6.2) has a unique solution. We write $P_t^*\mu = \mathscr{L}_{X(t)}$ if $\mathscr{L}_{X(0)} = \mu$. By [51, Theorem 3.1(2)], P_t^* has a unique invariant probability measure $\bar\mu \in \mathscr{P}_2(\mathbb{R}^d)$ such that

$$\mathbb{W}_2(P_t^*\nu, \bar\mu)^2 \le \mathrm{e}^{-(\kappa_1-\kappa_2)t}\mathbb{W}_2(\nu,\bar\mu)^2, \quad t \ge 0, \nu \in \mathscr{P}_2(\mathbb{R}^d). \tag{6.3}$$

Consider the reference SDE

$$\mathrm{d}\bar X(t) = b(\bar X(t), \bar\mu)\mathrm{d}t + \sigma(\bar\mu)\mathrm{d}W(t). \tag{6.4}$$

It is standard that under (H_6^1) the equation (6.4) has a unique solution $\bar X^x(t)$ for any starting point $x \in \mathbb{R}^d$, and $\bar\mu$ is the unique invariant probability measure of the associated Markov semigroup

$$\bar P_t f(x) := \mathbb{E}[f(\bar X^x(t))], \quad t \ge 0, x \in \mathbb{R}^d, f \in \mathscr{B}_b(\mathbb{R}^d).$$

Consequently, $\bar P_t$ uniquely extends to $L^\infty(\bar\mu)$. If $f \in L^\infty(\bar\mu)$ satisfies

$$\bar P_t f = f + \int_0^t \bar P_s g \mathrm{d}s, \quad \bar\mu\text{-a.e.}$$

for some $g \in L^\infty(\bar\mu)$ and all $t \ge 0$, we write $f \in \mathscr{D}(\bar{\mathscr{A}})$ and denote $\bar{\mathscr{A}} f = g$. Obviously, we have $\mathscr{D}(\bar{\mathscr{A}}) \supset C_c^\infty(\mathbb{R}^d) := \{f \in C_b^\infty(\mathbb{R}^d) : \nabla f \text{ has compact support}\}$ and

$$\bar{\mathscr{A}} f(x) = \frac{1}{2}\sum_{i,j=1}^d \{\sigma\sigma^*\}_{ij}(\bar\mu)\partial_i\partial_j f(x) + \sum_{i=1}^d b_i(x,\bar\mu)\partial_i f(x), \quad f \in C_c^\infty(\mathbb{R}^d).$$

The Donsker-Varadhan level 2 entropy function J for the diffusion process generated by $\bar{\mathscr{A}}$ has compact level sets in $\mathscr{P}(\mathbb{R}^d)$ under the τ and weak topologies, and by [40, 3.11], we have

$$J(\nu) = \begin{cases} \sup\left\{ \int_{\mathbb{R}^d} \frac{-\bar{\mathscr{A}} f}{f}\mathrm{d}\nu : \ 1 \le f \in \mathscr{D}(\bar{\mathscr{A}})\right\}, & \text{if } \nu \ll \mu, \\ \infty, & \text{otherwise.} \end{cases}$$

Theorem 6.1 ([40])**.** *Assume (H_6^1). For any $r, R > 0$, let $\mathscr{B}_{r,R} = \left\{\nu \in \mathscr{P}(\mathbb{R}^d) : \nu(\mathrm{e}^{|\cdot|^r}) \le R\right\}$.*

(1) *We have $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}} \in LDP_u(J)$ for all $r, R > 0$. If $\bar P_t$ is strong Feller and $\bar\mu$-irreducible for some $t > 0$, then $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}} \in LDP(J)$ for all $r, R > 0$.*

(2) *If there exist constants $\varepsilon, c_1, c_2 > 0$ such that*

$$\langle x, b(x,\nu)\rangle \le c_1 - c_2|x|^{2+\varepsilon}, \quad x \in \mathbb{R}^d, \nu \in \mathscr{P}_2(\mathbb{R}^d), \tag{6.5}$$

then $\{L_t^\nu\}_{\nu\in\mathscr{P}_2(\mathbb{R}^d)} \in LDP_u(J)$. If moreover $\bar P_t$ is strong Feller and $\bar\mu$-irreducible for some $t > 0$, then $\{L_t^\nu\}_{\nu\in\mathscr{P}_2(\mathbb{R}^d)} \in LDP(J)$.

To apply this result, we first recall some facts on the strong Feller property and the irreducibility of diffusion semigroups.

Remark 6.2. *(1) Let $\bar{P}_t$ be the (sub-)Markov semigroup generated by the second order differential operator*

$$\bar{\mathscr{A}} := \sum_{i=1}^{m} U_i^2 + U_0,$$

where $\{U_i\}_{i=1}^m$ are C^1-vector fields and U_0 is a continuous vector field. According to [31, Theorem 5.1], if $\{U_i : 1 \le i \le m\}$ together with their Lie brackets with U_0 span $\mathbb{R}^d$ at any point (i.e. the Hörmander condition holds), then the Harnack inequality

$$P_t f(x) \le \psi(t,s,x,y) P_{t+s} f(y), \quad t,s>0, x,y \in \mathbb{R}^d, f \in \mathscr{B}^+(\mathbb{R}^d)$$

for some map $\psi : (0,\infty)^2 \times (\mathbb{R}^d)^2 \to (0,\infty)$. Consequently, if moreover $\bar{P}_t$ has an invariant probability measure $\bar{\mu}$, then $\bar{P}_t$ is $\bar{\mu}$-irreducible for any $t>0$. Finally, if $\{U_i\}_{0\le i\le m}$ are smooth with bounded derivatives of all orders, then the above Hörmander condition implies that $\bar{P}_t$ has smooth heat kernel with respect to the Lebesgue measure, in particular it is strong Feller for any $t>0$.

(2) Let $\bar{P}_t$ be the Markov semigroup generated by

$$\bar{\mathscr{A}} := \sum_{i,j=1}^{d} \bar{a}_{ij} \partial_i \partial_j + \sum_{i=1}^{d} \bar{b}_i \partial_j,$$

where $(\bar{a}_{ij}(x))$ is strictly positive definite for any x, $\bar{a}_{ij} \in H^{p,1}_{loc}(\mathrm{d}x)$ and $\bar{b}_i \in L^p_{loc}(\mathrm{d}x)$ for some $p>d$ and all $1\le i,j\le d$. Moreover, let $\bar{\mu}$ be an invariant probability measure of $\bar{P}_t$. Then by [8, Theorem 4.1], $\bar{P}_t$ is strong Feller for all $t>0$. Moreover, as indicated in (1) that [31, Theorem 5.1] ensures the $\bar{\mu}$-irreducibility of $\bar{P}_t$ for $t>0$.

We present below two examples to illustrate this result, where the first is a distribution dependent perturbation of the Ornstein-Ulenbeck process, and the second is the distribution dependent stochastic Hamiltonian system.

Example 6.1. Let $\sigma(\nu) = I + \varepsilon\sigma_0(\nu)$ and $b(x,\nu) = -\frac{1}{2}(\sigma\sigma^*)(\nu)x$, where I is the identity matrix, $\varepsilon>0$ and σ_0 is a bounded Lipschitz continuous map from $\mathscr{P}_2(\mathbb{R}^d)$ to $\mathbb{R}^d \otimes \mathbb{R}^d$. When $\varepsilon>0$ is small enough, assumption (H_1) holds and that $\bar{P}_t$ satisfies conditions in Remark 6.2(2). So, Theorem 6.1(1) implies $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}} \in LDP(J)$ for all $r,R>0$.

If we take $b(x,\nu) = -x - c|x|^\theta x$ for some constants $c,\theta>0$, then when $\varepsilon>0$ is small enough such that (H_1) and (6.5) are satisfied, Theorem 6.1(2) and Remark 6.2(2) imply $\{L_t^\nu\}_{\nu\in\mathscr{P}_2(\mathbb{R}^d)} \in LDP(J)$.

Example 6.2. Let $d=2m$ and consider the following distribution dependent SDE for $X(t) = (X^{(1)}(t), X^{(2)}(t))$ on $\mathbb{R}^m \times \mathbb{R}^m$:

$$\begin{cases} \mathrm{d}X^{(1)}(t) = \{X^{(2)}(t) - \lambda X^{(1)}(t)\}\mathrm{d}t \\ \mathrm{d}X^{(2)}(t) = \{Z(X(t), \mathscr{L}_{X(t)}) - \lambda X^{(2)}(t)\}\mathrm{d}t + \sigma \mathrm{d}W(t), \end{cases},$$

were $\lambda > 0$ is a constant, σ is an invertible $m\times m$-matrix, $W(t)$ is the m-dimensional Brownian motion, and $Z:\mathbb{R}^{2m}\times\mathscr{P}_2(\mathbb{R}^{2m})\to\mathbb{R}^m$ satisfies

$$|Z(x_1,\nu_1)-Z(x_2,\nu_2)|\le \alpha_1|x_1^{(1)}-x_2^{(1)}|+\alpha_2|x_1^{(2)}-x_2^{(2)}|+\alpha_3\mathbb{W}_2(\nu_1,\nu_2)$$

for some constants $\alpha_1,\alpha_2,\alpha_3\ge 0$ and all $x_i=(x_i^{(1)},x_i^{(2)})\in\mathbb{R}^{2m},\nu_i\in\mathscr{P}_2(\mathbb{R}^{2m}),1\le i\le 2$. If

$$4\lambda > \inf_{s>0}\big\{2\alpha_3 s+\alpha_3 s^{-1}+2\alpha_2+\sqrt{4(1+\alpha_1)^2+(2\alpha_2+\alpha_3 s^{-1})^2}\big\}, \tag{6.6}$$

then $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}}\in LDP(J)$ for all $r,R>0$.

Indeed, $b(x,\nu):=(x^{(2)}-\lambda x^{(1)},Z(x,\nu)-\lambda x^{(2)})$ satisfies

$$\begin{aligned}&2\langle b(x_1,\nu_1)-b(x_2,\nu_2),x_1-x_2\rangle\\ &\le -2\lambda|x_1^{(1)}-x_2^{(1)}|^2-2(\lambda-\alpha_2)|x_1^{(2)}-x_2^{(2)}|^2\\ &\quad +2|x_1^{(2)}-x_2^{(2)}|\big\{(1+\alpha_1)|x_1^{(1)}-x_2^{(1)}|+\alpha_3\mathbb{W}_2(\nu_1,\nu_2)\big\}\\ &\le \alpha_3 s\mathbb{W}_2(\nu_1,\nu_2)^2-\{2\lambda-\delta(1+\alpha_1)\}|x_1^{(1)}-x_2^{(1)}|^2\\ &\quad -\{2\lambda-2\alpha_2-\delta^{-1}(1+\alpha_1)-\alpha_3 s^{-1}\}|x_1^{(2)}-x_2^{(2)}|^2,\quad s,\delta>0\end{aligned}$$

for all $x_1,x_2\in\mathbb{R}^{2m}$ and $\nu_1,\nu_2\in\mathscr{P}_2(\mathbb{R}^{2m})$. Taking

$$\delta=\frac{2\alpha_2+\alpha_3 s^{-1}+\sqrt{4(1+\alpha_1)^2+(2\alpha_2+\alpha_3 r^{-1})^2}}{2(1+\alpha_1)}$$

such that $\delta(1+\alpha_1)=2\alpha_2+\delta^{-1}(1+\alpha_1)+\alpha_3 s^{-1}$, we see that (H_6^1) holds for some $\kappa_1>\kappa_2$ provided $2\lambda-\delta(1+\alpha_1)>\alpha_3 s$ for some $s>0$, i.e. (6.6) implies (H_6^1). Moreover, it is easy to see that conditions in Remark 6.2(1) hold, see also [16, 52] for Harnack inequalities and gradeint estimates on stochastic Hamiltonian systems which also imply the strong Feller and $\bar\mu$-irreducibility of $\bar P_t$. Therefore, the claimed assertion follows from Theorem 6.1(1).

6.2 Distribution dependent SPDE

Consider the following distribution-dependent SPDE on a separable Hilbert space $\mathbb{H}$:

$$\mathrm{d}X(t)=\{AX(t)+b(X(t),\mathscr{L}_{X(t)})\}\mathrm{d}t+\sigma(\mathscr{L}_{X(t)})\mathrm{d}W(t), \tag{6.7}$$

where $(A,\mathscr{D}(A))$ is a linear operator on $\mathbb{H}$, $b:\mathbb{H}\times\mathscr{P}_2(\mathbb{H})\to\mathbb{H}$ and $\sigma:\mathscr{P}_2(\mathbb{H})\to\mathbb{L}(\tilde{\mathbb{H}};\mathbb{H})$ are measurable, and $W(t)$ is the cylindrical Brwonian motion on $\tilde{\mathbb{H}}$. We make the following assumption.

(H_6^2) $(-A,\mathscr{D}(A))$ is self-adjoint with discrete spectrum $0<\lambda_1\le\lambda_2\le\cdots$ counting multiplicities such that $\sum_{i=1}^\infty\lambda_i^{\gamma-1}<\infty$ holds for some constant $\gamma\in(0,1)$.

Moreover, b is Lipschitz continuous on $\mathbb{H}\times\mathscr{P}_2(\mathbb{H})$, σ is bounded and there exist constants $\alpha_1,\alpha_2\ge 0$ with $\lambda_1>\alpha_1+\alpha_2$ such that

$$2\langle x-y,b(x,\mu)-b(y,\nu)\rangle+\|\sigma(\mu)-\sigma(\nu)\|_{HS}^2\le 2\alpha_1|x-y|^2+2\alpha_2\mathbb{W}_2(\mu,\nu)^2$$

holds for all $x,y\in\mathbb{H}$ and $\mu,\nu\in\mathscr{P}_2(\mathbb{H})$.

According to Theorem [40, Theorem 3.1], assumption (H_6^2) implies that for any $X(0) \in L^2(\Omega \to \mathbb{H}, \mathscr{F}_0, \mathbb{P})$, the equation (6.7) has a unique mild solution $X(t)$. As before we denote by $X^\nu(t)$ the solution with initial distribution $\nu \in \mathscr{P}_2(\mathbb{H})$, and write $P_t^*\nu = \mathscr{L}_{X^\nu(t)}$. Moreover, by Itô's formula and $\kappa := \lambda_1 - (\alpha_1 + \alpha_2) > 0$, it is easy to see that P_t^* has a unique invariant probability measure $\bar{\mu} \in \mathscr{P}_2(\mathbb{H})$ and

$$\mathbb{W}_2(P_t^*\nu, \bar{\mu}) \le \mathrm{e}^{-\kappa t}\mathbb{W}_2(\nu, \bar{\mu}), \quad t \ge 0. \tag{6.8}$$

Consider the reference SPDE

$$\mathrm{d}\bar{X}(t) = \{A\bar{X}(t) + b(\bar{X}(t), \bar{\mu})\}\mathrm{d}t + \sigma(\bar{\mu})\mathrm{d}W(t),$$

which is again well-posed for any initial value $\bar{X}(0) \in L^2(\Omega \to \mathbb{H}, \mathscr{F}_0, \mathbb{P})$. Let J be the Donsker-Varadhan level 2 entropy function for the Markov process $\bar{X}(t)$, see [40, Section 3]. For any $r, R > 0$ let

$$\mathscr{B}_{r,R} := \big\{\nu \in \mathscr{P}(\mathbb{H}) : \nu(\mathrm{e}^{|\cdot|^r}) \le R\big\}.$$

Theorem 6.3 ([40]). *Assume (H_6^2). If there exist constants $\varepsilon \in (0,1)$ and $c > 0$ such that*

$$\langle (-A)^{\gamma-1}x, b(x,\mu)\rangle \le c + \varepsilon|(-A)^{\frac{\gamma}{2}}x|^2, \quad x \in \mathscr{D}((-A)^{\frac{\gamma}{2}}), \tag{6.9}$$

then $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}} \in LDP_u(J)$ for all $r, R > 0$. If moreover $\bar{P}_t$ is strong Feller and $\bar{\mu}$-irreducible for some $t > 0$, then $\{L_t^\nu\}_{\nu\in\mathscr{B}_{r,R}} \in LDP(J)$ for all $r, R > 0$.

Assumption (H_6^2) is standard to imply the well-posedness of (6.7) and the exponential convergence of P_t^* in $\mathbb{W}_2$. Condition (6.9) is implied by

$$|(-A)^{\frac{\gamma}{2}-1}b(x,\mu)| \le \varepsilon'|(-A)^{\frac{\gamma}{2}}x| + c', \quad x \in \mathscr{D}((-A)^{\frac{\gamma}{2}}) \tag{6.10}$$

for some constants $\varepsilon' \in (0,1)$ and $c' > 0$. In particular, (6.9) holds if $|b(x,\mu)| \le c_1 + c_2|x|$ for some constants $c_1 > 0$ and $c_2 \in (0, \lambda_1)$.

6.3 Path-distribution dependent SPDE with additive noise

Let $\tilde{\mathbb{H}} = \mathbb{H}$ and $\sigma \in \mathbb{L}(\mathbb{H})$. Then (6.1) becomes

$$\mathrm{d}X(t) = \big\{AX(t) + b(X_t, \mathscr{L}_{X_t})\big\}\mathrm{d}t + \sigma\mathrm{d}W(t). \tag{6.11}$$

Below we consider this equation with σ being invertible and non-invertible respectively.

6.3.1 Invertible σ

Since σ is constant, we are able to establish LDP for $b(\xi, \cdot)$ being Lipshcitz continuous in $\mathbb{W}_p$ for some $p \ge 1$ rather than just for $p = 2$ as in the last two results.

(H_6^3) $\sigma \in \mathbb{L}(\mathbb{H})$ is constant and $(A, \mathscr{D}(A))$ satisfies the corresponding condition in (H_2). Moreover, there exist constants $p \ge 1$ and $\alpha_1, \alpha_2 \ge 0$ such that

$$|b(\xi,\mu) - b(\eta,\nu)| \le \alpha_1\|\xi - \eta\|_\infty + \alpha_2\mathbb{W}_p(\mu,\nu), \quad \xi, \eta \in \mathscr{C}, \mu, \nu \in \mathscr{P}_p(\mathscr{C}).$$

Obviously, (H_6^3) implies assumption **(A)** in [40, Theorem 3.1], so that for any $X_0^\nu \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ with $\nu = \mathscr{L}_{X_0^\nu}$, the equation (6.11) has a unique mild segment solution X_t^ν with

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|X_t^\nu\|_\infty^p\Big] < \infty, \quad T > 0.$$

Let $P_t^* \nu = \mathscr{L}_{X_t^\nu}$ for $t \geq 0$ and $\nu \in \mathscr{P}_p(\mathscr{C})$.

When P_t^* has a unique invariant probability measure $\bar\mu \in \mathscr{P}_p(\mathscr{C})$, we consider the reference functional SPDE

$$\text{(6.12)} \qquad \mathrm{d}\bar X(t) = \big\{A\bar X(t) + b(\bar X_t, \bar\mu)\big\}\mathrm{d}t + \sigma \mathrm{d}W(t).$$

By [40, Theorem 3.1], this reference equation is well-posed for any initial value in $L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$. For any $\varepsilon, R > 0$, let

$$\mathscr{I}_{\varepsilon,R} = \big\{\nu \in \mathscr{P}(\mathscr{C}) : \nu(\mathrm{e}^{\varepsilon\|\cdot\|_\infty^2}) \leq R\big\}.$$

Theorem 6.4 ([40]). *Assume (H_6^3). Let $\theta \in [0, \lambda_1]$ such that*

$$\kappa_p := \theta - (\alpha_1 + \alpha_2)\mathrm{e}^{p\theta r_0} = \sup_{r\in[0,\lambda_1]} \big\{r - (\alpha_1 + \alpha_2)\mathrm{e}^{prr_0}\big\}.$$

(1) *For any $\nu_1, \nu_2 \in \mathscr{P}_p(\mathscr{C})$,*

$$\text{(6.13)} \qquad \mathbb{W}_p(P_t^*\nu_1, P_t^*\nu_2)^p \leq \mathrm{e}^{p\theta r_0 - p\kappa_p t}\mathbb{W}_p(\nu_1, \nu_2)^p, \quad t \geq 0.$$

In particular, if $\kappa_p > 0$, then P_t^* has a unique invariant probability measure $\bar\mu \in \mathscr{P}_p(\mathscr{C})$ such that

$$\text{(6.14)} \qquad \mathbb{W}_p(P_t^*\nu, \bar\mu)^p \leq \mathrm{e}^{p\theta r_0 - p\kappa_p t}\mathbb{W}_p(\nu, \bar\mu)^p, \quad t \geq 0, \nu \in \mathscr{P}_p(\mathscr{C}).$$

(2) *Let σ be invertible. If $\kappa_p > 0$ and $\sup_{s\in(0,\lambda_1]}(s - \alpha_1\mathrm{e}^{sr_0}) > 0$, then $\{L_t^\nu\}_{\nu\in\mathscr{I}_{\varepsilon,R}} \in LDP(J)$ for any $\varepsilon, R > 0$, where J is the Donsker-Varadhan level 2 entropy function for the Markov process $\bar X_t$ on $\mathscr{C}$.*

Example 6.3. For a bounded domain $D \subset \mathbb{R}^d$, let $\mathbb{H} = L^2(D; \mathrm{d}x)$ and $A = -(-\Delta)^\alpha$, where Δ is the Dirichlet Laplacian on D and $\alpha > \frac{d}{2}$ is a constant. Let $\sigma = I$ be the identity operator on $\mathbb{H}$, and

$$b(\xi, \mu) = b_0(\mu) + \alpha_1 \int_{-r_0}^0 \xi(r)\Theta(\mathrm{d}r), \quad (\xi, \mu) \in \mathscr{C} \times \mathscr{P}_1(\mathscr{C}),$$

where $\alpha_1 \geq 0$ is a constant, Θ is a signed measure on $[-r_0, 0]$ with total variation 1 (i.e. $|\Theta|([-r_0, 0]) = 1$), and b_0 satisfies

$$|b_0(\mu) - b_0(\nu)| \leq \alpha_2\mathbb{W}_1(\mu, \nu), \quad \mu, \nu \in \mathscr{P}_1(\mathscr{C})$$

for some constant $\alpha_2 \geq 0$. Then (H_6^3) holds for $p = 1$, and as shown in he proof of Example 1.1 in [3] that

$$\lambda_1 \geq \lambda := \frac{(d\pi^2)^\alpha}{R(D)^{2\alpha}},$$

where $R(D)$ is the diameter of D. Therefore, all assertions in Theorem 6.4 hold provided

$$\sup_{r\in(0,\lambda]} \{r - (\alpha_1 + \alpha_2)\mathrm{e}^{rr_0}\} > 0.$$

In particular, under this condition $\{L_t^\nu\}_{\nu\in\mathscr{I}_{\varepsilon,R}} \in LDP(J)$ for any $\varepsilon, R > 1$.

6.3.2 Non-invertible σ

Let $\mathbb{H} = \mathbb{H}_1 \times \mathbb{H}_2$ for two separable Hilbert spaces $\mathbb{H}_1$ and $\mathbb{H}_2$, and consider the following path-distribution dependent SPDE for $X(t) = (X^{(1)}(t), X^{(2)}(t))$ on $\mathbb{H}$:

$$\begin{cases} \mathrm{d}X^{(1)}(t) = \{A_1 X^{(1)}(t) + BX^{(2)}(t)\}\mathrm{d}t, \\ \mathrm{d}X^{(2)}(t) = \{A_2 X^{(2)}(t) + Z(X_t, \mathscr{L}_{X_t})\}\mathrm{d}t + \sigma \mathrm{d}W(t), \end{cases} \tag{6.15}$$

where $(A_i, \mathscr{D}(A_i))$ is a densely defined closed linear operator on $\mathbb{H}_i$ generating a C_0-semigroup e^{tA_i} $(i = 1, 2)$, $B \in \mathbb{L}(\mathbb{H}_2; \mathbb{H}_1)$, $Z : \mathscr{C} \mapsto \mathbb{H}_2$ is measurable, $\sigma \in \mathbb{L}(\mathbb{H}_2)$, and $W(t)$ is the cylindrical Wiener process on $\mathbb{H}_2$. Obviously, (6.15) can be reduced to (6.11) by taking $A = \mathrm{diag}\{A_1, A_2\}$ and using $\mathrm{diag}\{0, \sigma\}$ replacing σ, i.e. (6.15) is a special case of (6.11) with non-invertible σ.

For any $\alpha > 0$ and $p \geq 1$, define

$$\mathbb{W}_{p,\alpha}(\nu_1, \nu_2) := \inf_{\pi\in\mathscr{C}(\nu_1,\nu_2)} \left(\int_{\mathscr{C}\times\mathscr{C}} \left(\alpha\|\xi_1^{(1)} - \xi_2^{(1)}\|_\infty + \|\xi_1^{(2)} - \xi_2^{(2)}\|_\infty\right)^p \pi(\mathrm{d}\xi_1, \mathrm{d}\xi_2) \right)^{\frac{1}{p}}.$$

We assume

(H_6^4) Let $p \geq 1$ and $\alpha > 0$. $(-A_2, \mathscr{D}(A_2))$ is self-adjoint with discrete spectrum $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ counting multiplicities such that $\sum_{i=1}^\infty \lambda_i^{\gamma-1} < \infty$ for some $\gamma \in (0, 1)$. Moreover, $A_1 \leq \delta - \lambda_1$ for some constant $\delta \geq 0$; i.e., $\langle A_1 x, x\rangle \leq (\delta - \lambda_1)|x|^2$ holds for all $x \in \mathscr{D}(A_1)$.

Next, there exist constants $K_1, K_2 > 0$ such that

$$\begin{aligned} &|Z(\xi_1, \nu_1) - Z(\xi_2, \nu_2)| \\ &\leq K_1\|\xi_1^{(1)} - \xi_2^{(1)}\|_\infty + K_2\|\xi_1^{(2)} - \xi_2^{(2)}\|_\infty + K_3\mathbb{W}_{p,\alpha}(\nu_1, \nu_2), \quad (\xi_i, \nu_i) \in \mathscr{C} \times \mathscr{P}_p(\mathscr{C}). \end{aligned}$$

Finally, σ is invertible on $\mathbb{H}_2$, and there exists $A_0 \in \mathbb{L}(\mathbb{H}_1; \mathbb{H}_1)$ such that for any $t > 0$, $B\mathrm{e}^{tA_2} = \mathrm{e}^{tA_1}\mathrm{e}^{tA_0}B$ holds and

$$Q_t := \int_0^t \mathrm{e}^{sA_0}BB^*\mathrm{e}^{sA_0^*}\mathrm{d}s$$

is invertible on $\mathbb{H}_1$.

By [40, Theorem 3.2] for $\mathbb{H}_0 = \mathbb{H}_2$ and $\mathrm{diag}\{0, \sigma\}$ replacing σ, (H_6^4) implies that for any $X_0 \in L^p(\Omega \to \mathscr{C}, \mathscr{F}_0, \mathbb{P})$ equation (6.15) has a unique mild segment solution. Let $P_t^*\nu = \mathscr{L}_{X_t}$ for $\mathscr{L}_{X_0} = \nu \in \mathscr{P}_p(\mathscr{C})$.

Theorem 6.5 ([40]). *Assume (H_6^4) for some constants $p \geq 1$ and $\alpha > 0$ satisfying*

$$\alpha \leq \alpha' := \frac{1}{2\|B\|}\big\{\delta - K_2 + \sqrt{(\delta - K_2)^2 + 4K_1\|B\|}\big\}, \tag{6.16}$$

where $\|\cdot\|$ is the operator norm. If

$$\inf_{s\in(0,\lambda_1]} s\mathrm{e}^{-sr_0} > K_2 + \alpha'\|B\| + K_3, \tag{6.17}$$

then P_t^* has a unique invariant probability measure $\bar{\mu}$ such that

$$\mathbb{W}_p(P_t^*\nu, \bar{\mu})^2 \leq c_1 \mathrm{e}^{-c_2 t}\mathbb{W}_p(\nu, \bar{\mu}), \quad \nu \in \mathscr{P}_p(\mathscr{C}), t \geq 0 \tag{6.18}$$

holds for some constants $c_1, c_2 > 0$, and $\{L_t^\nu\}_{\nu\in\mathscr{I}_{\varepsilon,R}} \in LDP(J)$ for any $\varepsilon, R > 1$, where J is the Donsker-Varadhan level 2 entropy function for the associated reference equation for $\bar{X}(t)$.

Example 6.4. Consider the following equation for $X(t) = (X^{(1)}(t), X^{(2)}(t))$ on $\mathbb{H} = \mathbb{H}_0 \times \mathbb{H}_0$ for a separable Hilbert space $\mathbb{H}_0$:

$$\begin{cases} \mathrm{d}X^{(1)}(t) = \{\alpha_1 X^{(2)}(t) - \lambda_1 X^{(1)}(t)\}\mathrm{d}t \\ \mathrm{d}X^{(2)}(t) = \{Z(X_t, \mathscr{L}_{X_t}) - AX^{(2)}(t)\}\mathrm{d}t + \mathrm{d}W(t), \end{cases}$$

where $\alpha_1 \in \mathbb{R}\setminus\{0\}$, $W(t)$ is the cylindrical Brownian motion on $\mathbb{H}_0$, A is a self-adjoint operator on $\mathbb{H}_0$ with discrete spectrum such that all eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \cdots$ counting multiplicities satisfy

$$\sum_{i=1}^{\infty} \lambda_i^{\gamma-1} < \infty$$

for some $\gamma \in (0,1)$, and Z satisfies

$$|Z(\xi_1, \nu_1) - Z(\xi_2, \nu_2)| \leq \alpha_2\|\xi_1 - \xi_2\|_\infty + \alpha_3 \mathbb{W}_2(\nu_1, \nu_2), \quad (\xi_i, \nu_i) \in \mathscr{C} \times \mathscr{P}_2(\mathscr{C}), i = 1, 2.$$

Let

$$\alpha = \frac{1}{2\alpha_1}\Big(\sqrt{\alpha_2^2 + 4\alpha_1\alpha_2} - \alpha_2\Big).$$

Then P_t^* has a unique invariant probability measure $\bar{\mu} \in \mathscr{P}_2(\mathscr{C})$, and $\{L_t^\nu\}_{\nu\in\mathscr{I}_{R,q}} \in LDP(J)$ for any $R, q > 1$ if

$$\inf_{s\in[0,\lambda_1]} s\mathrm{e}^{-sr_0} > \alpha_2 + \alpha_1\alpha + \frac{\alpha_3}{1 \wedge \alpha}. \tag{6.19}$$

Indeed, it is easy to see that assumption (H_6^4) holds for $p = 2$, $\delta = 0$, $\|B\| = \alpha_1$, $K_1 = K_2 = \alpha_2$ and $K_3 = \frac{\alpha_3}{1\wedge\alpha}$. So, we have $\alpha = \alpha'$ and (6.19) is equivalent to (6.17). Then the desired assertion follows from Theorem 6.5.

7 Comparison theorem

The order preservation of stochastic processes is a crucial property for one to compare a complicated process with simpler ones, and a result to ensure this property is called "comparison theorem" in the literature. There are two different type order preservations, one is in the distribution (weak) sense and the other is in the pathwise (strong) sense, where the latter implies the former. The weak order preservation has been investigated for diffusion-jump Markov processes in [10, 54] and references within, as well as a class of super processes in [53]. There are also plentiful results on the strong order preservation, see, for instance, [4, 14, 26, 32, 34, 56] and references within for comparison theorems on forward/backward SDEs (stochastic differential equations), with jumps and/or with memory. Recently, sufficient and necessary conditions have been derived in [20] for the order preservation of SDEs with memory.

On the other hand, path-distribution dependent SDEs have been investigated in [19], see also [51] and references within for distribution-dependent SDEs without memory. In this section, sufficient and necessary conditions of the order preservations for path-distribution dependent SDEs are presented.

Let $r_0 \geq 0$ be a constant and $d \geq 1$ be a natural number. The path space $\mathbf{C} = C([-r_0, 0]; \mathbb{R}^d)$ is Polish under the uniform norm $\|\cdot\|_\infty$. For any continuous map $f : [-r_0, \infty) \to \mathbb{R}^d$ and $t \geq 0$, let $f_t \in \mathbf{C}$ be such that $f_t(\theta) = f(\theta + t)$ for $\theta \in [-r_0, 0]$. We call $(f_t)_{t\geq 0}$ the segment of $(f(t))_{t\geq -r_0}$. Next, let $\mathscr{P}(\mathbf{C})$ be the set of probability measures on $\mathbf{C}$ equipped with the weak topology. Finally, let $W(t)$ be an m-dimensional Brownian motion on a complete filtration probability space $(\Omega, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$.

We consider the following Distribution-dependent SDEs with memory:

$$\begin{cases} \mathrm{d}X(t) = b(t, X_t, \mathscr{L}_{X_t})\,\mathrm{d}t + \sigma(t, X_t, \mathscr{L}_{X_t})\,\mathrm{d}W(t), \\ \mathrm{d}\bar{X}(t) = \bar{b}(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t})\,\mathrm{d}t + \bar{\sigma}(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t})\,\mathrm{d}W(t), \end{cases} \tag{7.1}$$

where

$$b, \bar{b} : [0, \infty) \times \mathscr{C} \times \mathscr{P}(\mathbf{C}) \to \mathbb{R}^d; \quad \sigma, \bar{\sigma} : [0, \infty) \times \mathscr{C} \times \mathscr{P}(\mathbf{C}) \to \mathbb{R}^d \otimes \mathbb{R}^m$$

are measurable.

For any $s \geq 0$ and $\mathscr{F}_s$-measurable $\mathbf{C}$-valued random variables $\xi, \bar{\xi}$, a solution to (7.1) for $t \geq s$ with $(X_s, \bar{X}_s) = (\xi, \bar{\xi})$ is a continuous adapted process $(X(t), \bar{X}(t))_{t\geq s}$ such that for all $t \geq s$,

$$\begin{aligned} X(t) &= \xi(0) + \int_s^t b(r, X_r, \mathscr{L}_{X_r})\mathrm{d}r + \int_s^t \sigma(r, X_r, \mathscr{L}_{X_r})\mathrm{d}W(r), \\ \bar{X}(t) &= \bar{\xi}(0) + \int_s^t \bar{b}(r, \bar{X}_r, \mathscr{L}_{\bar{X}_r})\mathrm{d}r + \int_s^t \bar{\sigma}(r, \bar{X}_r, \mathscr{L}_{\bar{X}_r})\mathrm{d}W(r), \end{aligned}$$

where $(X_t, \bar{X}_t)_{t\geq s}$ is the segment process of $(X(t), \bar{X}(t))_{t\geq s-r_0}$ with $(X_s, \bar{X}_s) = (\xi, \bar{\xi})$.

Following the line of [19], we consider the class of probability measures of finite second moment:

$$\mathscr{P}_2(\mathbf{C}) = \left\{ \nu \in \mathscr{P}(\mathbf{C}) : \nu(\|\cdot\|_\infty^2) := \int_{\mathbf{C}} \|\xi\|_\infty^2 \nu(\mathrm{d}\xi) < \infty \right\}.$$

It is a Polish space under the Wasserstein distance

$$\mathbb{W}_2(\mu_1,\mu_2) := \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)} \left(\int_{\mathbf{C}\times\mathbf{C}} \|\xi-\eta\|_\infty^2 \pi(\mathrm{d}\xi,\mathrm{d}\eta)\right)^{\frac{1}{2}}, \quad \mu_1,\mu_2\in\mathscr{P}_2(\mathbf{C}),$$

where $\mathscr{C}(\mu_1,\mu_2)$ is the set of all couplings for μ_1 and μ_2.

To investigate the order preservation, we make the following assumptions.

(H_7^1) (Continuity) There exists an increasing function $\alpha:\mathbb{R}_+\to\mathbb{R}_+$ such that for any $t\geq 0;\xi,\eta\in\mathbf{C};\mu,\nu\in\mathscr{P}_2(\mathbf{C})$,

$$\begin{aligned}&|b(t,\xi,\mu)-b(t,\eta,\nu)|^2+|\bar{b}(t,\xi,\mu)-\bar{b}(t,\eta,\nu)|^2+\|\sigma(t,\xi,\mu)-\sigma(t,\eta,\nu)\|_{HS}^2\\&+\|\bar{\sigma}(t,\xi,\mu)-\bar{\sigma}(t,\eta,\nu)\|_{HS}^2\leq\alpha(t)\big(\|\xi-\eta\|_\infty^2+\mathbb{W}_2(\mu,\nu)^2\big).\end{aligned}$$

(H_7^1) (Growth) There exists an increasing function $K:\mathbb{R}_+\to\mathbb{R}_+$ such that

$$|b(t,0,\delta_0)|^2+|\bar{b}(t,0,\delta_0)|^2+\|\sigma(t,0,\delta_0)\|_{HS}^2+\|\bar{\sigma}(t,0,\delta_0)\|_{HS}^2\leq K(t),\quad t\geq 0,$$

where δ_0 is the Dirac measure at point $0\in\mathbf{C}$.

It is easy to see that these two conditions imply assumptions $(H1)$-$(H3)$ in [19], so that by [19, Theorem 3.1], for any $s\geq 0$ and $\mathscr{F}_s$-measurable $\mathbf{C}$-valued random variables $\xi,\bar{\xi}$ with finite second moment, the equation (7.1) has a unique solution $\{X(s,\xi;t),\bar{X}(s,\bar{\xi};t)\}_{t\geq s}$ with $X_s=\xi$ and $\bar{X}_s=\bar{\xi}$. Moreover, the segment process $\{X(s,\xi)_t,\bar{X}(s,\bar{\xi})_t\}_{t\geq s}$ satisfies

$$\mathbb{E}\sup_{t\in[s,T]}\big(\|X(s,\xi)_t\|_\infty^2+\|\bar{X}(s,\bar{\xi})_t\|_\infty^2\big)<\infty,\quad T\in[s,\infty). \tag{7.2}$$

To characterize the order-preservation for solutions of (7.1), we introduce the partial-order on $\mathbf{C}$. For $x=(x^1,\cdots,x^d)$ and $y=(y^1,\cdots,y^d)\in\mathbb{R}^d$, we write $x\leq y$ if $x^i\leq y^i$ holds for all $1\leq i\leq d$. Similarly, for $\xi=(\xi^1,\cdots,\xi^d)$ and $\eta=(\eta^1,\cdots,\eta^d)\in\mathbf{C}$, we write $\xi\leq\eta$ if $\xi^i(\theta)\leq\eta^i(\theta)$ holds for all $\theta\in[-r_0,0]$ and $1\leq i\leq d$. A function f on $\mathbf{C}$ is called increasing if $f(\xi)\leq f(\eta)$ for $\xi\leq\eta$. Moreover, for any $\xi_1,\xi_2\in\mathbf{C}$, $\xi_1\wedge\xi_2\in\mathbf{C}$ is defined by

$$(\xi_1\wedge\xi_2)^i := \min\{\xi_1^i,\xi_2^i\},\quad 1\leq i\leq d.$$

For two probability measures $\mu,\nu\in\mathscr{P}(\mathbf{C})$, we write $\mu\leq\nu$ if $\mu(f)\leq\nu(f)$ holds for any increasing function $f\in C_b(\mathbf{C})$. According to [29, Theorem 5], $\mu\leq\nu$ if and only if there exists $\pi\in\mathscr{C}(\mu,\nu)$ such that $\pi(\{(\xi,\eta)\in\mathbf{C}^2:\xi\leq\eta\})=1$.

Definition 7.1. The stochastic differential system (7.1) is called order-preserving, if for any $s\geq 0$ and $\xi,\bar{\xi}\in L^2(\Omega\to\mathbf{C},\mathscr{F}_s,\mathbb{P})$ with $\mathbb{P}(\xi\leq\bar{\xi})=1$,

$$\mathbb{P}\big(X(s,\xi;t)\leq\bar{X}(s,\bar{\xi};t),\ t\geq s\big)=1.$$

We first present the following sufficient conditions for the order preservation, which reduce back to the corresponding ones in [20] when the system is distribution-independent.

Theorem 7.1. *Assume* (H_7^1) *and* (H_7^2). *The system* (7.1) *is order-preserving provided the following two conditions are satisfied:*

(1) *For any* $1 \le i \le d$, $\mu, \nu \in \mathscr{P}_2(\mathbf{C})$ *with* $\mu \le \nu$, $\xi, \eta \in \mathbf{C}$ *with* $\xi \le \eta$ *and* $\xi^i(0) = \eta^i(0)$,

$$b^i(t, \xi, \mu) \le \bar{b}^i(t, \eta, \nu), \quad a.e.\ t \ge 0.$$

(2) *For a.e.* $t \ge 0$ *it holds:* $\sigma(t, \cdot, \cdot) = \bar{\sigma}(t, \cdot, \cdot)$ *and* $\sigma^{ij}(t, \xi, \mu) = \sigma^{ij}(t, \eta, \nu)$ *for any* $1 \le i \le d$, $1 \le j \le m$, $\mu, \nu \in \mathscr{P}_2(\mathbf{C})$ *and* $\xi, \eta \in \mathbf{C}$ *with* $\xi^i(0) = \eta^i(0)$.

Condition (2) means that for a.e. $t \ge 0$, $\sigma(t, \xi, \mu) = \bar{\sigma}(t, \xi, \mu)$ and the dependence of $\sigma^{ij}(t, \xi, \mu)$ on (ξ, μ) is only via $\xi^i(0)$.

On the other hand, the next result shows that these conditions are also necessary if all coefficients are continuous on $[0, \infty) \times \mathbf{C} \times \mathscr{P}_2(\mathbf{C})$, so that [20, Theorem 1.2] is covered when the system is distribution-independent.

Theorem 7.2. *Assume* (H_7^1), (H_7^2) *and that* (7.1) *is order-preserving for any complete filtered probability space* $(\Omega, \{\mathscr{F}_t\}_{t \ge 0}, \mathbb{P})$ *and* m*-dimensional Brownian motion* $W(t)$ *thereon. Then for any* $1 \le i \le d$, $\mu, \nu \in \mathscr{P}_2(\mathbf{C})$ *with* $\mu \le \nu$, *and* $\xi, \eta \in \mathbf{C}$ *with* $\xi \le \eta$ *and* $\xi^i(0) = \eta^i(0)$, *the following assertions hold:*

(1′) $b^i(t, \xi, \mu) \le \bar{b}^i(t, \eta, \nu)$ *if* b^i *and* $\bar{b}^i$ *are continuous at points* (t, ξ, μ) *and* (t, η, ν) *respectively.*

(2′) *For any* $1 \le j \le m$, $\sigma^{ij}(t, \xi, \mu) = \bar{\sigma}^{ij}(t, \eta, \nu)$ *if* σ^{ij} *and* $\bar{\sigma}^{ij}$ *are continuous at points* (t, ξ, μ) *and* (t, η, ν) *respectively.*

Consequently, when $b, \bar{b}, \sigma$ *and* $\bar{\sigma}$ *are continuous on* $[0, \infty) \times \mathbf{C} \times \mathscr{P}_2(\mathbf{C})$, *conditions* (1) *and* (2) *hold.*

References

[1] S. Albeverio, Y. G. Kondratiev, M. Röckner, Differential geometry of Poisson spaces. C R Acad Sci Paris Sér I Math, 323(1996), 1129–1134.

[2] J. Bao, P. Ren, F.-Y. Wang, *Bismut formulas for Lions derivative of McKean-Vlasov SDEs with memory,* arxiv: 2004.14629.

[3] J. Bao, F.-Y. Wang, C. Yuan, *Hypercontractivity for functional stochastic partial differential equations,* Comm. Elect. Probab. 20(2015), 1–15.

[4] J. Bao, C. Yuan, *Comparison theorem for stochastic differential delay equations with jumps,* Acta Appl. Math. 116(2011), 119–132.

[5] V. Barbu, M. Röckner, *Probabilistic representation for solutions to nonlinear Fokker-Planck equations,* SIAM J. Math. Anal. 502018, 4246–4260.

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

[7] S. G. Bobkov, I. Gentil, M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations,* J. Math. Pures Appl. 80(2001), 669–696.

[8] V. Bogachev, N.V. Krylov, M. Röckner, *On regularity of transition probabilities and invariant measures of singular diffusions under minimal conditions,* Comm. Part. Diff. Equat. 26(2001), 2037–2080.

[9] Cardaliaguet P. Notes on mean field games. P.-L. Lions lectures at College de France. Online at https://www.ceremade.dauphine.fr/∼cardaliaguet/MFG20130420.pdf.

[10] M.-F. Chen, F.-Y. Wang, *On order-preservation and positive correlations for multidimensional diffusion processes,* Probab. Theory Relat. Fields 95(1993), 421–428.

[11] J. A. Carrillo, R. J. McCann, C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates,* Rev. Mat. Iberoam. 19(2003), 971–1018.

[12] M. D. Donsker, S. R. S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time,* I-IV, Comm. Pure Appl. Math. 28(1975), 1–47, 279–301; 29(1976), 389–461; 36(1983), 183–212.

[13] A. Dembo, O. Zeitouni, *Large Deviations Techniques and Applications,* Second Edition, Springer, New York. 1998.

[14] L. Gal'cuk, M. Davis, *A note on a comparison theorem for equations with different diffusions,* Stochastics 6(1982), 147–149.

[15] A. Guillin, W. Liu, L. Wu, *Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems,* arXiv:1909.07051, to appear in Ann. Appl. Probab.

[16] A. Guillin, F.-Y. Wang, *Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality,* J. Diff. Equat. 253(2012), 20–40.

[17] W. Hammersley, D. Šiška, L. Szpruch, *McKean-Vlasov SDE under measure dependent Lyapunov conditions,* arXiv:1802.03974.

[18] X. Huang, C. Liu, F.-Y. Wang, *Order preservation for path-distribution dependent SDEs,* Comm. Pure Appl. Anal 17(2018), 2125–2133.

[19] X. Huang, M. Röckner, F.-Y. Wang, *Nonlinear Fokker-Planck equations for probability measures on path space and path-distribution dependent SDEs,* Discrete Contin. Dyn. Syst. A. 39(2019), 3017–3035.

[20] X. Huang, F.-Y. Wang, *Order-preservation for multidimensional stochastic functional differential equations with jumps,* J. Evol. Equat. 14(2014),445–460.

[21] X. Huang, F.-Y. Wang, *Distribution dependent SDEs with singular coefficients,* Stoch. Proc. Appl. 129(2019), 4747–4770.

[22] X. Huang, F.-Y. Wang, *McKean-Vlasov SDEs with drifts discontinuous under Wasserstein distance,* arXiv:2002.06877, to appear in Disc. Contin. Dyn. Sys. A.

[23] X. Huang, F.-Y. Wang, *Derivative estimates on distributions of McKean-Vlasov SDEs,* arXiv:2006.16731.

[24] X. Huang, F.-Y. Wang, *Well-posedness for singular McKean-Vlasov stochastic differential equations,* arXiv:2012.05014.

[25] X. Huang, C. Yuan, *Comparison theorem for distribution dependent neutral SFDEs,* arXiv:1903.02360, to appear in J. Evol. Equat.

[26] N. Ikeda, S. Watanabe, *A comparison theorem for solutions of stochastic differential equations and its applications,* Osaka J. Math. 14(1977), 619–633.

[27] M. Kac, *Foundations of kinetic theory,* Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954C1955, vol. III, University of California Press, 171–197.

[28] M. Kac, *Probability and Related Topics in the Physical Sciences,* New York 1959.

[29] T. Kamae, U.Krengel, G. L. O'Brien, *Stochastic inequalities on partially ordered spaces,* Ann. Probab. 5(1977), 899-912.

[30] T. Kurtz, *Weak and strong solutions of general stochastic models,* Electr. Comm. Probab. 19(2014), paper no. 58, 16 pp.

[31] E. Lanconelli, S. Polidoro, *On a class of hypoelliptic evolution operator,* Rend. Sem. Mat. Univ. Pol. Torino 52(1994), 29–63.

[32] X. Mao, *A note on comparison theorems for stochastic differential equations with respect to semimartingales,* Stochastics 37(1991), 49–59.

[33] H.P. McKean, *A class of Markov processes associated with nonlinear parabolic equations,* Proc. Nat. Acad. Sci. U.S.A. 56(1966), 1907–1911.

[34] G. O'Brien, *A new comparison theorem for solution of stochastic differential equations,* Stochastics 3(1980), 245–249.

[35] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality,* J. Funct. Anal. 173(2000), 361–400.

[36] S. Peng, Z. Yang, *Anticipated backward stochastic differential equations,* Ann. Probab. 37(2009), 877–902.

[37] P. Ren, H. Tang, F.-Y. Wang, *Distribution-path dependent nonlinear SPDEs with application to stochastic transport type equations,* arXiv:2007.09188.

[38] P. Ren, F.-Y. Wang, *Bismut formula for Lions derivative of distribution dependent SDEs and applications,* J. Diff. Equat. 267(2019), 4745–4777.

[39] P. Ren, F.-Y. Wang, *Derivative formulas in measure on Riemannian manifolds,* arXiv:1908.03711.

[40] P. Ren, F.-Y. Wang, *Donsker-Varadhan large deviations for path-distribution dependent SPDEs,* arXiv:2002.08652.

[41] P. Ren, F.-Y. Wang, *Exponential convergence in entropy and Wasserstein distance for McKean-Vlasov SDEs,* arXiv:2010.08950.

[42] M. Röckner, F.-Y. Wang, *Log-harnack inequality for stochastic differential equations in Hilbert spaces and its consequences*, Infin. Dim. Anal. Quat. Probab. Relat. Top. 13(2010), 27–37.

[43] M. Röckner, X. Zhang, *Well-posedness of distribution dependent SDEs with singular drifts,* arXiv:1809.02216, to appear in Bernoulli.

[44] P.-E. C. de Raynal, N. Frikha, *Well-posedness for some non-linear diffusion processes and related pde on the wasserstein space,* arXiv:1811.06904.

[45] A.-S. Sznitman, *Topics in propagations of chaos,* Lecture notes in Math. Vol. 1464, pp. 165–251, Springer, Berlin, 1991.

[46] M. Talagrand, *Transportation cost for Gaussian and other product measures,* Geom. Funct. Anal. 6(1996), 587–600.

[47] F.-Y. Wang, *Logarithmic Sobolev inequalities on noncompact Riemannian manifolds,* Probab. Theory Related Fields, 109(1997), 417–424.

[48] F.-Y. Wang, *Probability distance inequalities on Riemannian manifolds and path spaces,* J. Funct. Anal. 206(2004), 167–190.

[49] F.-Y. Wang, *Harnack inequalities on manifolds with boundary and applications,* J. Math. Pures Appl. 94(2010), 304–321.

[50] F.-Y. Wang, *Harnack Inequalities and Applications for Stochastic Partial Differential Equations*, Springer, 2013, Berlin.

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

[52] F.-Y. Wang, X. Zhang, *Derivative formula and applications for degenerate diffusion semigroups,* J. Math. Pures Appl. 99(2013), 726–740.

[53] F.-Y. Wang, *The stochastic order and critical phenomena for superprocesses,* Infin. Dimens. Anal. Quantum Probab. Relat. Top. 9(2006), 107–128.

[54] J.-M. Wang, *Stochastic comparison for Lévy-type processes,* J. Theor. Probab. 26(2013), 997-1019.

[55] T. Yamada, S. Watanabe, *On the uniqueness of solutions of stochastic differential equations,* J. Math. Kyoto Univ. 2(1971), 155–167.

[56] Z. Yang, X. Mao, C. Yuan, *Comparison theorem of one-dimensional stochastic hybrid systems,* Systems Control Lett. 57(2008), 56–63.

[57] X. Zhang, *Weak solutions of McKean-Vlasov SDEs with supercritical drifts,* arXiv:2010.15330.

[58] G. Zhao, *On distribution dependent SDEs with singular drifts,* arXiv:2003.04829v3.

[59] X. Zhu, *On the comparison theorem for multi-dimensional stochastic differential equations with jumps (in Chinese),* Sci. Sin. Math. 42(2012), 303–311.