Exponential Ergodicity and Propagation of Chaos for Path-Distribution Dependent Stochastic Hamiltonian System*

Xing Huang [a)], **Wujun Lv** [b)]

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

xinghuang@tju.edu.cn

b) Department of Statistics, College of Science, Donghua University, Shanghai, 201620, China

lvwujun@dhu.edu.cn

December 31, 2023

Abstract

By Girsanov's thoerem and using the existing log-Harnack inequality for distribution independent SDEs, the log-Harnack inequality is derived for path-distribution dependent stochastic Hamiltonian systems. As an application, the exponential ergodicity in relative entropy is obtained by combining with transportation cost inequality. In addition, the quantitative propagation of chaos in the sense of Wasserstein distance, which together with the coupling by change of measure implies the quantitative propagation of chaos in total variation norm as well as relative entropy are obtained.

AMS subject Classification: 60H10, 60H15.
Keywords: Stochastic Hamiltonian system, Path-Distribution dependent, Exponential ergodicity, Log-Harnack inequality, Propagation of chaos.

1 Introduction

The stochastic Hamiltonian system (SHS), which includes the kinetic Fokker-Planck equation (see [31]), has been extensively investigated in [6, 11, 15, 17, 34, 35, 37, 38] and references therein. More precisely, [11] has studied the regularity of stochastic kinetic equations; [15] investigated Bismut formula, gradient estimate and Harnack inequality for SHS by using

*Supported in part by NNSFC (12271398, 12301174).

coupling by change of measure; the derivative formula is extended to the case that the degenerate part is not linear by using Malliavin calculus in [35] and [37]; moreover, [37] derived the stochastic flows for SHS with linear degenerate part, and the diffusion only depends on the degenerate part; see also [38] for the results on the stochastic flows with singular coefficients; we refer to [34] for the hypercontractivity for SHS. For the path-dependent SHS, the derivative formula and Harnack inequality are established in [6], see also [17] for Harnack inequalities with singular drifts.

Recently, along with the application in nonlinear Fokker-Planck-Kolmogorov equations, McKean-Vlasov stochastic differential equations (SDEs), presented in [22], have gained much attention. There are plentiful results on these type SDEs, see for instance, [3, 4, 9, 16, 21, 28, 39] and references therein. In [27], the exponential ergodicity of McKean-Vlasov SDEs in relative entropy is derived by log-Harnack inequality and transportation cost inequality. The log-Harnack inequality for non-degenerate McKean-Vlasov SDEs is investigated in [33] by coupling by change of measure. One can also refer to [18] for the log-Harnack inequality of non-degenerate McKean-Vlasov SDEs with memory. In addition, there are lots of references on the well-posedness of McKean-Vlasov SDEs with singular coefficients, for instance, [9, 16, 19, 21, 24, 28, 39] and references therein. Since in this paper we do not plan to pay attention to the well-posedness for McKean-Vlasov SDEs with singular coefficients, we will not characterize the details of the well-posedness results in the above references and we will give the well-posedness result using the appendix in Section 5.

To obtain the log-Harnack inequality for the path-distribution dependent SHS, we will adopt Girsanov's transform and combine with the existing log-Harnack inequality in [32] and [17].

McKean-Vlasov SDEs can be viewed as the limit of the mean field interacting particle system. The so called propagation of chaos ([30]) means that the joint distribution of finite many particles converges to the product of the distribution of McKean-Vlasov SDEs as the number of interacting particle system tends to infinity, see [14, Definition 4.1] for more details. For the propagation of chaos, [21] obtain the convergence of the interacting particle system with non-degenerate noise in the total variation distance. In this paper, we obtain the convergence of the interacting particle system in the sense of Wasserstein distance, total variation norm and relative entropy, see Theorem 4.2 below. Since $\mathscr{C}^{m+d}$ is an infinite dimensional space, to obtain the quantitative propagation of chaos, we assume that the coefficients are Lipschitz continuous in $\mathbb{W}_\theta^\Gamma$ instead of L^θ-Wasserstein distance. For more results on the propagation of chaos, see [2, 7, 13, 14, 20, 25, 30] and references therein.

The main contributions of this paper mainly include: (1) The diffusion is degenerate. (2) The model is assumed to be both path and distribution dependent. (3) The quantitative propagation of chaos in the sense of total variation norm and relative entropy is obtained.

The paper is organized as follows: In Section 2, we prove the log-Harnack inequality for path-distribution dependent SHS; The exponential ergodicity in relative entropy is derived in Section 3, where the transportation cost inequality for the invariant probability measure is also investigated under the dissipative condition; In section 4, the quantitative propagation of chaos for path-distribution dependent SHS is studied. Finally, the well-posedness for general path-distribution dependent SDEs and mean field interacting particle system is provided in

Section 5.

Throughout the paper, fix a constant $r > 0$. For any $n \in \mathbb{N}^+$, let $\mathscr{C}^n = C([-r, 0]; \mathbb{R}^n)$ be equipped with the uniform norm $\|\xi\|_\infty =: \sup_{s\in[-r,0]} |\xi(s)|$. For any $f \in C([-r, \infty); \mathbb{R}^n)$, $t \geq 0$, define $f_t \in \mathscr{C}^n$ as $f_t(s) = f(t + s), s \in [-r, 0]$, which is called the segment process. Let $\mathscr{P}(\mathscr{C}^n)$ be the set of all probability measures in $\mathscr{C}^n$ equipped with the weak topology. For $\theta \geq 1$, define

$$\mathscr{P}_\theta(\mathscr{C}^n) = \big\{\mu \in \mathscr{P}(\mathscr{C}^n) : \mu(\|\cdot\|_\infty^\theta) < \infty\big\}.$$

It is well known that $\mathscr{P}_\theta(\mathscr{C}^n)$ is a Polish space under the Wasserstein distance

$$\mathbb{W}_\theta(\mu, \nu) := \inf_{\pi\in\mathbf{C}(\mu,\nu)} \bigg(\int_{\mathscr{C}^n\times\mathscr{C}^n} \|\xi - \eta\|_\infty^\theta \pi(\mathrm{d}\xi, \mathrm{d}\eta)\bigg)^{\frac{1}{\theta}}, \quad \mu, \nu \in \mathscr{P}_\theta(\mathscr{C}^n),$$

where $\mathbf{C}(\mu, \nu)$ is the set of all couplings of μ and ν.

Recall that for two probability measures μ, ν on some measurable space $(E, \mathscr{E})$, the entropy and total variation norm are defined as follows:

$$\mathrm{Ent}(\nu|\mu) := \begin{cases} \int_E (\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu})\mathrm{d}\nu, & \text{if } \nu \text{ is absolutely continuous with respect to } \mu, \\ \infty, & \text{otherwise;} \end{cases}$$

and

$$\|\mu - \nu\|_{var} := \sup_{|f|\leq 1} |\mu(f) - \nu(f))|.$$

By Pinsker's inequality (see [26]),

$$\|\mu - \nu\|_{var}^2 \leq 2\mathrm{Ent}(\nu|\mu), \quad \mu, \nu \in \mathscr{P}(E), \tag{1.1}$$

here $\mathscr{P}(E)$ denotes all probability measures on $(E, \mathscr{E})$. Throughout the paper, we will use C or c as a constant, the values of which may change from one place to another. For $n, k \in \mathbb{N}^+$, let 0_n and $0_{n\times k}$ denote the n dimensional vector and $n \times k$ matrix with all components being 0.

2 Log-Harnack Inequality

The log-Harnack inequality provides an estimate of the relative entropy for two probability measures, see for instance [32, Theorem 1.4.2 (2)]. For the path dependent SHS, the log-Harnack inequality has been established in [32, Theorem 4.4.5], see also [17] for the case with singular drifts. [18] studied log-Harnack inequality for path-distribution dependent SDEs with non-degenerate noise and the result is extended to the path-distribution dependent SDEs with singular drift in [16]. Moreover, by Girsanov's transform and Young's inequality, the log-Harnack inequality is obtained in [19], where the semi-linear SPDE with Dini continuous drift and non-degenerate noise is considered. In this section, we extend the method in [19] to the path-distribution dependent case including the path-distribution SHS. To this end, we first give a general result as follows.

2.1 A General Result

Let $T > r$ and $n, k \in \mathbb{N}^+$. Consider SDE on $\mathbb{R}^n$:

$$\mathrm{d}X(t) = H_0(t, X_t)\mathrm{d}t + \Sigma(t, X_t)H(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \Sigma(t, X_t)\mathrm{d}W(t), \tag{2.1}$$

where $H_0 : [0,\infty) \times \mathscr{C}^n \to \mathbb{R}^n, H : [0,\infty) \times \mathscr{C}^n \times \mathscr{P}(\mathscr{C}^n) \to \mathbb{R}^k$, $\Sigma : [0,\infty) \times \mathscr{C}^n \to \mathbb{R}^n \otimes \mathbb{R}^k$ are measurable and $W(t)$ is a k-dimensional Brownian motion on some complete filtration probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$.

Let $\hat{\mathscr{P}}(\mathscr{C}^n)$ be a subset of $\mathscr{P}(\mathscr{C}^n)$ containing all Dirac measures and it is equipped with some topology. Assume that (2.1) is well-posed in $\hat{\mathscr{P}}(\mathscr{C}^n)$, see Definition 5.1 and Theorem 5.1 for general result on the well-posedness of path-distribution depedndent SDEs. For any $\mu_0 \in \hat{\mathscr{P}}(\mathscr{C}^n)$, let $X_t^{\mu_0}$ be the unique solution to (2.1) with initial distribution μ_0 and define

$$P_t f(\mu_0) = (P_t^*\mu_0)(f) = \mathbb{E}f(X_t^{\mu_0}), \quad f \in \mathscr{B}_b(\mathscr{C}^n), t \geq 0. \tag{2.2}$$

For any $\mu \in C([0,T]; \hat{\mathscr{P}}(\mathscr{C}^n))$ and any $\mathscr{F}_0$-measurable random variable X_0 with $\mathscr{L}_{X_0} \in \hat{\mathscr{P}}(\mathscr{C}^n)$, suppose that the decoupled SDE

$$\mathrm{d}X^{X_0,\mu}(t) = H_0(t, X_t^{X_0,\mu})\mathrm{d}t + \Sigma(t, X_t^{X_0,\mu})H(t, X_t^{X_0,\mu}, \mu_t)\mathrm{d}t + \Sigma(t, X_t^{X_0,\mu})\mathrm{d}W(t) \tag{2.3}$$

with $X_0^{X_0,\mu} = X_0$ has a unique strong solution. Note that (2.3) reduces to a path dependent classical SDE, see [17, 36, 38] and references therein for the well-posedness with singular coefficients. Let P_t^μ be the associated semigroup to (2.3), i.e.

$$P_t^\mu f(\xi) = \mathbb{E}f(X_t^{\xi,\mu}), \quad \xi \in \mathscr{C}^n, f \in \mathscr{B}_b(\mathscr{C}^n), t \geq 0.$$

For $\nu \in C([0,T]; \hat{\mathscr{P}}(\mathscr{C}^n))$, let

$$\begin{aligned}
\zeta_t^{\mu,\nu} &= H(t, X_t^{X_0,\mu}, \mu_t) - H(t, X_t^{X_0,\mu}, \nu_t), \\
R_t^{\mu,\nu} &= \exp\left\{-\int_0^t \langle \zeta_s^{\mu,\nu}, \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^t |\zeta_s^{\mu,\nu}|^2 \mathrm{d}s\right\}, \quad t \in [0,T].
\end{aligned}$$

Theorem 2.1. *Assume that for any $\mu, \nu \in C([0,T], \hat{\mathscr{P}}(\mathscr{C}^n))$, $\{R_t^{\mu,\nu}\}_{t\in[0,T]}$ is a martingale and P_t^ν satisfies the log-Harnack inequality, i.e. there exists a function $C : (r,\infty) \to (0,\infty)$ such that for any $f \in \mathscr{B}_b(\mathscr{C}^n)$ with $f > 0$*

$$P_t^\nu \log f(\xi) \leq \log P_t^\nu f(\eta) + C(t)\|\xi - \eta\|_\infty^2, \quad r < t \leq T, \xi, \eta \in \mathscr{C}^n. \tag{2.4}$$

Then we have

$$\begin{aligned}
P_t \log f(\nu_0) &\leq \log P_t f(\mu_0) \\
&\quad + 2C(t)\mathbb{W}_2(\mu_0, \nu_0)^2 + \log \mathbb{E}(R_t^{\mu,\nu})^2, \quad r < t \leq T, \mu_0, \nu_0 \in \hat{\mathscr{P}}(\mathscr{C}^n).
\end{aligned} \tag{2.5}$$

Consequently,

$$\frac{1}{2}\|P_t^*\mu_0 - P_t^*\nu_0\|_{var}^2 \leq \mathrm{Ent}(P_t^*\mu_0|P_t^*\nu_0) \leq 2C(t)\mathbb{W}_2(\mu_0,\nu_0)^2 + \log \mathbb{E}(R_t^{\mu,\nu})^2, \quad r < t \leq T.$$

Proof. By [32, Theorem 1.4.2 (2)] and (1.1), it is sufficient to prove the log-Harnack inequality (2.5).

Let X_0 satisfy $\mathscr{L}_{X_0} = \mu_0$ and let $\mu_t = P_t^*\mu_0$ and $\nu_t = P_t^*\nu_0$, $\bar{W}(t) = W(t) + \int_0^t \zeta_s^{\mu,\nu}\mathrm{d}s$, $t \in [0,T]$. Since $\{R_t^{\mu,\nu}\}_{t\in[0,T]}$ is a martingale, it follows from Girsanov's theorem that $\{\bar{W}(t)\}_{t\in[0,T]}$ is a k-dimensional Brownian motion under $\mathbb{Q}_T = R_T^{\mu,\nu}\mathbb{P}$. So, (2.3) can be rewritten as

$$\mathrm{d}X^{X_0,\mu}(t) = H_0(t, X_t^{X_0,\mu}) + \Sigma(t, X_t^{X_0,\mu})H(t, X_t^{X_0,\mu}, \nu_t)\mathrm{d}t + \Sigma(t, X_t^{X_0,\mu})\mathrm{d}\bar{W}(t), \quad X_0^{X_0,\mu} = X_0.$$

Letting $\bar{\mu}_t = \mathscr{L}_{X_t^{X_0,\mu}}|\mathbb{Q}_T$ and noting that $\{R_t^{\mu,\nu}\}_{t\in[0,T]}$ is a martingale, we derive

$$\bar{\mu}_t(f) = \mathbb{E}^{\mathbb{Q}_T} f(X_t^{X_0,\mu}) = \mathbb{E}(R_t^{\mu,\nu} f(X_t^{X_0,\mu})), \quad f \in \mathscr{B}_b(\mathscr{C}^n), t \in [0,T],$$

which implies that for any $t \in [0,T]$, $\mathbb{P}$-a.s.

$$\frac{\mathrm{d}\bar{\mu}_t}{\mathrm{d}\mu_t}(X_t^{X_0,\mu}) = \mathbb{E}(R_t^{\mu,\nu}|X_t^{X_0,\mu}).$$

By Jensen's inequality for conditional expectation, we get

$$\bar{\mu}_t\left(\frac{\mathrm{d}\bar{\mu}_t}{\mathrm{d}\mu_t}\right) = \mathbb{E}(\mathbb{E}(R_t^{\mu,\nu}|X_t^{X_0,\mu})^2) \le \mathbb{E}(R_t^{\mu,\nu})^2, \quad t \in [0,T]. \tag{2.6}$$

On the other hand, taking expectation in (2.4) with respect to any $\pi \in \mathbf{C}(\nu_0,\mu_0)$, using Jensen's inequality and then taking infimum in $\pi \in \mathbf{C}(\nu_0,\mu_0)$, we get

$$(P_t^*\nu_0)(\log f) \le \log \bar{\mu}_t(f) + C(t)\mathbb{W}_2(\mu_0,\nu_0)^2, \quad r < t \le T.$$

This together with [32, Theorem 1.4.2 (2)] implies that

$$\mathrm{Ent}(P_t^*\nu_0|\bar{\mu}_t) = \bar{\mu}_t\left(\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\log\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\right) \le C(t)\mathbb{W}_2(\mu_0,\nu_0)^2.$$

It follows from Young's inequality (see for instance [1, Lemma 2.4]) and (2.6) that

$$\begin{aligned}
P_t\log f(\nu_0) &= \mu_t\left(\frac{\mathrm{d}\bar{\mu}_t}{\mathrm{d}\mu_t}\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\log f\right)\\
&\le \log P_t f(\mu_0) + \mu_t\left(\frac{\mathrm{d}\bar{\mu}_t}{\mathrm{d}\mu_t}\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\log\left(\frac{\mathrm{d}\bar{\mu}_t}{\mathrm{d}\mu_t}\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\right)\right)\\
&= \log P_t f(\mu_0) + \bar{\mu}_t\left(\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\log\frac{\mathrm{d}\bar{\mu}_t}{\mathrm{d}\mu_t}\right) + \bar{\mu}_t\left(\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\log\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\right)\\
&\le \log P_t f(\mu_0) + \log\bar{\mu}_t\left(\frac{\mathrm{d}\bar{\mu}_t}{\mathrm{d}\mu_t}\right) + 2\bar{\mu}_t\left(\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\log\frac{\mathrm{d}P_t^*\nu_0}{\mathrm{d}\bar{\mu}_t}\right)\\
&\le \log P_t f(\mu_0) + \log\mathbb{E}(R_t^{\mu,\nu})^2 + 2C(t)\mathbb{W}_2(\mu_0,\nu_0)^2.
\end{aligned}$$

Therefore, we complete the proof. □

2.2 Log-Harnack Inequality and Regularity for Path-Distribution Dependent SHS

Let $m, d \in \mathbb{N}^+$. In this section, consider the following path-distribution dependent stochastic Hamiltonian system on $\mathbb{R}^{m+d}$:

$$\begin{cases} \mathrm{d}X(t) = \{AX(t) + MY(t)\}\mathrm{d}t, \\ \mathrm{d}Y(t) = \{Z(X(t), Y(t), \mathscr{L}_{(X_t,Y_t)}) + B(X_t, Y_t, \mathscr{L}_{(X_t,Y_t)})\}\mathrm{d}t + \sigma \mathrm{d}W(t), \end{cases} \tag{2.7}$$

where $W = (W(t))_{t\geq 0}$ is a d-dimensional standard Brownian motion with respect to a complete filtration probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$, A is an $m \times m$ matrix, M is an $m \times d$ matrix, σ is a $d \times d$ matrix, $Z : \mathbb{R}^{m+d} \times \mathscr{P}(\mathscr{C}^{m+d}) \to \mathbb{R}^d$, $B : \mathscr{C}^{m+d} \times \mathscr{P}(\mathscr{C}^{m+d}) \to \mathbb{R}^d$. We should remark that the reason why we assume that the coefficients are time independent is only to coincide with the assertion in Section 3 and the result in Theorem 2.2 below can also be available in the time dependent case. To obtain the log-Harnack inequality, we make the following assumptions:

(A1) σ is invertible.

(A2) There exists $\theta \geq 1$ and $K_Z > 0$ such that

$$|Z(z,\gamma) - Z(\bar{z},\bar{\gamma})| \leq K_Z(|z-\bar{z}| + \mathbb{W}_\theta(\gamma,\bar{\gamma})), \quad z, \bar{z} \in \mathbb{R}^{m+d}, \gamma, \bar{\gamma} \in \mathscr{P}_\theta(\mathscr{C}^{m+d}).$$

(A3) Let θ be in **(A2)**. There exists a constant $K_B > 0$ such that

$$|B(\xi,\gamma) - B(\eta,\bar{\gamma})| \leq K_B(\|\xi-\eta\|_\infty + \mathbb{W}_\theta(\gamma,\bar{\gamma})), \quad \xi, \eta \in \mathscr{C}^{m+d}, \gamma, \bar{\gamma} \in \mathscr{P}_\theta(\mathscr{C}^{m+d}).$$

(A4) There exists an integer l with $0 \leq l \leq m-1$ such that

$$\mathrm{Rank}[M, AM, \cdots, A^l M] = m.$$

According to Remark 5.2 below, under **(A1)**-**(A3)**, (2.7) is well-posed in $\mathscr{P}_\theta(\mathscr{C}^{m+d})$. Denote the solution to (2.7) with $\mathscr{L}_{(X_0,Y_0)} = \mu_0 \in \mathscr{P}_\theta(\mathscr{C}^{m+d})$ by $(X_t^{\mu_0}, Y_t^{\mu_0})$. Let P_t and P_t^* be defined in the same way as in (2.2) for $(X_t^{\mu_0}, Y_t^{\mu_0})$ replacing $X_t^{\mu_0}$ there. The next result characterizes the log-Harnack inequality for (2.7).

Theorem 2.2. *Assume* **(A1)**-**(A4)** *and let $t > r$. Then for any $\mu_0, \nu_0 \in \mathscr{P}_\theta(\mathscr{C}^{m+d})$ and positive $f \in \mathscr{B}_b(\mathscr{C}^{m+d})$,*

$$P_t \log f(\mu_0) \leq \log P_t f(\nu_0) + C^2 \int_0^t \mathrm{e}^{2Cs}\mathrm{d}s\mathbb{W}_\theta(\mu_0,\nu_0)^2 + \Sigma(t,r,\|M\|,l)\mathbb{W}_2(\mu_0,\nu_0)^2,$$

where

$$\Sigma(t,r,\|M\|,l) = C\left\{\Big(\frac{1}{(t-r)\wedge 1} + \frac{\|M\|}{(t-r)^{(4l+3)}\wedge 1}\Big) + \Big(1 + \frac{\|M\|}{(t-r)^{2l+1}\wedge 1}\Big)^2\right\},$$

and $C>0$ is a constant. Consequently, it holds

$$\begin{aligned}&\frac{1}{2}\|P_t^*\mu_0-P_t^*\nu_0\|_{var}^2\le \mathrm{Ent}(P_t^*\mu_0|P_t^*\nu_0)\\&\le C^2\int_0^t \mathrm{e}^{2Cs}\mathrm{d}s\mathbb{W}_\theta(\mu_0,\nu_0)^2+\Sigma(t,r,\|M\|,l)\mathbb{W}_2(\mu_0,\nu_0)^2.\end{aligned}\tag{2.8}$$

Proof. Let $n=m+d$, $k=d$,

$$H_0(x,y)=\begin{pmatrix}Ax+My\\0_d\end{pmatrix},\quad H=\sigma^{-1}(Z+B),\quad \Sigma=\begin{pmatrix}0_{m\times d}\\\sigma\end{pmatrix},\quad x\in\mathbb{R}^m,y\in\mathbb{R}^d.$$

Let $\mu_t=P_t^*\mu_0$ and $\nu_t=P_t^*\nu_0$. For simplicity, we denote $(X_s,Y_s)=(X_s^{\mu_0},Y_s^{\mu_0})$. Set

$$\zeta_s^{\mu,\nu}=\sigma^{-1}[Z(X(s),Y(s),\mu_s)+B(X_s,Y_s,\mu_s)-Z(X(s),Y(s),\nu_s)-B(X_s,Y_s,\nu_s)].$$

By **(A2)**-**(A3)** and Remark 5.2 below, there exists a constant $C>0$ such that

$$|\zeta_s^{\mu,\nu}|\le\|\sigma^{-1}\|(K_Z+K_B)\mathbb{W}_\theta(\mu_s,\nu_s)\le C\mathrm{e}^{Cs}\mathbb{W}_\theta(\mu_0,\nu_0),\quad s\in[0,t].$$

Recalling the definition of $R_t^{\mu,\nu}$ in Theorem 2.1, we arrive at

$$\log\mathbb{E}(R_t^{\mu,\nu})^2\le\log\mathrm{esssup}_\Omega\mathrm{e}^{\int_0^t|\zeta_s^{\mu,\nu}|^2\mathrm{d}s}\le\int_0^t C^2\mathrm{e}^{2Cs}\mathbb{W}_\theta(\mu_0,\nu_0)^2\mathrm{d}s.$$

On the other hand, by [32, Theorem 4.4.5], we know

$$P_t^\nu\log f(\xi)\le\log P_t^\nu f(\eta)+\Sigma(t,r,\|M\|,k)\|\xi-\eta\|_\infty^2.$$

So, applying Theorem 2.1, we complete the proof. □

3 Exponential Ergodicity

In this section, we investigate the exponential ergodicity of (2.7) in L^2-Wasserstein distance as well as in relative entropy. To this end, we assume

(C) There exist $\lambda_1>0,\lambda_2,\lambda_3\ge0$ with $\lambda_2+\lambda_3<\sup_{\delta\in[0,\lambda_1]}\delta\mathrm{e}^{-\delta r}$ such that for any $\xi=(\xi^{(1)},\xi^{(2)}),\bar\xi=(\bar\xi^{(1)},\bar\xi^{(2)})\in\mathscr{C}^{m+d},\gamma,\bar\gamma\in\mathscr{P}_2(\mathscr{C}^{m+d})$,

$$\begin{aligned}&2\langle A(\xi^{(1)}(0)-\bar\xi^{(1)}(0))+M(\xi^{(2)}(0)-\bar\xi^{(2)}(0)),\ \ \xi^{(1)}(0)-\bar\xi^{(1)}(0)\rangle,\\&+2\langle Z(\xi(0),\gamma)-Z(\bar\xi(0),\bar\gamma)+B(\xi,\gamma)-B(\bar\xi,\bar\gamma),\ \ \xi^{(2)}(0)-\bar\xi^{(2)}(0)\rangle\\&\le-\lambda_1|\xi(0)-\bar\xi(0)|^2+\lambda_2\|\xi-\bar\xi\|_\infty^2+\lambda_3\mathbb{W}_2(\gamma,\bar\gamma)^2.\end{aligned}$$

Theorem 3.1. *Assume* **(C)** *and* **(A1)**-**(A4)** *with $\theta=2$. Then P_t^* has a unique invariant probability measure $\mu^*\in\mathscr{P}_2(\mathscr{C}^{m+d})$ with*

$$\begin{aligned}&\max(\mathbb{W}_2(P_t^*\nu,\mu^*)^2,\mathrm{Ent}(P_t^*\nu|\mu^*))\\&\le c\mathrm{e}^{-2\kappa t}\min(\mathbb{W}_2(\nu,\mu^*)^2,\mathrm{Ent}(\nu|\mu^*)),\quad \nu\in\mathscr{P}_2(\mathscr{C}^{m+d}),t>2r\end{aligned}$$

for some constants $c,\kappa>0$.

Proof. By [18, Remark 2.1], **(C)** implies that there exist constants $c_0, \kappa > 0$ such that

$$\mathbb{W}_2(P_t^*\mu_0, P_t^*\nu_0) \le c_0 \mathrm{e}^{-\kappa t}\mathbb{W}_2(\mu_0, \nu_0), \quad \mu_0, \nu_0 \in \mathscr{P}_2(\mathscr{C}^{m+d}), t > 0.$$

Then it is standard to prove that P_t^* has a unique invariant probability measure $\mu^* \in \mathscr{P}_2(\mathscr{C}^{m+d})$ with

$$\mathbb{W}_2(P_t^*\nu, \mu^*)^2 \le c_0^2 \mathrm{e}^{-2\kappa t}\mathbb{W}_2(\nu, \mu^*)^2, \quad \nu \in \mathscr{P}_2(\mathscr{C}^{m+d}), t > 0. \tag{3.1}$$

Combining this with (2.8) for $t = 2r$ and (3.2) below, we complete the proof by using [27, Theorem 2.1]. □

3.1 Transportation Cost Inequality

To obtain the exponential ergodicity in relative entropy, we also need to prove the transportation cost inequality for μ^*. [5] give a proof of transportation cost inequality for the solution to path dependent SDEs starting from dirac measure and the technique used there is also available in the present case. Furthermore, under the dissipative condition **(C)**, we can derive a uniform constant with respect to time variable T in the transportation cost inequality for the solution to (2.7) on $[0, T]$ starting from dirac measure, see (3.6) below. Then applying [10, Lemma 2.1] and [10, Lemma 2.2], the stability of transportation cost inequality, μ^* satisfies the transportation cost inequality due to (3.1).

Theorem 3.2. *Assume* **(C)**. *Then the transportation cost inequality holds for the invariant probability measure μ^*, i.e.*

$$\mathbb{W}_2(\nu, \mu^*)^2 \le 2\mathrm{e}^{(\lambda_1 - \epsilon)r}\frac{\|\sigma\|^2}{\epsilon}\mathrm{Ent}(\nu|\mu^*), \quad \nu \in \mathscr{P}_2(\mathscr{C}^{m+d}) \tag{3.2}$$

with some constant $\epsilon \in (0, \lambda_1)$.

Proof. Let $\xi \in \mathscr{C}^{m+d}$ and $X(t) = (X^1(t), X^2(t))$ solve

$$\begin{cases} \mathrm{d}X^1(t) = \{AX^1(t) + MX^2(t)\}\mathrm{d}t, \\ \mathrm{d}X^2(t) = \{Z(X(t), \mu^*) + B(X_t, \mu^*)\}\mathrm{d}t + \sigma \mathrm{d}W(t) \end{cases} \tag{3.3}$$

with $X_0 = (X_0^1, X_0^2) = \xi$. Let $P_t^{\mu^*}(\xi, \mathrm{d}\eta) = \mathscr{L}_{X_t}(\mathrm{d}\eta)$. According to **(C)** and [18, Remark 2.1], μ^* is the unique invariant probability measure of (3.3) and there exist constants $\tilde{c}, \tilde{\kappa} > 0$ such that

$$\mathbb{W}_2(P_t^{\mu^*}(\xi, \cdot), \mu^*) \le \tilde{c}\mathrm{e}^{-\tilde{\kappa}t}\mathbb{W}_2(\delta_\xi, \mu^*). \tag{3.4}$$

As in the proof of [5, Lemma 2.2], denote by Π_ξ^T as the distribution of $(X_t)_{t\in[0,T]}$. Define the distance

$$\rho_\infty^T(V, \tilde{V}) = \sup_{t\in[0,T]} \|V_t - \tilde{V}_t\|_\infty, \quad V, \tilde{V} \in C([0, T]; \mathscr{C}^{m+d}).$$

Let $(h(t))_{t\in[0,T]}$ be an $\mathbb{R}^d$-valued $\mathscr{F}_t$-predictable process and $Y(t)=(Y^1(t),Y^2(t))$ solve

$$\begin{cases}\mathrm{d}Y^1(t)=\{AY^1(t)+MY^2(t)\}\mathrm{d}t,\\ \mathrm{d}Y^2(t)=\{Z(Y(t),\mu^*)+B(Y_t,\mu^*)\}\mathrm{d}t-\sigma h(t)\mathrm{d}t+\sigma\mathrm{d}W(t)\end{cases}$$

with $Y_0=(Y_0^1,Y_0^2)=\xi$. Let $\alpha(\epsilon):=2\mathrm{e}^{(\lambda_1-\epsilon)r}\frac{\|\sigma\|^2}{\epsilon},\epsilon\in(0,\lambda_1)$. We claim that [5, Lemma 2.2] holds for $\alpha(\epsilon)$ with some constant $\epsilon\in(0,\lambda_1)$ replacing $\alpha(T)$. To this end, it is sufficient to prove [5, (14)] for $\alpha(\epsilon)$ with some constant $\epsilon\in(0,\lambda_1)$ instead of $\alpha(T)$, i.e.

$$\sup_{s\in[0,t]}\|X_s-Y_s\|_\infty^2\le\mathrm{e}^{(\lambda_1-\epsilon)r}\frac{\|\sigma\|^2}{\epsilon}\int_0^t|h(s)|^2\mathrm{d}s,\quad t\ge0. \tag{3.5}$$

In fact, it follows from Itô's formula and **(C)** that

$$\mathrm{d}|X(t)-Y(t)|^2\le\frac{\|\sigma\|^2}{\epsilon}|h(t)|^2\mathrm{d}t+(\epsilon-\lambda_1)|X(t)-Y(t)|^2\mathrm{d}t+\lambda_2\|X_t-Y_t\|_\infty^2\mathrm{d}t,\quad\epsilon\in(0,\lambda_1).$$

So, we get

$$\mathrm{d}[\mathrm{e}^{(\lambda_1-\epsilon)t}|X(t)-Y(t)|^2]\le\mathrm{e}^{(\lambda_1-\epsilon)t}\frac{\|\sigma\|^2}{\epsilon}|h(t)|^2\mathrm{d}t+\mathrm{e}^{(\lambda_1-\epsilon)t}\lambda_2\|X_t-Y_t\|_\infty^2\mathrm{d}t.$$

Let $\eta_t=\sup_{s\in[0,t]}\mathrm{e}^{(\lambda_1-\epsilon)s}|X(s)-Y(s)|^2$. It follows from $X_0=Y_0$ that

$$\eta_t\le\int_0^t\mathrm{e}^{(\lambda_1-\epsilon)s}\frac{\|\sigma\|^2}{\epsilon}|h(s)|^2\mathrm{d}s+\lambda_2\mathrm{e}^{(\lambda_1-\epsilon)r}\int_0^t\eta_s\mathrm{d}s.$$

Gronwall's inequality implies that

$$\begin{aligned}\eta_t&\le\int_0^t\exp\{\lambda_2\mathrm{e}^{(\lambda_1-\epsilon)r}(t-s)\}\mathrm{e}^{(\lambda_1-\epsilon)s}\frac{\|\sigma\|^2}{\epsilon}|h(s)|^2\mathrm{d}s\\&=\int_0^t\exp\{\lambda_2\mathrm{e}^{(\lambda_1-\epsilon)r}t\}\mathrm{e}^{[(\lambda_1-\epsilon)-\lambda_2\mathrm{e}^{(\lambda_1-\epsilon)r}]s}\frac{\|\sigma\|^2}{\epsilon}|h(s)|^2\mathrm{d}s.\end{aligned}$$

Noting that $\eta_t\ge\mathrm{e}^{(\lambda_1-\epsilon)(t-r)}\|X_t-Y_t\|_\infty^2$, we arrive at

$$\|X_t-Y_t\|_\infty^2\le\mathrm{e}^{(\lambda_1-\epsilon)r}\frac{\|\sigma\|^2}{\epsilon}\int_0^t\mathrm{e}^{-\mathrm{e}^{(\lambda_1-\epsilon)r}[(\lambda_1-\epsilon)\mathrm{e}^{-(\lambda_1-\epsilon)r}-\lambda_2](t-s)}|h(s)|^2\mathrm{d}s.$$

Since $\lambda_2<\sup_{\delta\in[0,\lambda_1]}\delta\mathrm{e}^{-\delta r}$ and $\delta\to\delta\mathrm{e}^{-\delta r}$ is a continuous function, there exists a constant $\epsilon\in(0,\lambda_1)$ such that $(\lambda_1-\epsilon)\mathrm{e}^{-(\lambda_1-\epsilon)r}-\lambda_2>0$. In the following, we fix this ϵ. We derive

$$\|X_t-Y_t\|_\infty^2\le\mathrm{e}^{(\lambda_1-\epsilon)r}\frac{\|\sigma\|^2}{\epsilon}\int_0^t|h(s)|^2\mathrm{d}s,\quad t\ge0,$$

which gives (3.5). So, [5, Lemma 2.2] holds for $\alpha(\epsilon)$ replacing $\alpha(T)$. Therefore, by [5, (7)] with $c_\mu = 0$, the transportation cost inequality for Π_ξ^T holds, i.e.

$$\mathbb{W}_{2,\rho_\infty^T}(\nu^T, \Pi_\xi^T)^2 \le 2\mathrm{e}^{(\lambda_1-\epsilon)r}\frac{\|\sigma\|^2}{\epsilon}\mathrm{Ent}(\nu^T|\Pi_\xi^T) \tag{3.6}$$

for any probability measure ν^T on $C([0,T];\mathscr{C}^{m+d})$ with $\nu^T(\sup_{t\in[0,T]}\|v_t\|_\infty^2) < \infty$.

Define the projection mapping $\pi_T : C([0,T];\mathscr{C}^{m+d}) \to \mathscr{C}^{m+d}$ as $\pi_T(v) = v_T, v \in C([0,T];\mathscr{C}^{m+d})$. Then by (3.6) and [10, Lemma 2.1] for $\Phi = \pi_T$, we obtain

$$\mathbb{W}_2(\nu, P_T^{\mu^*}(\xi,\cdot))^2 \le 2\mathrm{e}^{(\lambda_1-\epsilon)r}\frac{\|\sigma\|^2}{\epsilon}\mathrm{Ent}(\nu|P_T^{\mu^*}(\xi,\cdot)), \quad \nu \in \mathscr{P}_2(\mathscr{C}^{m+d}).$$

Finally, in view of (3.4) and [10, Lemma 2.2], we complete the proof. □

4 Propagation of Chaos

In this section, we consider path-distribution dependent SHS on $\mathbb{R}^{m+d}$:

$$\mathrm{d}X(t) = \begin{pmatrix} b(t, X_t, \mathscr{L}_{X_t}) \\ B(t, X_t, \mathscr{L}_{X_t}) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} 0_{m\times d} \\ \sigma(t, X_t, \mathscr{L}_{X_t}) \end{pmatrix}\mathrm{d}W(t), \tag{4.1}$$

where $W = (W(t))_{t\ge0}$ is a d-dimensional standard Brownian motion with respect to a complete filtration probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\ge0}, \mathbb{P})$, $b : [0,\infty)\times\mathscr{C}^{m+d}\times\mathscr{P}(\mathscr{C}^{m+d}) \to \mathbb{R}^m$, $B : [0,\infty)\times\mathscr{C}^{m+d}\times\mathscr{P}(\mathscr{C}^{m+d}) \to \mathbb{R}^d$ and $\sigma : [0,\infty)\times\mathscr{C}^{m+d}\times\mathscr{P}(\mathscr{C}^{m+d}) \to \mathbb{R}^d\otimes\mathbb{R}^d$ are measurable. Throughout this section, we fix $T > 0$ and consider the solution for (4.1) on time interval $[0,T]$.

Let X_0 be an $\mathscr{F}_0$-measurable $\mathscr{C}^{m+d}$-valued random variable, $N \ge 1$ be an integer and $(X_0^i, W^i(t))_{1\le i\le N}$ be i.i.d. copies of $(X_0, W(t))$. Consider the following non-interacting particle system:

$$\mathrm{d}X^i(t) = \begin{pmatrix} b(t, X_t^i, \mu_t^i) \\ B(t, X_t^i, \mu_t^i) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} 0_{m\times d} \\ \sigma(t, X_t^i, \mu_t^i) \end{pmatrix}\mathrm{d}W^i(t), \quad 1 \le i \le N, \tag{4.2}$$

where $\mu_t^i := \mathscr{L}_{X_t^i}$, and the mean field interacting particle system

$$\mathrm{d}X^{i,N}(t) = \begin{pmatrix} b(t, X_t^{i,N}, \hat\mu_t^N) \\ B(t, X_t^{i,N}, \hat\mu_t^N) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} 0_{m\times d} \\ \sigma(t, X_t^{i,N}, \hat\mu_t^N) \end{pmatrix}\mathrm{d}W^i(t), \quad X_0^{i,N} = X_0^i, \tag{4.3}$$

where $\hat\mu_t^N$ is the empirical distribution of $X_t^{1,N}, \cdots, X_t^{N,N}$, i.e.

$$\hat\mu_t^N = \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}.$$

To obtain the propagation of chaos, we make the following assumptions.

(H) There exist constants $K > 0$ and $\theta \geq 1$ such that the following conditions hold for all $t \in [0, T]$ and $\gamma \in \mathscr{P}_\theta(\mathscr{C}^{m+d})$:

(H1) For any $\xi, \eta \in \mathscr{C}^{m+d}$,

$$|b(t,\xi,\gamma) - b(t,\eta,\gamma)| + |B(t,\xi,\gamma) - B(t,\eta,\gamma)| + \|\sigma(t,\xi,\gamma) - \sigma(t,\eta,\gamma)\| \leq K\|\xi - \eta\|_\infty.$$

(H2) For any $\xi \in \mathscr{C}^{m+d}$ and $\bar{\gamma}, \tilde{\gamma} \in \mathscr{P}_\theta(\mathscr{C}^{m+d})$,

$$|b(t,\xi,\bar{\gamma}) - b(t,\xi,\tilde{\gamma})| + \|\sigma(t,\xi,\bar{\gamma}) - \sigma(t,\xi,\tilde{\gamma})\| + |B(t,\xi,\bar{\gamma}) - B(t,\xi,\tilde{\gamma})| \leq K\mathbb{W}_\theta(\bar{\gamma},\tilde{\gamma}),$$
$$|b(t,0,\delta_0)| + |B(t,0,\delta_0)| + |\sigma(t,0,\delta_0)| \leq K.$$

Under **(H)**, the well-posedness in $\mathscr{P}_\theta(\mathscr{C}^{m+d})$ for (4.1) holds due to Remark 5.2 below, which means that μ_t^i in (4.2) does not depend on i and we denote $\mu_t = \mu_t^i, t \in [0, T]$. Moreover, by Theorem 5.3 below, (4.3) is also well-posed.

To prove the propagation of chaos, we need the following lemma, which may be a known result. Since we have not found some references, we give a brief proof in the following.

Lemma 4.1. *Let $\{Z_i\}_{i\geq 1}$ be a sequence of i.i.d. non-negative random variables with $\mathbb{E}(Z_1) < \infty$. Then $\{\frac{1}{N}\sum_{i=1}^N Z_i\}_{N\geq 1}$ is uniformly integrable.*

Proof. Since $\mathbb{E}(Z_1) < \infty$, it follows from the strong law of large number that $\mathbb{P}$-a.s.

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N Z_i = \mathbb{E}(Z_1),$$

which yields $\mathbb{P}$-a.s.

$$\sup_{N\geq 1}\left\{\frac{1}{N}\sum_{i=1}^N Z_i\right\} < \infty.$$

This together with the fact that $\{Z_i\}_{i\geq 1}$ are i.i.d., $\mathbb{E}(Z_1) < \infty$ and the dominated convergence theorem yields that

$$\begin{aligned}
&\lim_{M\to\infty}\sup_{N\geq 1}\mathbb{E}\left\{\left(\frac{1}{N}\sum_{i=1}^N Z_i\right)1_{\left\{\frac{1}{N}\sum_{i=1}^N Z_i \geq M\right\}}\right\} \\
&= \lim_{M\to\infty}\sup_{N\geq 1}\mathbb{E}\left\{Z_1 1_{\left\{\frac{1}{N}\sum_{i=1}^N Z_i \geq M\right\}}\right\} \\
&\leq \lim_{M\to\infty}\mathbb{E}\left\{Z_1 1_{\left\{\sup_{N\geq 1}\{\frac{1}{N}\sum_{i=1}^N Z_i\} \geq M\right\}}\right\} = 0.
\end{aligned}$$

So, we complete the proof. □

To derive the quantitative propagation of chaos, we introduce the projection mappings

$$\pi(s)(\xi) = \xi(s), \quad s \in [-r, 0], \xi \in \mathscr{C}^{m+d}$$

and define $\mu^s = \mu \circ \pi(s)^{-1}, \mu \in \mathscr{P}(\mathscr{C}^{m+d})$. Then for any $\mu \in \mathscr{P}(\mathscr{C}^{m+d}), s \in [-r, 0]$, μ^s is a probability measure on $\mathbb{R}^{m+d}$. Let $\mathbb{W}_\theta^0$ be the L^θ-Wasserstein distance on $\mathscr{P}_\theta(\mathbb{R}^{m+d})$, the collection of all probability measures with finite θ-th moment on $\mathbb{R}^{m+d}$. Let Γ be a probability measure on $[-r, 0]$ and define

$$\mathbb{W}_\theta^\Gamma(\gamma, \bar{\gamma}) := \int_{-r}^0 \mathbb{W}_\theta^0(\gamma^s, \bar{\gamma}^s)\Gamma(\mathrm{d}s), \quad \gamma, \bar{\gamma} \in \mathscr{P}_\theta(\mathscr{C}^{m+d}). \tag{4.4}$$

Noting that for any $\gamma, \bar{\gamma} \in \mathscr{P}_\theta(\mathscr{C}^{m+d})$, it holds

$$\begin{aligned} |\mathbb{W}_\theta^0(\gamma^t, \bar{\gamma}^t) - \mathbb{W}_\theta^0(\gamma^s, \bar{\gamma}^s)| &\le |\mathbb{W}_\theta^0(\gamma^t, \bar{\gamma}^t) - \mathbb{W}_\theta^0(\gamma^t, \bar{\gamma}^s)| + |\mathbb{W}_\theta^0(\gamma^t, \bar{\gamma}^s) - \mathbb{W}_\theta^0(\gamma^s, \bar{\gamma}^s)| \\ &\le \mathbb{W}_\theta^0(\bar{\gamma}^t, \bar{\gamma}^s) + \mathbb{W}_\theta^0(\gamma^t, \gamma^s), \quad s, t \in [-r, 0]. \end{aligned}$$

So, $\mathbb{W}_\theta^0(\gamma^s, \bar{\gamma}^s)$ is continuous in s and the right hand side of (4.4) is well-defined. Moreover, it is clear that

$$\mathbb{W}_\theta^\Gamma(\gamma_1, \gamma_2) \le \mathbb{W}_\theta(\gamma_1, \gamma_2), \quad \gamma_1, \gamma_2 \in \mathscr{P}_\theta(\mathscr{C}^{m+d}). \tag{4.5}$$

In particular, when $\Gamma = \delta_0$, $\mathbb{W}_\theta^\Gamma(\gamma, \bar{\gamma}) = \mathbb{W}_\theta^0(\gamma^0, \bar{\gamma}^0)$. The main result in this section is as follows.

Theorem 4.2. *Assume* **(H)** *and* $\mathbb{E}\|X_0^i\|_\infty^\theta < \infty$. *Then the following assertions hold.*

(1) It holds

$$\lim_{N\to\infty} \mathbb{E} \sup_{t\in[0,T]} |X^i(t) - X^{i,N}(t)|^\theta = 0. \tag{4.6}$$

Consequently,

$$\lim_{N\to\infty} \mathbb{E} \sup_{t\in[0,T]} \mathbb{W}_\theta(\hat{\mu}_t^N, \mu_t)^\theta = 0. \tag{4.7}$$

If in addition, $b(t, \xi, \gamma)$ and $\sigma(t, \xi, \gamma)$ do not depend on γ and there exists a constant $K_0 > 0$ such that

$$\begin{aligned} &|B(t, \xi, \gamma) - B(t, \xi, \tilde{\gamma})| \le K_0[\mathbb{W}_\theta(\gamma, \tilde{\gamma}) \wedge 1], \\ &\|\sigma(t, \xi)^{-1}\| < K_0, \quad (t, \xi) \in [0, T] \times \mathscr{C}^{m+d}, \gamma, \tilde{\gamma} \in \mathscr{P}_\theta(\mathscr{C}^{m+d}), \end{aligned} \tag{4.8}$$

then for any $k \ge 1$,

$$\begin{aligned} &\lim_{N\to\infty} \sup_{t\in[0,T]} \|\mathscr{L}_{(X_t^{1,N}, X_t^{2,N}, \cdots, X_t^{k,N})} - \mu_t^{\otimes k}\|_{var}^2 \\ &\qquad \le 2 \lim_{N\to\infty} \sup_{t\in[0,T]} \mathrm{Ent}\left(\mathscr{L}_{(X_t^{1,N}, X_t^{2,N}, \cdots, X_t^{k,N})} | \mu_t^{\otimes k}\right) = 0, \end{aligned} \tag{4.9}$$

where $\mu_t^{\otimes k} = \prod_{i=1}^k \mu_t$, the k-independent product of μ_t.

(2) Assume that one of the following conditions holds:

(i) $\theta \geq 2$,

(ii) $\theta \in [1,2)$ *and* $\sigma(t,\xi,\gamma)$ *does not depend on* γ.

If $\mathbb{E}\|X_0^i\|_\infty^q < \infty$ *for some* $q > \theta$ *and there exists a probability measure* Γ *on* $[-r,0]$ *such that (H2) holds for* $\mathbb{W}_\theta^\Gamma$ *replacing* $\mathbb{W}_\theta$, *then there exists a constant* $C > 0$ *depending only on* $\theta, q, m+d, T$ *and* $\mathbb{E}\|X_0^i\|_\infty^q$ *such that*

$$\mathbb{E}\sup_{t\in[0,T]}|X^i(t)-X^{i,N}(t)|^\theta \leq CR_{m+d}(N), \tag{4.10}$$

where

$$R_{m+d}(N) = \begin{cases} N^{-\frac{1}{2}} + N^{-\frac{q-\theta}{q}}, & \theta > \frac{m+d}{2}, q \neq 2\theta, \\ N^{-\frac{1}{2}}\log(1+N) + N^{-\frac{q-\theta}{q}}, & \theta = \frac{m+d}{2}, q \neq 2\theta, \\ N^{-\frac{\theta}{m+d}} + N^{-\frac{q-\theta}{q}}, & \theta \in [1, \frac{m+d}{2}), q \neq \frac{m+d}{m+d-\theta}, \end{cases}$$

and consequently

$$\sup_{t\in[0,T]}\mathbb{E}\mathbb{W}_\theta^\Gamma(\hat{\mu}_t^N, \mu_t)^\theta \leq CR_{m+d}(N). \tag{4.11}$$

If in addition, $b(t,\xi,\gamma)$ *and* $\sigma(t,\xi,\gamma)$ *do not depend on* γ *and* (4.8) *holds for* $\mathbb{W}_\theta^\Gamma$ *replacing* $\mathbb{W}_\theta$, *then there exists a constant* $C > 0$ *depending on* $\theta, q, m+d, T$ *and* $\mathbb{E}\|X_0^i\|_\infty^q$ *such that for any* $k \geq 1$,

$$\begin{aligned} &\sup_{t\in[0,T]}\|\mathscr{L}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{k,N})} - \mu_t^{\otimes k}\|_{var}^2 \\ &\leq 2\sup_{t\in[0,T]}\mathrm{Ent}\left(\mathscr{L}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{k,N})}|\mu_t^{\otimes k}\right) \\ &\leq CkR_{m+d}(N)1_{\{\theta\in[1,2)\}} + CkR_{m+d}(N)^{\frac{2}{\theta}}1_{\{\theta\geq 2\}}. \end{aligned} \tag{4.12}$$

Proof. (1) If $\mathbb{E}\|X_0^i\|_\infty^p < \infty$ for some $p \geq \theta$, it is standard to derive from **(H)** that

$$\mathbb{E}\sup_{t\in[0,T]}\|X_t^i\|_\infty^p < C_0(1+\mathbb{E}(\|X_0^i\|_\infty^p)) \tag{4.13}$$

for some constant $C_0 > 0$. Let $\eta^{i,N}(t) = \sup_{s\in[0,t]}|X^{i,N}(s) - X^i(s)|$. Applying the BDG inequality and Hölder's inequality, we derive from **(H)** that

$$\begin{aligned} \mathbb{E}\eta^{i,N}(t)^\theta &\leq c_0\int_0^t \mathbb{E}(\eta^{i,N}(s)^\theta + \mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s)^\theta)\mathrm{d}s \\ &\quad + c_0\mathbb{E}\left(\int_0^t(\eta^{i,N}(s) + \mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s))^2\mathrm{d}s\right)^{\frac{\theta}{2}} \end{aligned} \tag{4.14}$$

for some constant $c_0 > 0$. Let $\tilde{\mu}_t^N = \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}$. Noting that

$$\mathbb{W}_\theta(\hat{\mu}_s^N, \tilde{\mu}_s^N) \le \left(\frac{1}{N}\sum_{i=1}^N \|X_s^{i,N} - X_s^i\|_\infty^\theta\right)^{\frac{1}{\theta}}, \tag{4.15}$$

we obtain

$$\begin{aligned}\mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s) &\le \mathbb{W}_\theta(\hat{\mu}_s^N, \tilde{\mu}_s^N) + \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)\\ &\le \left(\frac{1}{N}\sum_{i=1}^N \|X_s^{i,N} - X_s^i\|_\infty^\theta\right)^{\frac{1}{\theta}} + \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s).\end{aligned} \tag{4.16}$$

Next, we divide into two cases: $\theta \ge 2$ and $\theta \in [1, 2)$ to estimate the second term on the right hand side of (4.14).

If $\theta \ge 2$, by Hölder's inequality, we have

$$c_0\mathbb{E}\left(\int_0^t (\eta^{i,N}(s) + \mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s))^2 \mathrm{d}s\right)^{\frac{\theta}{2}} \le c_1\int_0^t \mathbb{E}\eta^{i,N}(s)^\theta \mathrm{d}s + c_1\mathbb{E}\int_0^t \mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s)^\theta \mathrm{d}s$$

for some constant $c_1 > 0$. This together with (4.14) and (4.16) implies that there exists a constant $c_2 > 0$ such that

$$\mathbb{E}\eta^{i,N}(t)^\theta \le c_2\int_0^t \mathbb{E}\eta^{i,N}(s)^\theta \mathrm{d}s + c_2\mathbb{E}\int_0^t \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^\theta \mathrm{d}s.$$

(4.13) for $p = \theta$ and Gronwall's inequality give

$$\mathbb{E}\eta^{i,N}(t)^\theta \le c_3\mathbb{E}\int_0^t \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^\theta \mathrm{d}s \tag{4.17}$$

for some constant $c_3 > 0$.

If $\theta \in [1, 2)$, it follows from (4.16), the inequality $\sqrt{|ab|} \le \frac{|a|+|b|}{2}$ and Hölder's inequality that

$$\begin{aligned}&c_0\mathbb{E}\left(\int_0^t (\eta^{i,N}(s) + \mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s))^2 \mathrm{d}s\right)^{\frac{\theta}{2}}\\ &\le c_0\mathbb{E}\left(\int_0^t \left(\eta^{i,N}(s) + \left(\frac{1}{N}\sum_{i=1}^N \|X_s^{i,N} - X_s^i\|_\infty^\theta\right)^{\frac{1}{\theta}} + \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)\right)^2 \mathrm{d}s\right)^{\frac{\theta}{2}}\\ &\le c_1'\int_0^t \mathbb{E}\eta^{i,N}(s)^\theta \mathrm{d}s + \frac{1}{2}\mathbb{E}\eta^{i,N}(t)^\theta + c_1'\mathbb{E}\left(\int_0^t \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^2 \mathrm{d}s\right)^{\frac{\theta}{2}}\end{aligned} \tag{4.18}$$

for some constant $c_1' > 0$. So, this combined with (4.14) and (4.16) derives

$$\begin{aligned}\mathbb{E}\eta^{i,N}(t)^\theta \le c_2' \int_0^t \mathbb{E}\eta^{i,N}(s)^\theta \mathrm{d}s + c_2' \mathbb{E}\int_0^t \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^\theta \mathrm{d}s \\ + c_2' \mathbb{E}\left(\int_0^t \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^2 \mathrm{d}s\right)^{\frac{\theta}{2}}\end{aligned} \tag{4.19}$$

for some constant $c_2' > 0$. Therefore, using Grönwall's inequality for (4.19), there exists a constant $c_3' > 0$ such that

$$\mathbb{E}\eta^{i,N}(t)^\theta \le c_3' \mathbb{E}\int_0^t \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^\theta \mathrm{d}s + c_3' \mathbb{E}\left(\int_0^t \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^2 \mathrm{d}s\right)^{\frac{\theta}{2}}. \tag{4.20}$$

Let $\mathscr{C}_T^{m+d} = C([-r, T]; \mathbb{R}^{m+d})$ be equipped with the uniform norm and $\mathscr{P}(\mathscr{C}_T^{m+d})$ be the set of the probability measures on $\mathscr{C}_T^{m+d}$. Define

$$\mathscr{P}_\theta(\mathscr{C}_T^{m+d}) = \left\{\mu^T \in \mathscr{P}(\mathscr{C}_T^{m+d}) : \int_{\mathscr{C}_T^{m+d}} \sup_{s\in[-r,T]} |\xi(s)|^\theta \mu^T(\mathrm{d}\xi) < \infty\right\}$$

and denote $\mathbb{W}_{\theta,T}$ as the L^θ-Wasserstein distance on $\mathscr{P}_\theta(\mathscr{C}_T^{m+d})$. So, $(\mathscr{P}_\theta(\mathscr{C}_T^{m+d}), \mathbb{W}_{\theta,T})$ is a Polish space.

Next, by the triangle inequality, we arrive at

$$\begin{aligned}\sup_{s\in[0,T]} \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s) &\le \mathbb{W}_{\theta,T}\left(\frac{1}{N}\sum_{i=1}^N \delta_{X^i([-r,T])}, \mathscr{L}_{X^i([-r,T])}\right) \\ &\le \mathbb{W}_{\theta,T}\left(\frac{1}{N}\sum_{i=1}^N \delta_{X^i([-r,T])}, \delta_0\right) + \mathbb{W}_{\theta,T}\left(\delta_0, \mathscr{L}_{X^i([-r,T])}\right) \\ &\le \left(\frac{1}{N}\sum_{i=1}^N \sup_{s\in[0,T]} \|X_s^i\|_\infty^\theta\right)^{\frac{1}{\theta}} + \left(\mathbb{E}\sup_{s\in[0,T]} \|X_s^i\|_\infty^\theta\right)^{\frac{1}{\theta}}.\end{aligned} \tag{4.21}$$

Thanks to the generalized Glivenko-Cantelli-Varadarajan theorem, see for instance [29, Corollary 12.2.2], it holds $\mathbb{P}$-a.s.

$$\lim_{N\to\infty} \mathbb{W}_{\theta,T}\left(\frac{1}{N}\sum_{i=1}^N \delta_{X^i([-r,T])}, \mathscr{L}_{X^i([-r,T])}\right) = 0. \tag{4.22}$$

Therefore, it follows from (4.13) for $p = \theta$, (4.21), (4.22), Lemma 4.1 for $Z_i = \sup_{s\in[0,T]} \|X_s^i\|_\infty^\theta$ and the dominated convergence theorem that

$$\lim_{N\to\infty} \mathbb{E}\sup_{s\in[0,T]} \mathbb{W}_\theta(\tilde{\mu}_s^N, \mu_s)^\theta$$

$$\leq \lim_{N\to\infty} \mathbb{E}\left[\mathbb{W}_{\theta,T}\left(\frac{1}{N}\sum_{i=1}^N \delta_{X^i([-r,T])}, \mathscr{L}_{X^i([-r,T])}\right)^\theta\right] = 0. \tag{4.23}$$

This together with (4.17) or (4.20) derives (4.6). Finally, by (4.6), (4.23) and (4.16), we get (4.7).

When $b(t,\xi,\gamma)$ and $\sigma(t,\xi,\gamma)$ do not depend on γ, we can rewrite (4.2) as

$$\mathrm{d}X^i(t) = \begin{pmatrix} b(t,X_t^i) \\ B(t,X_t^i,\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}) \end{pmatrix}\mathrm{d}t + \begin{pmatrix} 0_{m\times d} \\ \sigma(t,X_t^i)\end{pmatrix}\mathrm{d}\tilde{W}^i(t), \quad 1\leq i\leq N,$$

with

$$\mathrm{d}\tilde{W}^i(t) = \mathrm{d}W^i(t) - \tilde{\Gamma}^i(t)\mathrm{d}t, \quad 1\leq i\leq N$$

and

$$\tilde{\Gamma}^i(t) = \sigma(t,X_t^i)^{-1}[B(t,X_t^i,\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i}) - B(t,X_t^i,\mu_t)], \quad 1\leq i\leq N.$$

It follows from (4.8) that

$$|\tilde{\Gamma}^i(t)| \leq K_0^2(\mathbb{W}_\theta(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^i},\mu_t)\wedge 1), \quad t\in[0,T], 1\leq i\leq N. \tag{4.24}$$

Let

$$R_t = \exp\left\{\sum_{i=1}^N \int_0^t \langle\tilde{\Gamma}^i(s),\mathrm{d}W^i(s)\rangle - \frac{1}{2}\sum_{i=1}^N\int_0^t |\tilde{\Gamma}^i(s)|^2\mathrm{d}s\right\}, \quad t\in[0,T].$$

(4.24) and Girsanov's theorem imply that $\{R_t\}_{t\in[0,T]}$ is a martingale and $((\tilde{W}^i(t))_{1\leq i\leq N})_{t\in[0,T]}$ is an Nd-dimensional Brownian motion under $\mathbb{Q}_T = R_T\mathbb{P}$ and

$$\mathscr{L}_{(X_t^1,X_t^2,\cdots,X_t^N)}|\mathbb{Q}_T = \mathscr{L}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{N,N})}|\mathbb{P}, \quad t\in[0,T].$$

This implies that

$$\begin{aligned}\mathbb{E}[f(X_t^{1,N},X_t^{2,N},\cdots,X_t^{N,N})] &= \mathbb{E}[R_T f(X_t^1,X_t^2,\cdots,X_t^N)] \\ &= \mathbb{E}[R_t f(X_t^1,X_t^2,\cdots,X_t^N)], \quad f\in\mathscr{B}_b(\mathscr{C}^{N(m+d)}), t\in[0,T].\end{aligned}$$

So, there exists a constant $C>0$ such that

$$\begin{aligned}&\mathrm{Ent}(\mathscr{L}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{N,N})}|\mathbb{P}|\mu_t^{\otimes N}) \\ &\leq \mathbb{E}(R_t\log R_t) = \frac{1}{2}\sum_{i=1}^N\int_0^t \mathbb{E}^{\mathbb{Q}_T}|\tilde{\Gamma}^i(s)|^2\mathrm{d}s \\ &\leq C^2N\int_0^t \mathbb{E}^{\mathbb{Q}_T}(\mathbb{W}_\theta(\frac{1}{N}\sum_{i=1}^N\delta_{X_s^i},\mu_s)\wedge 1)^2\mathrm{d}s\end{aligned}$$

$$= C^2 N \int_0^t \mathbb{E}(\mathbb{W}_\theta(\frac{1}{N}\sum_{i=1}^N \delta_{X_s^{i,N}}, \mu_s) \wedge 1)^2 \mathrm{d}s$$
$$= C^2 N \int_0^t \mathbb{E}(\mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s) \wedge 1)^2 \mathrm{d}s, \quad t \in [0,T].$$

This together with [23, Lemma 3.9] implies that for any $k \geq 1$ and $N \geq k$,

$$\mathrm{Ent}(\mathscr{L}_{(X_t^{1,N}, X_t^{2,N}, \cdots, X_t^{k,N})} | \mu_t^{\otimes k}) \leq 2C^2 k \int_0^t \mathbb{E}(\mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s) \wedge 1)^2 \mathrm{d}s.$$

So, Pinsker's inequality (1.1) yields

$$\begin{aligned} \|\mathscr{L}_{(X_t^{1,N}, X_t^{2,N}, \cdots, X_t^{k,N})} - \mu_t^{\otimes k}\|_{var}^2 &\leq 2\mathrm{Ent}(\mathscr{L}_{(X_t^{1,N}, X_t^{2,N}, \cdots, X_t^{k,N})} | \mu_t^{\otimes k}) \\ &\leq 4C^2 k \int_0^t \mathbb{E}(\mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s) \wedge 1)^2 \mathrm{d}s. \end{aligned} \tag{4.25}$$

Note that

$$\mathbb{E}(\mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s) \wedge 1)^2 \leq \mathbb{E}(\mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s)^\theta) 1_{\{\theta \in [1,2)\}} + \big(\mathbb{E}\mathbb{W}_\theta(\hat{\mu}_s^N, \mu_s)^\theta\big)^{\frac{2}{\theta}} 1_{\{\theta \geq 2\}}. \tag{4.26}$$

By (4.7) and (4.25), we prove (4.9).

(2) Assume that (H2) holds for $\mathbb{W}_\theta^\Gamma$ replacing $\mathbb{W}_\theta$. When $\theta \geq 2$, repeating the proof to get (4.17), we derive

$$\mathbb{E}\eta^{i,N}(t)^\theta \leq c_4 \int_0^t \mathbb{E}\mathbb{W}_\theta^\Gamma(\tilde{\mu}_s^N, \mu_s)^\theta \mathrm{d}s \tag{4.27}$$

for some constant $c_4 > 0$. When $\theta \in [1,2)$ and $\sigma(t,\xi,\gamma)$ does not depend on γ, (4.18) is replaced by

$$c_0 \mathbb{E}\left(\int_0^t \eta^{i,N}(s)^2 \mathrm{d}s\right)^{\frac{\theta}{2}} \leq c_4' \int_0^t \mathbb{E}\eta^{i,N}(s)^\theta \mathrm{d}s + \frac{1}{2}\mathbb{E}\eta^{i,N}(t)^\theta$$

for some constant $c_4' > 0$. Then (4.27) instead of (4.20) holds. Next, by the definition of $\mathbb{W}_\theta^\Gamma$, we have

$$\mathbb{E}\mathbb{W}_\theta^\Gamma(\tilde{\mu}_s^N, \mu_s)^\theta \leq \int_{-r}^0 \mathbb{E}\left[\mathbb{W}_\theta^0\left(\frac{1}{N}\sum_{i=1}^N \delta_{X_s^i(u)}, \mathscr{L}_{X_s^i(u)}\right)^\theta\right] \Gamma(\mathrm{d}u). \tag{4.28}$$

Note that $\sup_{t\in[0,T]} \mu_t(\|\cdot\|_\infty^q) < \infty$ due to (4.13) for $p = q$. By [12, Theorem 1] for $p = \theta, q = q$, see [8, Theorem 5.8] for the special case $\theta = 2, q > 4$, there exists a constant $C_0 > 0$ depending only on $\theta, q, m + d$ such that

$$\mathbb{E}\left[\mathbb{W}_\theta^0\left(\frac{1}{N}\sum_{i=1}^N \delta_{X_s^i(u)}, \mathscr{L}_{X_s^i(u)}\right)^\theta\right]$$

$$\leq C_0 \left(\sup_{t\in[0,T]} \mu_t(\|\cdot\|_\infty^q) \right)^{\frac{\theta}{q}} R_{m+d}(N), \quad s \in [0,T], u \in [-r,0].$$

Substituting this into (4.28), we derive from (4.13) for $p = q$ that there exists a constant $C > 0$ depending only on $\theta, q, m+d, T$ and $\mathbb{E}\|X_0^i\|_\infty^q$ such that

$$\sup_{s\in[0,T]} \mathbb{E}\mathbb{W}_\theta^\Gamma(\tilde{\mu}_s^N, \mu_s)^\theta \leq C_0 \left(\sup_{t\in[0,T]} \mu_t(\|\cdot\|_\infty^q) \right)^{\frac{\theta}{q}} R_{m+d}(N) \leq C R_{m+d}(N). \tag{4.29}$$

So, (4.10) follows from (4.27) and (4.29). Moreover, it follows from (4.5) and (4.15) that

$$\begin{aligned}
\mathbb{W}_\theta^\Gamma(\hat{\mu}_s^N, \mu_s)^\theta &\leq 2^{\theta-1}\mathbb{W}_\theta^\Gamma(\hat{\mu}_s^N, \tilde{\mu}_s^N)^\theta + 2^{\theta-1}\mathbb{W}_\theta^\Gamma(\tilde{\mu}_s^N, \mu_s)^\theta \\
&\leq 2^{\theta-1}\mathbb{W}_\theta(\hat{\mu}_s^N, \tilde{\mu}_s^N)^\theta + 2^{\theta-1}\mathbb{W}_\theta^\Gamma(\tilde{\mu}_s^N, \mu_s)^\theta \\
&\leq 2^{\theta-1}\frac{1}{N}\sum_{i=1}^N \|X_s^{i,N} - X_s^i\|_\infty^\theta + 2^{\theta-1}\mathbb{W}_\theta^\Gamma(\tilde{\mu}_s^N, \mu_s)^\theta,
\end{aligned}$$

which implies (4.11) due to (4.10) and (4.29).

Finally, if $b(t,\xi,\gamma)$ and $\sigma(t,\xi,\gamma)$ do not depend on γ and (4.8) holds for $\mathbb{W}_\theta^\Gamma$ replacing $\mathbb{W}_\theta$, then (4.25) holds for $\mathbb{W}_\theta^\Gamma$ replacing $\mathbb{W}_\theta$. Moreover, by (4.26) for $\mathbb{W}_\theta^\Gamma$ replacing $\mathbb{W}_\theta$ and (4.11), we derive (4.12) and the proof is completed. □

5 Appendix

In this section, we give the well-posedness of general path-distribution dependent SDEs as well as mean field interacting particle system, and then apply it to the path-distribution dependent SHS. Fix $T > 0$. Let $n, k \in \mathbb{N}^+$ and $\theta \geq 1$. Consider path-distribution dependent SDEs on $\mathbb{R}^n$:

$$\mathrm{d}X(t) = H(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \Sigma(t, X_t, \mathscr{L}_{X_t})\mathrm{d}W(t), \quad t \in [0,T]. \tag{5.1}$$

where $H : [0,T] \times \mathscr{C}^n \times \mathscr{P}(\mathscr{C}^n) \to \mathbb{R}^n$, $\Sigma : [0,T] \times \mathscr{C}^n \times \mathscr{P}(\mathscr{C}^n) \to \mathbb{R}^n \otimes \mathbb{R}^k$ are measurable and $W(t)$ is a k-dimensional Brownian motion on some complete filtration probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$. Let $\hat{\mathscr{P}}(\mathscr{C}^n)$ be a subset of $\mathscr{P}(\mathscr{C}^n)$ and it is equipped with some topology.

Definition 5.1. The SDE (5.1) is called well-posed for distributions in $\hat{\mathscr{P}}(\mathscr{C}^n)$, if for any $\mathscr{F}_0$-measurable initial value X_0 with $\mathscr{L}_{X_0} \in \hat{\mathscr{P}}(\mathscr{C}^n)$ (respectively any initial distribution $\gamma \in \hat{\mathscr{P}}(\mathscr{C}^n)$), it has a unique strong solution (respectively weak solution) such that $\mathscr{L}_{X_\cdot} \in C([0,T]; \hat{\mathscr{P}}(\mathscr{C}^n))$, the space of continuous maps from $[0,T]$ to $\hat{\mathscr{P}}(\mathscr{C}^n)$. In particular, (5.1) is called well-posed for distributions in $\mathscr{P}_\theta(\mathscr{C}^n)$, if the above holds for $(\mathscr{P}_\theta(\mathscr{C}^n), \mathbb{W}_\theta)$ replacing $\hat{\mathscr{P}}(\mathscr{C}^n)$.

Theorem 5.1. *Assume that there exists some constant $K \geq 0$ such that*

$$\begin{aligned} &|H(s,\xi,\gamma_1)-H(s,\eta,\gamma_2)|+|\Sigma(s,\xi,\gamma_1)-\Sigma(s,\eta,\gamma_2)| \leq K(\|\xi-\eta\|_\infty + \mathbb{W}_\theta(\gamma_1,\gamma_2)),\\ &|H(s,0,\delta_0)|+|\Sigma(s,0,\delta_0)| \leq K, \quad s\in[0,T], \xi,\eta\in\mathscr{C}^n, \gamma_1,\gamma_2\in\mathscr{P}_\theta(\mathscr{C}^n). \end{aligned} \tag{5.2}$$

Then (5.1) *is strongly well-posed in $\mathscr{P}_\theta(\mathscr{C}^n)$ and there exists a constant $C>0$ such that*

$$\mathbb{W}_\theta(P_t^*\mu_0, P_t^*\nu_0) \leq C\mathrm{e}^{Ct}\mathbb{W}_\theta(\mu_0,\nu_0), \quad t\in[0,T], \mu_0,\nu_0\in\mathscr{P}_\theta(\mathscr{C}^n),$$

here $P_t^*\mu_0$ is the distribution of the solution to (5.1) *with initial distribution $\mu_0\in\mathscr{P}_\theta(\mathscr{C}^n)$.*

Proof. It follows from (5.2) that for any $\mu\in C([0,T],\mathscr{P}_\theta(\mathscr{C}^n))$, the classical SDE

$$\mathrm{d}X^\mu(t) = H(t,X_t^\mu,\mu_t)\mathrm{d}t + \Sigma(t,X_t^\mu,\mu_t)\mathrm{d}W(t), \quad t\in[0,T] \tag{5.3}$$

is well-posed. For any $\mathscr{F}_0$-measurable random variable X_0 with $\mathscr{L}_{X_0}\in\mathscr{P}_\theta(\mathscr{C}^n)$, let X_t^{μ,X_0} be the unique solution to (5.3) starting from X_0. Define the mapping $\Phi^{X_0}: C([0,T],\mathscr{P}_\theta(\mathscr{C}^n)) \to C([0,T],\mathscr{P}_\theta(\mathscr{C}^n))$ as

$$\Phi_t^{X_0}(\mu) = \mathscr{L}_{X_t^{\mu,X_0}}, \quad t\in[0,T].$$

By (5.2) and the inequality

$$(|a|+|b|+|c|)^\theta \leq 3^{\theta-1}(|a|^\theta+|b|^\theta+|c|^\theta),$$

we arrive at

$$\begin{aligned} |X^{\nu,\tilde{X}_0}(t)-X^{\mu,X_0}(t)|^\theta &\leq 3^{\theta-1}|\tilde{X}(0)-X(0)|^\theta\\ &\quad + 3^{\theta-1}\left|\int_0^t [H(s,X_s^{\nu,\tilde{X}_0},\nu_s)-H(s,X_s^{\mu,X_0},\mu_s)]\mathrm{d}s\right|^\theta\\ &\quad + 3^{\theta-1}\left|\int_0^t [\Sigma(s,X_s^{\nu,\tilde{X}_0},\nu_s)-\Sigma(s,X_s^{\mu,X_0},\mu_s)]\mathrm{d}W(s)\right|^\theta. \end{aligned} \tag{5.4}$$

Let $\xi_t = \sup_{s\in[-r,t]}|X^{\mu,X_0}(s)-X^{\nu,\tilde{X}_0}(s)|$. By (5.2), it follows from BDG's inequality, the inequality $\sqrt{|ab|}\leq\frac{|a|+|b|}{2}$ and Hölder's inequality that

$$\begin{aligned} &3^{\theta-1}\mathbb{E}\sup_{v\in[0,t]}\left|\int_0^v\{\Sigma(s,X_s^{\mu,X_0},\mu_s)-\Sigma(s,X_s^{\nu,\tilde{X}_0},\nu_s)\}\mathrm{d}W(s)\right|^\theta\\ &\leq C_0\mathbb{E}\left(\int_0^t(\xi_s^2+\mathbb{W}_\theta(\mu_s,\nu_s)^2)\mathrm{d}s\right)^{\frac{\theta}{2}}\\ &\leq \frac{1}{2}\mathbb{E}\xi_t^\theta + +C_1\int_0^t\mathbb{E}\xi_s^\theta\mathrm{d}s + C_1\left(\int_0^t\mathbb{W}_\theta(\mu_s,\nu_s)^2\mathrm{d}s\right)^{\frac{\theta}{2}} \end{aligned}$$

for some constant $C_1 > 0$. Again by (5.2) and Hölder's inequality, there exists a constant $C_2 > 0$ such that

$$3^{\theta-1}\mathbb{E}\sup_{v\in[0,t]}\left|\int_0^v [H(s, X_s^{\mu,X_0}, \mu_s) - H(s, X_s^{\nu,\tilde{X}_0}, \nu_s)]\mathrm{d}s\right|^\theta \le C_2\mathbb{E}\int_0^t (\xi_s^\theta + \mathbb{W}_\theta(\mu_s,\nu_s)^\theta)\mathrm{d}s.$$

As a result, we obtain from (5.4) and Hölder's inequality that

$$\begin{aligned}\mathbb{E}\xi_t^\theta &\le 2^{\theta-1}\mathbb{E}\|X_0 - \tilde{X}_0\|_\infty^\theta + 2^{\theta-1}\mathbb{E}\sup_{s\in[0,t]}|X^{\mu,X_0}(s) - X^{\nu,\tilde{X}_0}(s)|^\theta \\ &\le C_3\mathbb{E}\|X_0 - \tilde{X}_0\|_\infty^\theta + C_3\int_0^t \mathbb{E}\xi_s^\theta \mathrm{d}s + C_3\int_0^t \mathbb{W}_\theta(\mu_s,\nu_s)^\theta \mathrm{d}s + C_3\left(\int_0^t \mathbb{W}_\theta(\mu_s,\nu_s)^2\mathrm{d}s\right)^{\frac{\theta}{2}}\end{aligned}$$

for some constant $C_3 > 0$. So, Gronwall's inequality yields that there exists a constant $C_4 > 0$ such that

$$\begin{aligned}\mathbb{W}_\theta(\Phi_t^{X_0}(\mu), \Phi_t^{\tilde{X}_0}(\nu))^\theta \le \mathbb{E}\xi_t^\theta &\le C_4\mathbb{E}\|X_0 - \tilde{X}_0\|_\infty^\theta + C_4\int_0^t \mathbb{W}_\theta(\mu_s,\nu_s)^\theta \mathrm{d}s \\ &\quad + C_4\left(\int_0^t \mathbb{W}_\theta(\mu_s,\nu_s)^2\mathrm{d}s\right)^{\frac{\theta}{2}}, \quad t\in[0,T].\end{aligned} \tag{5.5}$$

Therefore, for any $\delta > 0$, we have

$$\sup_{t\in[0,T]} \mathrm{e}^{-\delta\theta t}\mathbb{W}_\theta(\Phi_t^{X_0}(\mu), \Phi_t^{X_0}(\nu))^\theta \le \sup_{t\in[0,T]} \mathrm{e}^{-\delta\theta t}\mathbb{W}_\theta(\mu_t,\nu_t)^\theta C_4[(\delta\theta)^{-1} + (2\delta)^{-\frac{\theta}{2}}].$$

Take δ_0 satisfying $\left(C_4[(\delta_0\theta)^{-1} + (2\delta_0)^{-\frac{\theta}{2}}]\right)^{\frac{1}{\theta}} < \frac{1}{2}$ and let $E^{X_0} := \{\mu \in C([0,T];\mathscr{P}_\theta(\mathscr{C}^n)) : \mu_0 = \mathscr{L}_{X_0}\}$ equipped with the complete metric

$$\rho_{\delta_0}(\nu,\mu) := \sup_{t\in[0,T]} \mathrm{e}^{-\delta_0 t}\mathbb{W}_\theta(\nu_t,\mu_t), \quad \mu,\nu\in E^{X_0}.$$

Then we conclude that

$$\rho_{\delta_0}(\Phi_\cdot^{X_0}(\mu), \Phi_\cdot^{X_0}(\nu)) < \frac{1}{2}\rho_{\delta_0}(\mu,\nu), \quad \mu,\nu\in E^{X_0},$$

and the Banach fixed point theorem yields that

$$\Phi_t^{X_0}(\mu) = \mu_t, \quad t\in[0,T]$$

has a unique solution $\mu \in E^{X_0}$. This means that (5.1) has a unique strong solution on $[0,T]$ with initial value X_0.

Next, applying (5.5) for $\mu_t = P_t^*\mu_0, \nu_t = P_t^*\nu_0$ and $X_0, \tilde{X}_0$ satisfying $\mathscr{L}_{X_0} = \mu_0, \mathscr{L}_{\tilde{X}_0} = \nu_0$ and noting that

$$C_4\left(\int_0^t \mathbb{W}_\theta(\mu_s,\nu_s)^2\mathrm{d}s\right)^{\frac{\theta}{2}} \le \frac{1}{2}\sup_{s\in[0,t]}\mathbb{W}_\theta(P_s^*\mu_0, P_s^*\nu_0)^\theta + C_5\left(\int_0^t \mathbb{W}_\theta(P_s^*\mu_0, P_s^*\nu_0)\mathrm{d}s\right)^\theta$$

$$\leq \frac{1}{2}\sup_{s\in[0,t]}\mathbb{W}_\theta(P_s^*\mu_0,P_s^*\nu_0)^\theta + C_6\int_0^t \mathbb{W}_\theta(P_s^*\mu_0,P_s^*\nu_0)^\theta \mathrm{d}s$$

for some constant $C_6>0$, there exists a constant $C_7>0$ such that

$$\sup_{s\in[0,t]}\mathbb{W}_\theta(P_s^*\mu_0,P_s^*\nu_0)^\theta \leq C_7\mathbb{E}\|X_0-\tilde{X}_0\|_\infty^\theta + C_7\int_0^t \mathbb{W}_\theta(P_s^*\mu_0,P_s^*\nu_0)^\theta \mathrm{d}s, \quad t\in[0,T].$$

So, by Grönwall inequality and taking infimum for all $X_0,\tilde{X}_0$ satisfying $\mathscr{L}_{X_0}=\mu_0, \mathscr{L}_{\tilde{X}_0}=\nu_0$, we complete the proof. □

Remark 5.2. *Under* **(A2)**-**(A3)**, *the assertions in Theorem 5.1 hold for* (2.7) *replacing* (5.1) *by applying Theorem 5.1 for* $n=m+d$, $k=d$ *and*

$$H(t,\xi,\gamma)=\begin{pmatrix} A\xi^{(1)}(0)+M\xi^{(2)}(0) \\ Z(\xi(0),\gamma)+B(\xi,\gamma)\end{pmatrix}, \quad \Sigma=\begin{pmatrix} 0_{m\times d}\\ \sigma\end{pmatrix}.$$

Similarly, under **(H)**, *the assertions in Theorem 5.1 hold for* (4.1) *replacing* (5.1).

Next, consider the mean field interacting particle system:

$$\mathrm{d}X^{i,N}(t)=H(t,X_t^{i,N},\hat{\mu}_t^N)+\Sigma(t,X_t^{i,N},\hat{\mu}_t^N)\mathrm{d}W^i(t), \quad 1\leq i\leq N, \tag{5.6}$$

with $\hat{\mu}_t^N=\frac{1}{N}\sum_{i=1}^N\delta_{X_t^{i,N}}$ and $(W^i)_{1\leq i\leq N}$ are independent k-dimensional standard Brownian motions. We give a result on the well-posedness of (5.6).

Theorem 5.3. *Under* (5.2), (5.6) *is well-posed.*

Proof. For any $\xi=\begin{pmatrix}\xi_1\\ \xi_2\\ \cdots\\ \xi_N\end{pmatrix}\in(\mathscr{C}^n)^N$, let $\mu_N^\xi=\frac{1}{N}\sum_{i=1}^N\delta_{\xi_i}$ and define

$$\tilde{H}(t,\xi)=\begin{pmatrix} H(t,\xi_1,\mu_N^\xi)\\ H(t,\xi_2,\mu_N^\xi)\\ \cdots\\ H(t,\xi_N,\mu_N^\xi)\end{pmatrix}, \tilde{\Sigma}(t,\xi)=\begin{pmatrix} \Sigma(t,\xi_1,\mu_N^\xi) & 0_{n\times k} & \cdots & 0_{n\times k}\\ 0_{n\times k} & \Sigma(t,\xi_2,\mu_N^\xi) & \cdots & 0_{n\times k}\\ \cdots & \cdots & \cdots & \cdots\\ 0_{n\times k} & 0_{n\times k} & \cdots & \Sigma(t,\xi_N,\mu_N^\xi)\end{pmatrix}.$$

Note that for $\xi,\eta\in(\mathscr{C}^n)^N$, it holds

$$\begin{aligned}\mathbb{W}_\theta(\mu_N^\xi,\mu_N^\eta)&=\mathbb{W}_\theta\left(\frac{1}{N}\sum_{i=1}^N\delta_{\xi_i},\frac{1}{N}\sum_{i=1}^N\delta_{\eta_i}\right)\\ &\leq\left(\frac{1}{N}\sum_{i=1}^N\|\xi_i-\eta_i\|_\infty^\theta\right)^{\frac{1}{\theta}}\leq c(\theta,N)\|\xi-\eta\|_\infty\end{aligned} \tag{5.7}$$

for some constant $c(\theta, N) > 0$. Consider path dependent SDE on $\mathbb{R}^{nN}$:

$$\mathrm{d}X(t) = \tilde{H}(t, X_t)\mathrm{d}t + \tilde{\Sigma}(t, X_t)\mathrm{d}W_N(t), \tag{5.8}$$

where $W_N = \begin{pmatrix} W^1 \\ W^2 \\ \cdots \\ W^N \end{pmatrix}$ is a kN-dimensional Brownian motion. By (5.2) and (5.7), we have

$$|\tilde{H}(t,\xi) - \tilde{H}(t,\eta)| + \|\tilde{\Sigma}(t,\xi) - \tilde{\Sigma}(t,\eta)\| \le C\|\xi - \eta\|_\infty, \quad \xi, \eta \in (\mathscr{C}^n)^N.$$

So, it is standard that (5.8) is well-posed and so is (5.6). □

References

[1] M. Arnaudon, A. Thalmaier, F.-Y. Wang, *Gradient estimates and Harnack inequalities on non-compact Riemannian manifolds,* Stochastic Process. Appl. 119(2009), 3653-3670.

[2] G. B. Arous, O. Zeitouni, *Increasing propagation of chaos for mean field models,* Ann. Inst. Henri Poincaré Probab. Stat. 35(1999), 85-102.

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

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

[5] J. Bao, F.-Y. Wang, C. Yuan, *Transportation Cost Inequalities for Neutral Functional Stochastic Equations,* Z. Anal. Anwend. 32(2013), 457-475.

[6] J. Bao, F.-Y. Wang, C. Yuan, *Derivative formula and Harnack inequality for degenerate functionals SDEs,* Stoch. Dyn. 13(2013), 943-951.

[7] R. J. Berman, M. Önnheim, *Propagation of Chaos for a Class of First Order Models with Singular Mean Field Interactions,* SIAM J. Math. Anal. 51(2019), 159-196.

[8] R. Carmona, F. Delarue, *Probabilistic theory of mean field games with applications. I. Mean field FBSDEs, control, and games. Probability Theory and Stochastic Modelling,* 83(2018), Springer.

[9] P.-E. Chaudry De Raynal, N. Frikha, *Well-posedness for some non-linear SDEs and related PDE on the Wasserstein space*, arXiv:1811.06904, to appear in J. Math. Pures Appl..

[10] H. Djellout, A. Guillin, L. Wu, *Transportation cost-information inequalities and applications to random dynamical systems and diffusions,* Ann. Probab. 32(2004), 2702-2732.

[11] E. Fedrizzi, F. Flandoli, E. Priola, J. Vovelle, *Regularity of stochastic kinetic equations,* Electron. J. Probab. 22(2017), 1-48.

[12] N. Fournier, A. Guillin, *On the rate of convergence in Wasserstein distance of the empirical measure,* Probab. Theory Related Fields 162(2015), 707-738.

[13] J. Gärtner, *On the McKean-Vlasov limit for interacting diffusions,* Mathematische Nachrichten 137(1988), 197-248.

[14] C. Graham, T. Kurtz, S. Méléard, P. Protter, M. Pulvirenti, D. Talay, *Probabilistic Models for Non-linear Partial Differential Equations, Lecture Notes in Mathematics 1627,* Springer-Verlag (1996).

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

[16] X. Huang, *Path-distribution dependent SDEs with singular coefficients,* Electron. J. Probab. 26(2021), 1-21.

[17] X. Huang, W. Lv, *Stochastic functional Hamiltonian system with singular coefficients,* Commun. Pur. Appl. Anal. 19(2020), 1257-1273.

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

[19] X. Huang, Y. Song, *Well-posedness and regularity for distribution dependent SPDEs with singular drifts,* Nonlinear Anal. 203(2021), 112167.

[20] B. Jourdain, J. Reygner, *Propagation of chaos for rank-based interacting diffusions and long time behaviour of a scalar quasilinear parabolic equation,* Stoch. PDE: Anal. Comp. 1(2013), 455-506.

[21] D. Lacker, *On a strong form of propagation of chaos for McKean-Vlasov equations,* Electron. Commun. Probab. 23(2018), 1-11.

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

[23] L. Miclo, P. Del Moral, *Genealogies and Increasing Propagation of Chaos For Feynman-Kac and Genetic Models,* Ann. Appl. Probab. 11(2001), 1166-1198.

[24] Yu. S. Mishura, A. Yu. Veretennikov, *Existence and uniqueness theorems for solutions of McKean-Vlasov stochastic equations,* Theor. Probability and Math. Statist. 103(2020), 59-101.

[25] M. Nagasawat, H. Tanaka, *Diffusion with Interactions and Collisions Between Coloured Particles and the Propagation of Chaos,* Probab. Theory Related Fields 74(1987), 161-198.

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

[27] P. Ren, F.-Y. Wang, *Exponential convergence in entropy and Wasserstein for McKean-Vlasov SDEs,* Nonlinear Anal. 206(2021), 112259.

[28] M. Röckner, X. Zhang, *Well-posedness of distribution dependent SDEs with singular drifts,* Bernoulli 27(2021), 1131-1158.

[29] S. T. Rachev, L. B. Klebanov, S. V. Stoyanov, F. J. Fabozzi, *Glivenko-Cantelli Theorem and Bernstein-Kantorovich Invariance Principle,* The Methods of Distances in the Theory of Probability and Statistics. Springer, New York, 2013.

[30] A.-S. Sznitman, *Topics in propagation of chaos,* In "École d'Été de Probabilités de Sain-Flour XIX-1989", Lecture Notes in Mathematics 1464, p. 165-251, Springer, Berlin, 1991.

[31] C. Villani, *Hypocoercivity,* Mem. Amer. Math. Soc. 202(2009).

[32] F.-Y. Wang, *Harnack Inequality for Stochastic Partial Differential Equations,* Springer, New York, 2013.

[33] F.-Y. Wang, *Distribution-dependent SDEs for Landau type equations,* Stochastic Process Appl. 128(2018), 595-621.

[34] F.-Y. Wang, *Hypercontractivity and Applications for Stochastic Hamiltonian Systems,* J. Funct. Anal. 272(2017), 5360-5383.

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

[36] F.-Y. Wang, X. Zhang, *Degenerate SDE with Hölder-Dini Drift and Non-Lipschitz Noise Coefficient,* SIAM J. Math. Anal. 48(2016), 2189-2226.

[37] X. Zhang, *Stochastic flows and Bismut formulas for stochastic Hamiltonian systems,* Stochastic Process Appl. 120(2010), 1929-1949.

[38] X. Zhang, *Stochastic hamiltonian flows with singular coefficients,* Sci. China Math. 61(2018),1353-1384.

[39] X. Zhang, *Second order McKean-Vlasov SDEs and kinetic Fokker-Planck-Kolmogorov equations,* arXiv:2109.01273.