Coupling by Change of Measure for Conditional McKean-Vlasov SDEs and Applications*

Xing Huang

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

xinghuang@tju.edu.cn

July 13, 2024

Abstract

The couplings by change of measure are applied to establish log-Harnack inequality(equivalently the entropy-cost estimate) for conditional McKean-Vlasov SDEs and derive the quantitative conditional propagation of chaos in relative entropy for mean field interacting particle system with common noise. For the log-Harnack inequality, two different types of couplings will be constructed for non-degenerate conditional McKean-Vlasov SDEs with multiplicative noise. As to the quantitative conditional propagation of chaos in relative entropy, the initial distribution of interacting particle system is allowed to be singular with that of limit equation. The above results are also extended to conditional distribution dependent stochastic Hamiltonian system.

AMS subject Classification: 60H10, 60G44.
Keywords: Conditional McKean-Vlasov SDEs, Log-Harnack inequality, Coupling by change of measure, Quantitative conditional propagation of chaos, Relative entropy.

1 Introduction

Distribution dependent stochastic differential equations(SDEs) can be viewed as the limit equation of a single particle in the mean field interacting particle system as the number of particles goes to infinity, see [29]. It is applied extensively in mean field games [22]. It is also

*Supported in part by National Key R&D Program of China (No. 2022YFA1006000) and NNSFC (12271398).

called a McKean-Vlasov SDE in the literature due to the work in [23]. Different from the classical Itô stochastic differential equation, the distribution of McKean-Vlasov SDEs solves a nonlinear Fokker-Planck-Kolmogorov equation. When there exists a common noise in the mean field interacting particle system, which is independent of the private noise of all particles, the limit equation of a single particle turns into a conditional distribution dependent SDE, which is called conditional McKean-Vlasov SDE, see [5]. Moreover, the conditional distribution of the solution with respect to the common noise is a probability measure-valued stochastic process, which solves a stochastic nonlinear Fokker-Planck-Kolmogorov equation, see for instance [5, 21]. Compared with the McKean-Vlasov SDEs, there are fewer results on conditional McKean-Vlasov ones. One can refer to [1, 3, 4, 5, 7, 13, 19, 21, 26, 27, 33] for well-posedness, [7, 17, 19] for the study of stochastic nonlinear Fokker-Planck-Kolmogorov equations and [1, 3, 6, 10, 17, 26, 27, 28, 30] for conditional propagation of chaos.

Let $\mathscr{P}(\mathbb{R}^d)$ be the space of all probability measures on $\mathbb{R}^d$ equipped with the weak topology. For $k \geq 1$, let

$$\mathscr{P}_k(\mathbb{R}^d) := \big\{\mu \in \mathscr{P}(\mathbb{R}^d) : \ \|\mu\|_k := \mu(|\cdot|^k)^{\frac{1}{k}} < \infty\big\},$$

which is a Polish space under the L^k-Wasserstein distance

$$\mathbb{W}_k(\mu,\nu) = \inf_{\pi\in\mathscr{C}(\mu,\nu)} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^k \pi(\mathrm{d}x,\mathrm{d}y)\right)^{\frac{1}{k}}, \quad \mu,\nu\in\mathscr{P}_k(\mathbb{R}^d),$$

where $\mathscr{C}(\mu,\nu)$ is the set of all couplings of μ and ν. The relative entropy of two probability measures is defined as

$$\mathrm{Ent}(\nu|\mu) = \begin{cases} \nu(\log(\frac{\mathrm{d}\nu}{\mathrm{d}\mu})), & \nu \ll \mu; \\ \infty, & \text{otherwise.} \end{cases}$$

Fix $T > 0$. As in [5, Section 2.1.3], let $(\Omega^i, \mathscr{F}^i, (\mathscr{F}^i_t)_{t\geq 0}, \mathbb{P}^i), i = 0,1$ be two complete filtration probability spaces and $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq0}, \mathbb{P})$ be the completion of the product structure generated by them, i.e. $\Omega = \Omega^0\times\Omega^1$, $\mathscr{F}$ and $(\mathscr{F}_t)_{t\geq0}$ are the completions of $\mathscr{F}^0\otimes\mathscr{F}^1$ and $(\mathscr{F}^0_t\otimes\mathscr{F}^1_t)_{t\geq0}$ with respect to the product measure $\mathbb{P} = \mathbb{P}^0\times\mathbb{P}^1$. Denote $\mathbb{E}$ the expectation associated to $\mathbb{P}$. W_t is a d_W-dimensional Brownian motion on $(\Omega^1, \mathscr{F}^1, (\mathscr{F}^1_t)_{t\geq0}, \mathbb{P}^1)$ while B_t is a d_B-dimensional Brownian motion on $(\Omega^0, \mathscr{F}^0, (\mathscr{F}^0_t)_{t\geq0}, \mathbb{P}^0)$. Let $\{\mathscr{F}^B_t\}_{t\in[0,T]}$ be the completion of $\sigma\{B_s, s\in[0,t]\}\otimes\{\emptyset,\Omega^1\}, t\in[0,T]$ with respect to $\mathbb{P}$. Consider conditional McKean-Vlasov SDEs:

$$\mathrm{d}X_t = b_t(X_t, \mathscr{L}_{X_t|\mathscr{F}^B_t})\mathrm{d}t + \sigma_t(X_t, \mathscr{L}_{X_t|\mathscr{F}^B_t})\mathrm{d}W_t + \tilde{\sigma}_t(X_t, \mathscr{L}_{X_t|\mathscr{F}^B_t})\mathrm{d}B_t, \tag{1.1}$$

where $\mathscr{L}_{X_t|\mathscr{F}^B_t}$ stands for the regular conditional distribution of X_t with respect to $\mathscr{F}^B_t$ under $\mathbb{P}$, $b : [0,T]\times\mathbb{R}^d\times\mathscr{P}(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : [0,T]\times\mathbb{R}^d\times\mathscr{P}(\mathbb{R}^d)\to\mathbb{R}^d\otimes\mathbb{R}^{d_W}$ and $\tilde{\sigma} : [0,T]\times\mathbb{R}^d\times\mathscr{P}(\mathbb{R}^d)\to\mathbb{R}^d\otimes\mathbb{R}^{d_B}$ are measurable and bounded on bounded set. The noise B_t is usually called the common noise while W_t is the private noise. Throughout the paper, we assume that the initial value X_0 is $(\Omega^1, \mathscr{F}^1_0)$-measurable. Note that when (1.1) is well-posed, see Definition 1.1 below, [5, Proposition 2.9 and Lemma 2.5] tells that $\{\mathscr{L}_{X_t|\mathscr{F}^B_t}\}_{t\in[0,T]}$ is a version of $\{\mathcal{L}^1(X_t)\}_{t\in[0,T]}$ in [5, (2.6)]. For more other assumptions on the initial value X_0, one can refer to [5, Remark 2.10].

Definition 1.1. For any $\xi \in L^2(\Omega^1 \to \mathbb{R}^d, \mathscr{F}_0^1, \mathbb{P}^1)$, we call a continuous and $(\mathscr{F}_t)_{t\geq 0}$-adapted process $(X_t)_{t\geq 0}$ with $\mathbb{E}\sup_{t\in[0,T]}|X_t|^2 < \infty, T > 0$ a solution to (1.1) with initial value ξ, if $\mathscr{L}_{X_t|\mathscr{F}_t^B}$ is a continuous, $\mathscr{F}_t^B$-adapted and $\mathscr{P}_2(\mathbb{R}^d)$-valued process and it holds

$$\mathbb{E}\int_0^T \{|b_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})| + \|\sigma_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})\|_{HS}^2 + \|\tilde{\sigma}_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})\|_{HS}^2\}\mathrm{d}t < \infty, \quad T > 0$$

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

$$X_s = \xi + \int_0^s b_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}t + \int_0^s \sigma_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}W_t + \int_0^s \tilde{\sigma}_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}B_t, \quad s \geq 0.$$

We call (1.1) is well-posed, if for any $\xi \in L^2(\Omega^1 \to \mathbb{R}^d, \mathscr{F}_0^1, \mathbb{P}^1)$, it has a unique solution starting from ξ which will be denoted by X_t^ξ in the sequel.

When (1.1) has a solution, Itô's formula implies that the conditional time-marginal distribution $\mu_t := \mathscr{L}_{X_t|\mathscr{F}_t^B}$ solves measured-valued stochastic evolution equation, i.e. for any $f \in C_0^\infty(\mathbb{R}^d)$, the smooth functions with compact support on $\mathbb{R}^d$,

$$\mathrm{d}(\mu_t(f)) = \mu_t(L_{t,\mu_t}f)\mathrm{d}t + \mu_t(\langle\tilde{\sigma}_t(\cdot, \mu_t), \nabla f\rangle)\mathrm{d}B_t,$$

where $L_{t,\mu}f(x) := \frac{1}{2}\mathrm{tr}[(\sigma_t\sigma_t^* + \tilde{\sigma}_t\tilde{\sigma}_t^*)(x,\mu)\nabla^2 f(x)] + \langle b_t(x,\mu), \nabla f(x)\rangle$ and $\mu_t(f) = \int_{\mathbb{R}^d} f\mathrm{d}\mu_t$, the study of which can be dated to [8], see also [5, 21]. Since then, it has been intensively investigated. [3] derived the well-posedness of mean reflected forward and backward SDEs and obtained the propagation of chaos in Wasserstein distance for the associated interacting particle system. Moreover, in the forward case, the conditional mean reflected SDEs and conditional propagation of chaos in Wasserstein distance are also studied; [4] studied a systemic risk control problem by the central bank, which stabilizes the interbank system with borrowing and lending activities and the mean field optimal control is shown to satisfy a stochastic Fokker-Planck-Kolmogorov equation driven by the common noise; In [19], the uniqueness for the stochastic nonlinear Fokker-Planck-Kolmogorov equation is proved in the class of solutions with squarely integrable density with respect to the Lebesgue measure; In [7], the uniqueness is shown by means of a duality argument to a backward stochastic PDE and [9] verifies the uniqueness of solutions by a dual method, coupling arguments as well as the Krylov-Rozovskii variational approach to SPDE. In [21], the superposition principle and mimicking theorem for conditional McKean-Vlasov SDE are derived, which establish the correspondence between conditional McKean-Vlasov SDE and stochastic nonlinear Fokker-Planck-Kolmogorov equation under reasonable condition and also show that the conditional time-marginals of an Itô process can be constructed by those of the solution to a conditional McKean-Vlasov SDE with Markovian coefficients. This provides a probability method to investigate stochastic nonlinear Fokker-Planck-Kolmogorov equation.

In recent years, the study of (1.1) attracts much attention. [13] proved that (1.1) is well-posed if (1.1) with $b = 0$ is well-posed, σ and $\tilde{\sigma}$ are distribution free and $\sigma^{-1}b$ is bounded and Lipchitz continuous under total variation distance. In [17], the quantitative conditional propagation of chaos in weak convergence is provided, where σ and $\tilde{\sigma}$ are distribution free and all the coefficients are regular enough in spatial-measure arguments. [6] proved conditional propagation of chaos in Wasserstein distance when $\sigma = 0$, $b_t(x,\mu) = \mu(f(x-\cdot))$ for some Lipschitz function f. [10] investigated conditional propagation of chaos for one dimensional SDEs driven by Poisson random measure and common Brownian motion noise, where $\tilde{\sigma} = \sqrt{\mu(f)}$ for some positive Lipschitz function f. For moderately interacting particle systems with environmental noise and singular interaction kernel such as the Biot-Savart and repulsive Poisson kernels, [12] proved that the mollified empirical measures converge in strong norms to the unique (local) solutions of nonlinear Fokker-Planck-Kolmogorov equations. [33] studied the well-posedness in the case $\sigma = 0$ by constructing image dependent SDE. In [30], the quantitative conditional propagation of chaos in the sense of Wasserstein distance is studied for stochastic spatial epidemic model, where the evolution of infection states are driven by the Poisson point processes and the displacement of individuals contains a common noise. Quite recently, adopting the technology of disintegration and the entropy method developed by [15], [28] established quantitative conditional propagation of chaos in relative entropy for the stochastic 2-dimensional Navier-Stokes equation in torus. One can also refer to [26, 27] for the (conditional)propagation of chaos for (conditional)McKean-Vlasov SDEs with regime-switching.

The propagation of chaos is a hot topic in the McKean-Vlasov frame($\tilde{\sigma} = 0$). The quantitative propagation of chaos in strong sense is studied in [29] by using synchronous coupling argument, where the coefficients are assumed to be Lipschitz continuous and the initial value of interacting particle system coincides with that of the limit equation. [2, 15, 16] apply the entropy method to derive the quantitative propagation of chaos in relative entropy with additive noise and singular interaction, for which the initial distribution of interacting particle system is assumed to be absolutely continuous with that of limit equation. In [20], the authors give the sharp rate of propagation of chaos for some models such as bounded or uniformly continuous interaction by BBGKY hierarchy. We should also mention that in [24], the (uniform in time)quantitative propagation of chaos for genetic-type interacting particle system approximating model in the sense of relative entropy as well as $\mathbb{L}_\alpha(\alpha \in [1,\infty])$ estimate and thus in the sense of total variation distance are obtained.

As far as we know, the regularity estimate of conditional McKean-Vlasov SDEs with respect to the initial value such as the entropy-cost estimate is still open. In this paper, we try to construct the coupling by change of measure for conditional McKean-Vlasov SDEs (1.1). We will present two different couplings in the case with non-degenerate and multiplicative noise to derive the log-Harnack inequality, which is equivalent to entropy-cost estimate. In the distribution independent case, the log-Harnack inequality associated to a Markov semigroup P_t is formulated as

$$P_t \log f(x) \le \log P_t f(y) + c(t)|x-y|^2, \quad f \in \mathscr{B}_b^+(\mathbb{R}^d), t \in (0,T], x, y \in \mathbb{R}^d$$

for some nonnegative function c with $\lim_{t\to 0} c(t) = \infty$, which implies the gradient-L^2 estimate:

$$|\nabla P_t f|^2 \le c(t) P_t |f|^2, \quad t \in (0,T].$$

In the case of non-degenerate diffusion, it is also equivalent to the gradient-gradient estimate:

$$|\nabla P_t f|^2 \leq C P_t |\nabla f|^2, \quad t \in [0,T]$$

for some constant $C > 0$. One can refer to [32, Chapter 1] for more details.

Different from the McKean-Vlasov frame, the conditional distribution with respect to the common noise is a functional of common noise so that we have to overcome essential difficulties produced by this crucial difference. For instance, in the procedure of constructing coupling processes, we usually view the conditional distribution with respect to the common noise as a known functional of common noise so that the common noise need also be fixed. Hence, we can only construct a new private noise in coupling process. Moreover, since the private noise and the common noise are independent, when we calculate the expectation for a functional of (W, B), we can firstly take conditional expectation with respect to the common noise in which the common noise can be viewed as a constant and then use the tower property of conditional expectation to realize this goal.

We will also investigate the quantitative conditional propagation of chaos in the sense of Wasserstein distance, which together with coupling by change of measure implies the quantitative conditional propagation of chaos in relative entropy. Different from [2, 15, 16], the initial distribution of interacting particle system is allowed to be singular with that of the limit equation. The main tool is an entropy inequality in [25, Lemma 2.1] as well as Wang's Harnack inequality with power, see for instance [31] and the monograph [32]. Furthermore, the associated assertions are derived by the method of coupling by change of measure for the conditional distribution dependent stochastic Hamiltonian system and mean field interacting stochastic Hamiltonian system with common noise.

When the conditional distribution is involved, an inequality is often used:

$$\mathbb{E}\mathbb{W}_2(\mathscr{L}_{\xi|\mathscr{G}}, \mathscr{L}_{\eta|\mathscr{G}})^2 \leq \mathbb{E}\left\{\mathbb{E}(|\xi - \eta|^2|\mathscr{G})\right\} = \mathbb{E}|\xi - \eta|^2$$

for any random variables ξ, η with finite second monents and any sub-σ-algebra $\mathscr{G} \subset \mathscr{F}$. Using Banach's fixed point theorem and repeating the proof of [5, Proposition 2.11], it is standard to obtain Lemma 1.1 below under the following monotonicity condition **(H)**, see for instance [18, Theorem 2.1]. One can also refer to [33] for image-dependent SDE, a special type of conditional McKean-Vlasov SDE aforementioned. When (1.1) is well-posed and for any $\gamma \in \mathscr{P}_2(\mathbb{R}^d)$ and any $\xi, \tilde{\xi} \in L^2(\Omega^1 \to \mathbb{R}^d, \mathscr{F}_0^1, \mathbb{P}^1)$ with $\mathscr{L}_\xi = \mathscr{L}_{\tilde{\xi}} = \gamma$, it holds $\mathscr{L}_{X_t^\xi} = \mathscr{L}_{X_t^{\tilde{\xi}}}$, then we denote $P_t^*\gamma = \mathscr{L}_{X_t^\xi}$ and

$$P_t f(\gamma) := \int_{\mathbb{R}^d} f(x)(P_t^*\gamma)(\mathrm{d}x), \quad f \in \mathscr{B}_b(\mathbb{R}^d).$$

(H) For any $t \in [0,T]$, $b_t, \sigma_t, \tilde{\sigma}_t$ are continuous in $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$. There exists a constant $K \geq 0$ such that

$$\begin{aligned}&\|\sigma_t(x,\mu) - \sigma_t(y,\nu)\|_{HS}^2 + \|\tilde{\sigma}_t(x,\mu) - \tilde{\sigma}_t(y,\nu)\|_{HS}^2 + 2\langle b_t(x,\mu) - b_t(y,\nu), x - y\rangle \\ &\leq K(|x-y|^2 + \mathbb{W}_2(\mu,\nu)^2), \quad t \in [0,T],\ x, y \in \mathbb{R}^d,\ \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d).\end{aligned}$$

Lemma 1.1. *Assume* **(H)**. *Then* (1.1) *is well-posed and* $\mathscr{L}_{X_t^\xi|\mathscr{F}_t^B} = \mathscr{L}_{X_t^{\tilde\xi}|\mathscr{F}_t^B}$ *for any initial values* $\xi, \tilde\xi \in L^2(\Omega^1 \to \mathbb{R}^d, \mathscr{F}_0^1, \mathbb{P}^1)$ *with* $\mathscr{L}_\xi = \mathscr{L}_{\tilde\xi}$. *Moreover, there exists a constant* $C_T = e^{2KT}$ *such that for all* $s \in [0,T]$, *and* $\xi, \tilde\xi \in L^2(\Omega^1 \to \mathbb{R}^d, \mathscr{F}_0^1, \mathbb{P}^1)$,

$$\mathbb{E}\mathbb{W}_2(\mathscr{L}_{X_s^\xi|\mathscr{F}_s^B}, \mathscr{L}_{X_s^{\tilde\xi}|\mathscr{F}_s^B})^2 + \mathbb{W}_2(\mathscr{L}_{X_s^\xi}, \mathscr{L}_{X_s^{\tilde\xi}})^2 \leq C_T \mathbb{W}_2(\mathscr{L}_\xi, \mathscr{L}_{\tilde\xi})^2.$$

When there are different probability measures on $(\Omega, \mathscr{F})$, we use $\mathscr{L}_\xi^{\mathbb{P}}$ and $\mathscr{L}_{\xi|\mathscr{G}}^{\mathbb{P}}$ to denote the distribution and regular conditional distribution of a random variable ξ with respect to sub-σ-algebra $\mathscr{G} \subset \mathscr{F}$ respectively under probability measure $\mathbb{P}$.

The remaining of the paper is organized as follows: In section 2, we establish the log-Harnack inequality and thus the entropy-cost estimate for conditional McKean-Vlasov SDEs with non-degenerate and multiplicative noise and two different cases are considered. Moreover, we investigate the quantitative conditional propagation of chaos in Wasserstein distance and relative entropy. The corresponding results are derived in Section 3 for conditional distribution dependent stochastic Hamiltonian system and mean field interacting stochastic Hamiltonian system with common noise.

2 Non-degenerate case

2.1 Log-Harnack inequality

To apply the coupling by change of measure to establish the log-Harnack inequality for conditional McKean-Vlasov SDEs, we assume σ is distribution free and consider

$$\mathrm{d}X_t = b_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}t + \sigma_t(X_t)\mathrm{d}W_t + \tilde\sigma_t(X_t, \mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}B_t. \tag{2.1}$$

In the following, we will investigate two different cases and construct corresponding couplings by change of measure to derive the log-Harnack inequality for (2.1).

2.1.1 State-dependent case: $\tilde\sigma_t(x,\mu) = \tilde\sigma_t(x)$

(A) For any $t \in [0,T], x \in \mathbb{R}^d$, $(\sigma_t\sigma_t^*)(x)$ is invertible and b_t is continuous in $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$. There exist $\lambda \in (0,1]$ and $K, \tilde K \geq 0$ such that

$$\begin{aligned}&\lambda^{-1} \geq (\sigma_t\sigma_t^*)(x) \geq \lambda, \quad \|\sigma_t(x) - \sigma_t(y)\|_{HS}^2 \leq K|x-y|^2,\\ &\langle b_t(x,\mu) - b_t(y,\nu), x-y\rangle \leq K(|x-y|\mathbb{W}_2(\mu,\nu) + |x-y|^2),\\ &\|\tilde\sigma_t(x) - \tilde\sigma_t(y)\|_{HS}^2 \leq \tilde K|x-y|^2, \quad t \in [0,T],\ x,y \in \mathbb{R}^d,\ \mu,\nu \in \mathscr{P}_2(\mathbb{R}^d).\end{aligned}$$

Theorem 2.1. *Assume* **(A)**. *Then there exists a constant* $c > 0$ *such that*

$$P_t \log f(\nu_0) \leq \log P_t f(\mu_0) + c\left\{\frac{(3K+\tilde K)}{1 - e^{-(3K+\tilde K)t}} + t\right\}\mathbb{W}_2(\mu_0,\nu_0)^2,$$
$$0 < f \in \mathscr{B}_b(\mathbb{R}^d), \mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^d), t \in (0,T].$$

Proof. We will follow essentially the line of [31, Theorem 1.1(1)] to complete the proof of Theorem 2.1. Let $X_0^{\mu_0}, X_0^{\nu_0}$ be $(\Omega^1, \mathscr{F}_0^1)$-measurable such that

$$\mathscr{L}_{X_0^{\mu_0}} = \mu_0, \quad \mathscr{L}_{X_0^{\nu_0}} = \nu_0, \quad \mathbb{E}|X_0^{\mu_0} - X_0^{\nu_0}|^2 = \mathbb{W}_2(\mu_0, \nu_0)^2. \tag{2.2}$$

Let $X_t^{\mu_0}$ and $X_t^{\nu_0}$ solve (2.1) with initial values $X_0^{\mu_0}$ and $X_0^{\nu_0}$ respectively. Denote

$$\nu_t = \mathscr{L}^{\mathbb{P}}_{X_t^{\nu_0}|\mathscr{F}_t^B}, \quad \mu_t = \mathscr{L}^{\mathbb{P}}_{X_t^{\mu_0}|\mathscr{F}_t^B}, \quad t \in [0, T]. \tag{2.3}$$

Then it holds

$$\mathrm{d}X_t^{\mu_0} = b_t(X_t^{\mu_0}, \mu_t)\mathrm{d}t + \sigma_t(X_t^{\mu_0})\mathrm{d}W_t + \tilde{\sigma}_t(X_t^{\mu_0})\mathrm{d}B_t, \quad t \in [0, T]. \tag{2.4}$$

Let $t_0 \in (0, T]$ and $\xi_t = \frac{1}{3K+\tilde{K}}(1 - \mathrm{e}^{(3K+\tilde{K})(t-t_0)}), t \in [0, t_0]$, which satisfies

$$-\xi_t' + (3K + \tilde{K})\xi_t = 1. \tag{2.5}$$

Consider the following SDE:

$$\begin{aligned} \mathrm{d}Y_t &= b_t(Y_t, \nu_t)\mathrm{d}t + \sigma_t(Y_t)\mathrm{d}W_t + \tilde{\sigma}_t(Y_t)\mathrm{d}B_t \\ &\quad + \sigma_t(Y_t)[\sigma_t^*(\sigma_t\sigma_t^*)^{-1}](X_t^{\mu_0})\frac{X_t^{\mu_0} - Y_t}{\xi_t}\mathrm{d}t, \qquad t \in [0, t_0), \ Y_0 = X_0^{\nu_0}. \end{aligned} \tag{2.6}$$

Let $\tau_n = t_0 \wedge \inf\{t \in [0, t_0), |X_t^{\mu_0}| \vee |Y_t| \geq n\}$. Then $\mathbb{P}$-a.s. $\tau_n \uparrow t_0$ as $n \uparrow \infty$. Let

$$\begin{aligned} &\gamma_t := [\sigma_t^*(\sigma_t\sigma_t^*)^{-1}](X_t^{\mu_0})\frac{Y_t - X_t^{\mu_0}}{\xi_t}, \\ &\hat{W}_t := W_t - \int_0^t \gamma_s \mathrm{d}s, \quad R_t := \mathrm{e}^{\int_0^t \langle \gamma_r, \mathrm{d}W_r\rangle - \frac{1}{2}\int_0^t |\gamma_r|^2 \mathrm{d}r}, \\ &\mathbb{Q}_t := R_t\mathbb{P}, \quad t \in [0, t_0). \end{aligned} \tag{2.7}$$

Fix $s \in [0, t_0)$. According to Girsanov's theorem, under the weighted probability $\mathbb{Q}_{s\wedge\tau_n}$, $(\hat{W}_t, B_t)$ is a $(d_W + d_B)$-dimensional Brownian motion up to time $s \wedge \tau_n$.

Then (2.4) and (2.6) can be rewritten as

$$\mathrm{d}X_t^{\mu_0} = b_t(X_t^{\mu_0}, \mu_t)\mathrm{d}t + \sigma_t(X_t^{\mu_0})\mathrm{d}\hat{W}_t + \tilde{\sigma}_t(X_t^{\mu_0})\mathrm{d}B_t + \frac{Y_t - X_t^{\mu_0}}{\xi_t}\mathrm{d}t, \quad t \in [0, s \wedge \tau_n],$$

and

$$\mathrm{d}Y_t = b_t(Y_t, \nu_t)\mathrm{d}t + \sigma_t(Y_t)\mathrm{d}\hat{W}_t + \tilde{\sigma}_t(Y_t)\mathrm{d}B_t, \qquad t \in [0, s \wedge \tau_n], \quad Y_0 = X_0^{\nu_0}.$$

It follows from Itô's formula that

$$\begin{aligned} &\mathrm{d}\frac{|Y_t - X_t^{\mu_0}|^2}{\xi_t} \\ &= -\frac{\xi_t'|Y_t - X_t^{\mu_0}|^2}{\xi_t^2}\mathrm{d}t + \frac{2\langle b_t(Y_t, \nu_t) - b_t(X_t^{\mu_0}, \mu_t), Y_t - X_t^{\mu_0}\rangle}{\xi_t}\mathrm{d}t - 2\frac{|Y_t - X_t^{\mu_0}|^2}{\xi_t^2}\mathrm{d}t \\ &+ \frac{2\langle [\sigma_t(Y_t) - \sigma_t(X_t^{\mu_0})]\mathrm{d}\hat{W}_t + [\tilde{\sigma}_t(Y_t) - \tilde{\sigma}_t(X_t^{\mu_0})]\mathrm{d}B_t, Y_t - X_t^{\mu_0}\rangle}{\xi_t} \\ &+ \frac{\|\sigma_t(Y_t) - \sigma_t(X_t^{\mu_0})\|_{HS}^2 + \|\tilde{\sigma}_t(Y_t) - \tilde{\sigma}_t(X_t^{\mu_0})\|_{HS}^2}{\xi_t}\mathrm{d}t, \quad t \in [0, s \wedge \tau_n]. \end{aligned} \tag{2.8}$$

In view of **(A)** and the inequality $ab \le \frac{1}{2}a^2 + \frac{1}{2}b^2, a, b \in \mathbb{R}$, we conclude that

$$\begin{aligned}
&\frac{2\langle b_t(Y_t,\nu_t) - b_t(X_t^{\mu_0},\mu_t), Y_t - X_t^{\mu_0}\rangle}{\xi_t}\\
&\le \frac{2K|Y_t - X_t^{\mu_0}|\mathbb{W}_2(\mu_t,\nu_t)}{\xi_t} + \frac{2K|Y_t - X_t^{\mu_0}|^2}{\xi_t}\\
&\le \frac{1}{2}\frac{|Y_t - X_t^{\mu_0}|^2}{\xi_t^2} + 2K^2\mathbb{W}_2(\mu_t,\nu_t)^2 + \frac{2K\xi_t|Y_t - X_t^{\mu_0}|^2}{\xi_t^2}, \quad t \in [0, s\wedge\tau_n],
\end{aligned}$$

and

$$\frac{\|\sigma_t(Y_t) - \sigma_t(X_t^{\mu_0})\|_{HS}^2 + \|\tilde\sigma_t(Y_t) - \tilde\sigma_t(X_t^{\mu_0})\|_{HS}^2}{\xi_t} \le \frac{(K+\tilde K)\xi_t|Y_t - X_t^{\mu_0}|^2}{\xi_t^2}, \quad t \in [0, s\wedge\tau_n].$$

This together with (2.8) gives

$$\begin{aligned}
\text{(2.9)} \qquad \mathrm{d}\frac{|Y_t - X_t^{\mu_0}|^2}{\xi_t} &\le \frac{[-\xi_t' + (3K+\tilde K)\xi_t - \frac{3}{2}]|Y_t - X_t^{\mu_0}|^2}{\xi_t^2}\mathrm{d}t\\
&\quad + 2K^2\mathbb{W}_2(\mu_t,\nu_t)^2\mathrm{d}t + \mathrm{d}M_t, \quad t \in [0, s\wedge\tau_n],
\end{aligned}$$

where

$$\mathrm{d}M_t = \frac{2\langle[\sigma_t(Y_t) - \sigma_t(X_t^{\mu_0})]\mathrm{d}\hat W_t + [\tilde\sigma_t(Y_t) - \tilde\sigma_t(X_t^{\mu_0})]\mathrm{d}B_t, Y_t - X_t^{\mu_0}\rangle}{\xi_t}.$$

Since W is independent of B, we have

$$\mathbb{E}(R_{s\wedge\tau_n}|\mathscr{F}_s^B) = 1,$$

which together with (2.3), the definition of μ_t, ν_t and Lemma 1.1 implies

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_{s\wedge\tau_n}}\int_0^{s\wedge\tau_n} \mathbb{W}_2(\mu_t,\nu_t)^2\mathrm{d}t\\
&\le \mathbb{E}\left\{\mathbb{E}(R_{s\wedge\tau_n}|\mathscr{F}_s^B)\int_0^s \mathbb{W}_2(\mu_t,\nu_t)^2\mathrm{d}t\right\}\\
&= \mathbb{E}\int_0^s \mathbb{W}_2(\mu_t,\nu_t)^2\mathrm{d}t \le Cs\mathbb{W}_2(\mu_0,\nu_0)^2.
\end{aligned}$$

Combining this with (2.5) and (2.9), we derive

$$\begin{aligned}
&\mathbb{E}_{\mathbb{Q}_{s\wedge\tau_n}}\int_0^{s\wedge\tau_n}\frac{|Y_t - X_t^{\mu_0}|^2}{\xi_t^2}\mathrm{d}t \le 2\mathbb{E}_{\mathbb{Q}_{s\wedge\tau_n}}\frac{|Y_0 - X_0^{\mu_0}|^2}{\xi_0} + \mathbb{E}_{\mathbb{Q}_{s\wedge\tau_n}}\int_0^{s\wedge\tau_n} 4K^2\mathbb{W}_2(\mu_t,\nu_t)^2\mathrm{d}t\\
&\le 2\frac{\mathbb{W}_2(\mu_0,\nu_0)^2}{\xi_0} + 4K^2Cs\mathbb{W}_2(\mu_0,\nu_0)^2.
\end{aligned}$$

Hence, by (2.7) and **(A)**, we find a constant $c_1 > 0$ such that

$$\mathbb{E}[R_{s\wedge\tau_n}\log R_{s\wedge\tau_n}] = \mathbb{E}_{\mathbb{Q}_{s\wedge\tau_n}}[\log R_{s\wedge\tau_n}] = \frac{1}{2}\mathbb{E}_{\mathbb{Q}_{s\wedge\tau_n}}\int_0^{s\wedge\tau_n}|\gamma_t|^2\mathrm{d}t$$

$$\leq c_1\frac{\mathbb{W}_2(\mu_0,\nu_0)^2}{\xi_0}+c_1s\mathbb{W}_2(\mu_0,\nu_0)^2.$$

Consequently, $\{R_{s\wedge\tau_n}\}_{n\geq 1}$ is a uniformly integrable martingale under $\mathbb{P}$, which together with the martingale convergence theorem and Fatou's lemma implies that

$$\frac{1}{2}\mathbb{E}_{\mathbb{Q}_s}\int_0^s|\gamma_t|^2\mathrm{d}t=\mathbb{E}[R_s\log R_s]\leq c_1\frac{\mathbb{W}_2(\mu_0,\nu_0)^2}{\xi_0}+c_1s\mathbb{W}_2(\mu_0,\nu_0)^2,\quad s\in[0,t_0). \tag{2.10}$$

This means that $\{R_s\}_{s\in[0,t_0]}$ is a uniformly integrable martingale under $\mathbb{P}$ and Girsanov's theorem yields that under the weighted probability $\mathbb{Q}_{t_0}$, $(\hat{W}_t,B_t)_{t\in[0,t_0]}$ is a (d_W+d_B)-dimensional Brownian motion. Moreover, $\mathbb{Q}_{t_0}$-a.s. $Y_{t_0}=X_{t_0}^{\mu_0}$ by (2.10) for $s=t_0$ due to Fatou's lemma. On the other hand, consider the conditional McKean-Vlasov SDE

$$\mathrm{d}\tilde{Y}_t=b_t(\tilde{Y}_t,\mathscr{L}^{\mathbb{Q}_{t_0}}_{\tilde{Y}_t|\mathscr{F}_t^B})\mathrm{d}t+\sigma_t(\tilde{Y}_t)\mathrm{d}\hat{W}_t+\tilde{\sigma}_t(\tilde{Y}_t)\mathrm{d}B_t,\qquad t\in[0,t_0],\quad \tilde{Y}_0=X_0^{\nu_0}. \tag{2.11}$$

According to [5, Proposition 2.11], we derive $\nu_t=\mathscr{L}^{\mathbb{Q}_{t_0}}_{\tilde{Y}_t|\mathscr{F}_t^B}$ and $\mathscr{L}^{\mathbb{P}}_{X_t^{\nu_0}}=\mathscr{L}^{\mathbb{Q}_{t_0}}_{\tilde{Y}_t}$ so that (2.11) can be rewritten as

$$\mathrm{d}\tilde{Y}_t=b_t(\tilde{Y}_t,\nu_t)\mathrm{d}t+\sigma_t(\tilde{Y}_t)\mathrm{d}\hat{W}_t+\tilde{\sigma}_t(\tilde{Y}_t)\mathrm{d}B_t,\qquad t\in[0,t_0],\tilde{Y}_0=X_0^{\nu_0}. \tag{2.12}$$

The strong uniqueness of (2.12) implies $Y_t=\tilde{Y}_t, t\in[0,t_0]$. In fact, (2.12) is an SDE with random coefficients, the well-posedness of which can be proved by standard argument under the assumption **(A)**. Therefore, $\mathscr{L}^{\mathbb{Q}_{t_0}}_{Y_t|\mathscr{F}_t^B}=\mathscr{L}^{\mathbb{P}}_{X_t^{\nu_0}|\mathscr{F}_t^B}=\nu_t$ and $\mathscr{L}^{\mathbb{Q}_{t_0}}_{Y_t}=\mathscr{L}^{\mathbb{P}}_{X_t^{\nu_0}}$. Combining this with Young's inequality and (2.10) for $s=t_0$ due to Fatou's lemma, we derive

$$\begin{aligned}
&P_{t_0}\log f(\nu_0)=\mathbb{E}_{\mathbb{Q}_{t_0}}[\log f(Y_{t_0})]=\mathbb{E}[R_{t_0}\log f(X_{t_0}^{\mu_0})]\\
&\leq\log\mathbb{E}[f(X_{t_0}^{\mu_0})]+\mathbb{E}[R_{t_0}\log R_{t_0}]\\
&\leq\log P_{t_0}f(\mu_0)+c_1\frac{\mathbb{W}_2(\mu_0,\nu_0)^2}{\xi_0}+c_1t_0\mathbb{W}_2(\mu_0,\nu_0)^2,\quad 0<f\in\mathscr{B}_b(\mathbb{R}^d).
\end{aligned}$$

Therefore, we complete the proof by the definition of ξ_0. □

2.1.2 Measure-dependent case: $\tilde{\sigma}_t(x,\mu)=\tilde{\sigma}_t(\mu)$

In the second case, we assume that $\tilde{\sigma}$ only depends on the time-distribution arguments, i.e. consider

$$\mathrm{d}X_t=b_t(X_t,\mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}t+\sigma_t(X_t)\mathrm{d}W_t+\tilde{\sigma}_t(\mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}B_t,\quad t\in[0,T]. \tag{2.13}$$

To establish the log-Harnack inequality, we make the following Lipschitz assumption on b instead of the monotonicity condition on b in **(A)**.

(B) For any $t \in [0,T], x \in \mathbb{R}^d$, $(\sigma_t\sigma_t^*)(x)$ is invertible and there exist $\lambda \in (0,1]$ and $K, \tilde{K} \geq 0$ such that

$$\begin{aligned}
&\lambda^{-1} \geq (\sigma_t\sigma_t^*)(x) \geq \lambda, \quad \|\sigma_t(x) - \sigma_t(y)\|_{HS}^2 \leq K|x-y|^2,\\
&|b_t(x,\mu) - b_t(y,\nu)| \leq K(|x-y| + \mathbb{W}_2(\mu,\nu)),\\
&\|\tilde{\sigma}_t(\mu) - \tilde{\sigma}_t(\nu)\|_{HS}^2 \leq \tilde{K}\mathbb{W}_2(\mu,\nu)^2, \quad t \in [0,T],\ x,y \in \mathbb{R}^d,\ \mu,\nu \in \mathscr{P}_2(\mathbb{R}^d).
\end{aligned}$$

In **(A)**, $\tilde{\sigma}$ is only allowed to depend on the time and spatial variables while $\tilde{\sigma}$ only depends on time and measure variables in **(B)** and the condition for b in **(B)** can derive that for b in **(A)**. Since assumption **(B)** implies **(H)**, Lemma 1.1 holds for SDE (2.13) replacing (1.1). In the case that $\tilde{\sigma}_t(x,\mu) = \tilde{\sigma}_t(\mu)$, the coupling used in Section 2.1 is unavailable so that we need to construct a new coupling by change of measure which involves in conditional probability with respect to $\mathscr{F}_t^B$.

Theorem 2.2. *Assume* **(B)**. *Then there exists a constant $c > 0$ such that for any $0 < f \in \mathscr{B}_b(\mathbb{R}^d), \mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^d), t \in (0,T]$ and $\xi, \tilde{\xi} \in L^2(\Omega^1 \to \mathbb{R}^d, \mathscr{F}_0^1, \mathbb{P}^1)$ with $\mathscr{L}_\xi = \mu_0, \mathscr{L}_{\tilde{\xi}} = \nu_0$,*

$$\mathbb{E}\{\mathrm{Ent}(\mathscr{L}_{X_t^\xi|\mathscr{F}_t^B}|\mathscr{L}_{X_t^{\tilde{\xi}}|\mathscr{F}_t^B})\} \leq c\left\{\frac{4K}{1-\mathrm{e}^{-4Kt}} + \int_0^t \frac{4Ks}{1-\mathrm{e}^{-4Ks}}\mathrm{d}s\right\}\mathbb{W}_2(\mu_0,\nu_0)^2,$$

and consequently,

$$P_t \log f(\nu_0) \leq \log P_t f(\mu_0) + c\left\{\frac{4K}{1-\mathrm{e}^{-4Kt}} + \int_0^t \frac{4Kr}{1-\mathrm{e}^{-4Kr}}\mathrm{d}r\right\}\mathbb{W}_2(\mu_0,\nu_0)^2. \tag{2.14}$$

Proof. For fixed $\mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^d)$, let $X_0^{\mu_0}, X_0^{\nu_0}$ be chosen in (2.2). Let $X_t^{\mu_0}$ and $X_t^{\nu_0}$ solve (2.13) with initial value $X_0^{\mu_0}$ and $X_0^{\nu_0}$ respectively and μ_t and ν_t be defined in (2.3). Define

$$\eta_t^\mu := \int_0^t \tilde{\sigma}_s(\mu_s)\mathrm{d}B_s, \quad t \in [0,T]. \tag{2.15}$$

By **(B)**, BDG's inequality and Lemma 1.1, we find a constant $C_1 > 0$ such that

$$\mathbb{E}\Big[\sup_{t\in[0,T]}|\eta_t^\mu - \eta_t^\nu|^2\Big] \leq \tilde{K}^2\mathbb{E}\int_0^T \mathbb{W}_2(\mu_s,\nu_s)^2\mathrm{d}s \leq C_1\tilde{K}^2T\mathbb{W}_2(\mu_0,\nu_0)^2, \tag{2.16}$$

and

$$\mathbb{E}|\eta_{t_1}^\mu - \eta_{t_2}^\mu - (\eta_{t_1}^\nu - \eta_{t_2}^\nu)|^2 \leq C_1\tilde{K}^2\mathbb{W}_2(\mu_0,\nu_0)^2|t_1-t_2|, \quad t_1,t_2 \in [0,T]. \tag{2.17}$$

Moreover, we derive from (2.3) that

$$\mathrm{d}X_t^{\mu_0} = b_t(X_t^{\mu_0},\mu_t)\mathrm{d}t + \sigma_t(X_t^{\mu_0})\mathrm{d}W_t + \tilde{\sigma}_t(\mu_t)\mathrm{d}B_t, \quad t \in [0,T].$$

Then $\hat{X}_t^{\mu_0} := X_t^{\mu_0} - \eta_t^\mu, \quad t \in [0,T]$ solves

$$\mathrm{d}\hat{X}_t^{\mu_0} = b_t(\hat{X}_t^{\mu_0} + \eta_t^\mu, \mu_t)\mathrm{d}t + \sigma_t(\hat{X}_t^{\mu_0} + \eta_t^\mu)\mathrm{d}W_t, \quad t \in [0,T]. \tag{2.18}$$

Let $t_0 \in (0,T]$ and $\xi_t = \frac{1}{4K}(1-\mathrm{e}^{4K(t-t_0)})$ and it holds

$$-\xi_t' + 4K\xi_t = 1. \tag{2.19}$$

Now, we construct the coupling process:

$$\begin{aligned}\mathrm{d}\hat{Y}_t^\nu &= b_t(\hat{Y}_t^\nu + \eta_t^\nu, \nu_t)\mathrm{d}t + \sigma_t(\hat{Y}_t^\nu + \eta_t^\nu)\mathrm{d}W_t \\ &\quad + \sigma_t(\hat{Y}_t^\nu + \eta_t^\nu)[\sigma_t^*(\sigma_t\sigma_t^*)^{-1}](\hat{X}_t^{\mu_0} + \eta_t^\mu)\frac{(\hat{X}_t^{\mu_0} + \eta_{t_0}^\mu) - (\hat{Y}_t^\nu + \eta_{t_0}^\nu)}{\xi_t}\mathrm{d}t, \\ &\quad t \in [0,t_0),\ \hat{Y}_0^\nu = X_0^{\nu_0}.\end{aligned} \tag{2.20}$$

Define

$$\mathbb{P}^B := \mathbb{P}(\ \cdot\ |\mathscr{F}_T^B), \quad \mathbb{E}^B := \mathbb{E}(\ \cdot\ |\mathscr{F}_T^B). \tag{2.21}$$

Set $\tau_n = t_0 \wedge \inf\{t \in [0,t_0), |\hat{X}_t^\mu + \eta_{t_0}^\mu| \vee |\hat{Y}_t^\nu + \eta_{t_0}^\nu| \ge n\}$. Then we have $\mathbb{P}^B$-a.s. $\tau_n \uparrow t_0$ as $n \uparrow \infty$. Let

$$\begin{aligned}&\beta_t := [\sigma_t^*(\sigma_t\sigma_t^*)^{-1}](\hat{X}_t^{\mu_0} + \eta_t^\mu)\frac{(\hat{Y}_t^\nu + \eta_{t_0}^\nu) - (\hat{X}_t^{\mu_0} + \eta_{t_0}^\mu)}{\xi_t}, \\ &\tilde{W}_t := W_t - \int_0^t \beta_s \mathrm{d}s, \quad R_t := \mathrm{e}^{\int_0^t \langle \beta_r, \mathrm{d}W_r\rangle - \frac{1}{2}\int_0^t |\beta_r|^2 \mathrm{d}r}, \\ &\mathbb{Q}_t^B := R_t\mathbb{P}^B, \quad t \in [0,t_0).\end{aligned} \tag{2.22}$$

Fix $s \in [0,t_0)$. Girsanov's theorem yields that under the weighted conditional probability $\mathbb{Q}_{s\wedge\tau_n}^B$, $\tilde{W}_t$ is a d_W-dimensional Brownian motion on $[0, s\wedge\tau_n]$. Hence, (2.18) and (2.20) can be reformulated as

$$\begin{aligned}\mathrm{d}[\hat{X}_t^{\mu_0} + \eta_{t_0}^\mu - \eta_{t_0}^\nu)] &= b_t(\hat{X}_t^{\mu_0} + \eta_t^\mu, \mu_t)\mathrm{d}t + \sigma_t(\hat{X}_t^{\mu_0} + \eta_t^\mu)\mathrm{d}\tilde{W}_t \\ &\quad + \frac{(\hat{Y}_t^\nu + \eta_{t_0}^\nu) - (\hat{X}_t^{\mu_0} + \eta_{t_0}^\mu)}{\xi_t}\mathrm{d}t, \quad t \in [0, s\wedge\tau_n],\end{aligned}$$

and

$$\mathrm{d}\hat{Y}_t^\nu = b_t(\hat{Y}_t^\nu + \eta_t^\nu, \nu_t)\mathrm{d}t + \sigma_t(\hat{Y}_t^\nu + \eta_t^\nu)\mathrm{d}\tilde{W}_t, \qquad t \in [0, s\wedge\tau_n],\ \hat{Y}_0^\nu = X_0^{\nu_0}. \tag{2.23}$$

By Itô's formula, we obtain

$$\begin{aligned}&\mathrm{d}\frac{|(\hat{Y}_t^\nu + \eta_{t_0}^\nu) - (\hat{X}_t^{\mu_0} + \eta_{t_0}^\mu)|^2}{\xi_t}\mathrm{d}t \\ &= -\frac{\xi_t'|(\hat{Y}_t^\nu + \eta_{t_0}^\nu) - (\hat{X}_t^{\mu_0} + \eta_{t_0}^\mu)|^2}{\xi_t^2}\mathrm{d}t \\ &\quad + \frac{2\langle b_t(\hat{Y}_t^\nu + \eta_t^\nu, \nu_t) - b_t(\hat{X}_t^{\mu_0} + \eta_t^\mu, \mu_t), (\hat{Y}_t^\nu + \eta_{t_0}^\nu) - (\hat{X}_t^{\mu_0} + \eta_{t_0}^\mu)\rangle}{\xi_t}\mathrm{d}t\end{aligned} \tag{2.24}$$

$$+\frac{2\langle[\sigma_t(\hat{Y}_t^\nu+\eta_t^\nu)-\sigma_t(\hat{X}_t^{\mu_0}+\eta_t^\mu)]\mathrm{d}\tilde{W}_t,(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)\rangle}{\xi_t}$$
$$-2\frac{|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|^2}{\xi_t^2}\mathrm{d}t$$
$$+\frac{\|\sigma_t(\hat{Y}_t^\nu+\eta_t^\nu)-\sigma_t(\hat{X}_t^{\mu_0}+\eta_t^\mu)\|_{HS}^2}{\xi_t}\mathrm{d}t,\quad t\in[0,s\wedge\tau_n].$$

(B) implies that

$$\begin{aligned}&\frac{2\langle b_t(\hat{Y}_t^\nu+\eta_t^\nu,\nu_t)-b_t(\hat{X}_t^{\mu_0}+\eta_t^\mu,\mu_t),(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)\rangle}{\xi_t}\\
&\leq\frac{2K|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|^2}{\xi_t}\\
&+\frac{2K[\mathbb{W}_2(\nu_t,\mu_t)+|\eta_t^\nu-\eta_{t_0}^\nu-(\eta_t^\mu-\eta_{t_0}^\mu)|]|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|}{\xi_t}\\
&\leq\frac{[2K\xi_t+\frac{1}{2}]|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|^2}{\xi_t^2}+2K^2[\mathbb{W}_2(\nu_t,\mu_t)+|\eta_t^\nu-\eta_{t_0}^\nu-(\eta_t^\mu-\eta_{t_0}^\mu)|]^2,\end{aligned}$$

and

$$\begin{aligned}&\frac{\|\sigma_t(\hat{Y}_t^\nu+\eta_t^\nu)-\sigma_t(\hat{X}_t^{\mu_0}+\eta_t^\mu)\|_{HS}^2}{\xi_t}\\
&\leq\frac{2K\xi_t|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|^2}{\xi_t^2}+\frac{2K|\eta_t^\nu-\eta_{t_0}^\nu-(\eta_t^\mu-\eta_{t_0}^\mu)|^2}{\xi_t}.\end{aligned}$$

This together with (2.24) yields that

$$\begin{aligned}&\mathrm{d}\frac{|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|^2}{\xi_t}\\
&\leq\frac{[-\xi_t'+4K\xi_t-\frac{3}{2}]|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|^2}{\xi_t^2}\mathrm{d}t\\
&+2K^2[\mathbb{W}_2(\nu_t,\mu_t)+|\eta_t^\nu-\eta_{t_0}^\nu-(\eta_t^\mu-\eta_{t_0}^\mu)|]^2\mathrm{d}t\\
&+\frac{2K|\eta_t^\nu-\eta_{t_0}^\nu-(\eta_t^\mu-\eta_{t_0}^\mu)|^2}{\xi_t}\mathrm{d}t\\
&+\frac{2\langle[\sigma_t(\hat{Y}_t^\nu+\eta_t^\nu)-\sigma_t(X_t^\mu+\eta_t^\mu)]\mathrm{d}\tilde{W}_t,(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^\mu+\eta_{t_0}^\mu)\rangle}{\xi_t},\quad t\in[0,s\wedge\tau_n].\end{aligned}$$

Combining this with (2.19), we deduce

$$\mathbb{E}_{\mathbb{Q}^B_{s\wedge\tau_n}}\int_0^{s\wedge\tau_n}\frac{|(\hat{Y}_t^\nu+\eta_{t_0}^\nu)-(\hat{X}_t^{\mu_0}+\eta_{t_0}^\mu)|^2}{\xi_t^2}\mathrm{d}t$$

$$\leq \frac{2\mathbb{E}^B|(\hat{Y}_0^\nu + \eta_{t_0}^\nu) - (\hat{X}_0^{\mu_0} + \eta_{t_0}^\mu)|^2}{\xi_0} + 8K^2 \int_0^s \mathbb{W}_2(\nu_t, \mu_t)^2 \mathrm{d}t$$
$$+ \int_0^s 8K^2 |\eta_t^\nu - \eta_{t_0}^\nu - (\eta_t^\mu - \eta_{t_0}^\mu)|^2 \mathrm{d}t + \int_0^s \frac{4K|\eta_t^\nu - \eta_{t_0}^\nu - (\eta_t^\mu - \eta_{t_0}^\mu)|^2}{\xi_t} \mathrm{d}t.$$

As a result, it follows from (2.22) and **(B)** that

$$\mathbb{E}^B[R_{s\wedge\tau_n} \log R_{s\wedge\tau_n}] = \mathbb{E}_{\mathbb{Q}^B_{s\wedge\tau_n}}[\log R_{s\wedge\tau_n}] = \frac{1}{2}\mathbb{E}_{\mathbb{Q}^B_{s\wedge\tau_n}} \int_0^{s\wedge\tau_n} |\beta_t|^2 \mathrm{d}t$$
$$\leq \lambda^{-1} \frac{\mathbb{E}^B|(\hat{Y}_0^\nu + \eta_{t_0}^\nu) - (\hat{X}_0^\mu + \eta_{t_0}^\mu)|^2}{\xi_0} + 4\lambda^{-1}K^2 \int_0^s \mathbb{W}_2(\nu_t, \mu_t)^2 \mathrm{d}t$$
$$+ 4\lambda^{-1}K^2 \int_0^s |\eta_t^\nu - \eta_{t_0}^\nu - (\eta_t^\mu - \eta_{t_0}^\mu)|^2 \mathrm{d}t + 2\lambda^{-1}K \int_0^s \frac{|\eta_t^\nu - \eta_{t_0}^\nu - (\eta_t^\mu - \eta_{t_0}^\mu)|^2}{\xi_t} \mathrm{d}t.$$

This means that $\{R_{s\wedge\tau_n}\}_{n\geq 1}$ is a uniform integrable martingale under $\mathbb{P}^B$ so that we derive from the martingale convergence theorem and Fatou's lemma that

$$\begin{aligned} &\frac{1}{2}\mathbb{E}_{\mathbb{Q}^B_s} \int_0^s |\beta_t|^2 \mathrm{d}t = \mathbb{E}^B[R_s \log R_s] \\ &\leq \lambda^{-1} \frac{\mathbb{E}^B|(\hat{Y}_0^\nu + \eta_{t_0}^\nu) - (\hat{X}_0^\mu + \eta_{t_0}^\mu)|^2}{\xi_0} + 4\lambda^{-1}K^2 \int_0^s \mathbb{W}_2(\nu_t, \mu_t)^2 \mathrm{d}t \\ &+ 4\lambda^{-1}K^2 \int_0^s |\eta_t^\nu - \eta_{t_0}^\nu - (\eta_t^\mu - \eta_{t_0}^\mu)|^2 \mathrm{d}t \\ &+ 2\lambda^{-1}K \int_0^s \frac{|\eta_t^\nu - \eta_{t_0}^\nu - (\eta_t^\mu - \eta_{t_0}^\mu)|^2}{\xi_t} \mathrm{d}t, \quad s \in [0, t_0). \end{aligned} \tag{2.25}$$

This combined with (2.17) implies that $\{R_s\}_{s\in[0,t_0]}$ is a uniform integrable martingale under $\mathbb{P}^B$ and $\mathbb{Q}^B_{t_0}$-a.s. $\hat{Y}_{t_0}^\nu + \eta_{t_0}^\nu = \hat{X}_{t_0}^{\mu_0} + \eta_{t_0}^\mu$ in view of the definition of β_t and ξ_t and (2.25) for $s = t_0$ due to Fatou's lemma. It follows from the weak uniqueness to (2.23) on $[0, t_0]$ due to **(B)** that

$$\mathscr{L}^{\mathbb{Q}^B_{t_0}}_{\hat{Y}_t^\nu} = \mathscr{L}^{\mathbb{P}^B}_{\hat{X}_t^{\nu_0}}, \quad t \in [0, t_0], \tag{2.26}$$

where $\hat{X}_t^{\nu_0} = X_t^{\nu_0} - \eta_t^\nu, t \in [0, T]$ solves

$$\mathrm{d}\hat{X}_t^{\nu_0} = b_t(\hat{X}_t^{\nu_0} + \eta_t^\nu, \nu_t)\mathrm{d}t + \sigma_t(\hat{X}_t^{\nu_0} + \eta_t^\nu)\mathrm{d}W_t, \quad t \in [0, T].$$

Let $Y_t^\nu := \hat{Y}_t^\nu + \eta_t^\nu, t \in [0, t_0]$. Since $\eta_{t_0}^\nu$ is measurable with respect to $\mathscr{F}^B_{t_0}$, it follows from (2.26) that

$$\mathscr{L}^{\mathbb{Q}^B_{t_0}}_{Y_{t_0}^\nu} = \mathscr{L}^{\mathbb{Q}^B_{t_0}}_{\hat{Y}_{t_0} + \eta_{t_0}^\nu} = \mathscr{L}^{\mathbb{P}^B}_{\hat{X}_{t_0}^{\nu_0} + \eta_{t_0}^\nu} = \mathscr{L}^{\mathbb{P}^B}_{X_{t_0}^{\nu_0}}. \tag{2.27}$$

Again by (2.25) for $s=t_0$, (2.27), the fact $\mathbb{Q}_{t_0}^B$-a.s. $Y_{t_0}^\nu=\hat Y_{t_0}^\nu+\eta_{t_0}^\nu=\hat X_{t_0}^{\mu_0}+\eta_{t_0}^\mu=X_{t_0}^{\mu_0}$ and Young's inequality, we conclude that

$$\begin{aligned}&\mathbb{E}^B[\log f(X_{t_0}^{\nu_0})]=\mathbb{E}_{\mathbb{Q}_{t_0}^B}[\log f(Y_{t_0}^\nu)]\\&=\mathbb{E}^B[R_{t_0}\log f(X_{t_0}^{\mu_0})]\le\log\mathbb{E}^B[f(X_{t_0}^{\mu_0})]+\mathbb{E}^B[R_{t_0}\log R_{t_0}]\\&\le\log\mathbb{E}^B[f(X_{t_0}^{\mu_0})]+\lambda^{-1}\frac{\mathbb{E}^B|(\hat Y_0^\nu+\eta_{t_0}^\nu)-(\hat X_0^\mu+\eta_{t_0}^\mu)|^2}{\xi_0}+4\lambda^{-1}K^2\int_0^{t_0}\mathbb{W}_2(\nu_t,\mu_t)^2\mathrm{d}t\\&+4\lambda^{-1}K^2\int_0^{t_0}|\eta_t^\nu-\eta_{t_0}^\nu-(\eta_t^\mu-\eta_{t_0}^\mu)|^2\mathrm{d}t\\&+2\lambda^{-1}K\int_0^{t_0}\frac{|\eta_t^\nu-\eta_{t_0}^\nu-(\eta_t^\mu-\eta_{t_0}^\mu)|^2}{\xi_t}\mathrm{d}t=:\log\mathbb{E}^B[f(X_{t_0}^{\mu_0})]+\Phi_{t_0},\quad 0<f\in\mathscr{B}_b(\mathbb{R}^d).\end{aligned}$$

So, by [32, Theorem 1.4.2(2)], Lemma 1.1, (2.16), (2.17) and (2.22), we find a constant $c>0$ such that

$$\mathbb{E}\{\mathrm{Ent}(\mathscr{L}_{X_{t_0}^{\nu_0}|\mathscr{F}_{t_0}^B}|\mathscr{L}_{X_{t_0}^{\mu_0}|\mathscr{F}_{t_0}^B})\}\le\mathbb{E}\Phi_{t_0}\le c\left\{\frac{4K}{1-\mathrm{e}^{-4Kt_0}}+\int_0^{t_0}\frac{4Kt}{1-\mathrm{e}^{-4Kt}}\mathrm{d}t\right\}\mathbb{W}_2(\mu_0,\nu_0)^2.$$

This together with the fact

$$\mathrm{Ent}(\mathscr{L}_{X_t^{\nu_0}}|\mathscr{L}_{X_t^{\mu_0}})\le\mathbb{E}\{\mathrm{Ent}(\mathscr{L}_{X_t^{\nu_0}|\mathscr{F}_t^B}|\mathscr{L}_{X_t^{\mu_0}|\mathscr{F}_t^B})\}\tag{2.28}$$

implies (2.14), which combined with [32, Theorem 1.4.2(2)] completes the proof. □

2.2 Conditional propagation of chaos

Fix $T>0$. Let $(\Omega^i,\mathscr{F}^i,(\mathscr{F}_t^i)_{t\ge0},\mathbb{P}^i),i=0,1$ and $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\ge0},\mathbb{P})$ be defined in Section 1. Let $N\ge1$ be an integer, $(W_t^i)_{1\le i\le N}$ be N independent d_W-dimensional Brownian motions on $(\Omega^1,\mathscr{F}^1,(\mathscr{F}_t^1)_{t\ge0},\mathbb{P}^1)$, B_t be a d_B-dimensional Brownian motion on $(\Omega^0,\mathscr{F}^0,(\mathscr{F}_t^0)_{t\ge0},\mathbb{P}^0)$, and $(X_0^i)_{1\le i\le N}$ be i.i.d. $(\Omega^1,\mathscr{F}_0^1)$-measurable $\mathbb{R}^d$-valued random variables. Let $b:[0,T]\times\mathbb{R}^d\times\mathscr{P}(\mathbb{R}^d)\to\mathbb{R}^d$, $\sigma:[0,T]\times\mathbb{R}^d\times\mathscr{P}(\mathbb{R}^d)\to\mathbb{R}^d\otimes\mathbb{R}^{d_W}$ and $\tilde\sigma:[0,T]\to\mathbb{R}^d\otimes\mathbb{R}^{d_B}$ be measurable. Consider conditional McKean-Vlasov SDEs

$$\mathrm{d}X_t^i=b_t(X_t^i,\mathscr{L}_{X_t^i|\mathscr{F}_t^B})\mathrm{d}t+\sigma_t(X_t^i,\mathscr{L}_{X_t^i|\mathscr{F}_t^B})\mathrm{d}W_t^i+\tilde\sigma_t\mathrm{d}B_t,\quad 1\le i\le N,\tag{2.29}$$

and the mean field interacting particle system with common noise

$$\mathrm{d}X_t^{i,N}=b_t(X_t^{i,N},\hat\mu_t^N)\mathrm{d}t+\sigma_t(X_t^{i,N},\hat\mu_t^N)\mathrm{d}W_t^i+\tilde\sigma_t\mathrm{d}B_t,\quad 1\le i\le N,\tag{2.30}$$

where for any $1\le i\le N$, $X_0^{i,N}$ is an $(\Omega^1,\mathscr{F}_0^1)$-measurable $\mathbb{R}^d$-valued random variable, the distribution of $(X_0^{1,N},X_0^{2,N},\cdots,X_0^{N,N})$ is exchangeable and $\hat\mu_t^N$ is the empirical distribution of $(X_t^{i,N})_{1\le i\le N}$, i.e.

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

Theorem 2.3. *Assume that there exists a constant $K \geq 0$ such that*

$$\begin{aligned}&|b_t(0,\delta_0)| + \|\sigma_t(0,\delta_0)\|_{HS} + \|\tilde{\sigma}_t\|_{HS} \leq K, \quad t\in[0,T],\\ &|b_t(x,\mu)-b_t(y,\nu)| + \|\sigma_t(x,\mu)-\sigma_t(y,\nu)\|_{HS}\\ &\leq K(|x-y| + \mathbb{W}_2(\mu,\nu)), \quad t\in[0,T], x,y\in\mathbb{R}^d, \mu,\nu\in\mathscr{P}_2(\mathbb{R}^d).\end{aligned} \tag{2.31}$$

Then the following assertions hold.

(1) Assume that $\mathbb{E}|X_0^{1,N}|^2 < \infty$ and $\mathbb{E}|X_0^1|^q < \infty$ for some $q > 2$. Then there exists a constant $C > 0$ depending only on d, T and $\mathbb{E}|X_0^1|^q$ such that

$$\begin{aligned}&\sup_{t\in[0,T]} \mathbb{E}\mathbb{W}_2(\hat{\mu}_t^N, \mathscr{L}^{\mathbb{P}}_{X_t^1|\mathscr{F}_t^B})^2\\ &\leq C\frac{1}{N}\mathbb{W}_2(\mathscr{L}_{(X_0^1,X_0^2,\cdots,X_0^N)}, \mathscr{L}_{(X_0^{1,N},X_0^{2,N},\cdots,X_0^{N,N})})^2 + CR_{d,q}(N),\end{aligned} \tag{2.32}$$

and

$$\begin{aligned}&\mathbb{E}\mathbb{W}_2(\mathscr{L}^{\mathbb{P}}_{(X_t^1,X_t^2,\cdots,X_t^k)|\mathscr{F}_t^B}, \mathscr{L}^{\mathbb{P}}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{k,N})|\mathscr{F}_t^B})^2\\ &\leq C\frac{k}{N}\mathbb{W}_2(\mathscr{L}_{(X_0^1,X_0^2,\cdots,X_0^N)}, \mathscr{L}_{(X_0^{1,N},X_0^{2,N},\cdots,X_0^{N,N})})^2 + CkR_{d,q}(N), \quad 1\leq k\leq N,\end{aligned} \tag{2.33}$$

where

$$R_{d,q}(N) = \begin{cases} N^{-\frac{1}{2}} + N^{-\frac{q-2}{q}}, & d<4, q\neq 4,\\ N^{-\frac{1}{2}}\log(1+N) + N^{-\frac{q-2}{q}}, & d=4, q\neq 4,\\ N^{-\frac{2}{d}} + N^{-\frac{q-2}{q}}, & d>4, q\neq \frac{d}{d-2}.\end{cases}$$

(2) If in addition, $\sigma_t(x,\mu)$ does not depend on μ and $\lambda^{-1} \geq \sigma\sigma^* \geq \lambda$ for some $\lambda\in(0,1]$, then for any $k \geq 1$ and $k \leq N$ and $t \in (0,T]$, it holds

$$\begin{aligned}&\mathbb{E}\mathrm{Ent}(\mathscr{L}^{\mathbb{P}}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{k,N})|\mathscr{F}_t^B} | \mathscr{L}^{\mathbb{P}}_{(X_t^1,X_t^2,\cdots,X_t^k)|\mathscr{F}_t^B})\\ &\leq CkR_{d,q}(N) + \frac{C}{1-\mathrm{e}^{-(K^2+2K)t}}\frac{k}{N}\mathbb{W}_2(\mathscr{L}_{(X_0^1,X_0^2,\cdots,X_0^N)}, \mathscr{L}_{(X_0^{1,N},X_0^{2,N},\cdots,X_0^{N,N})})^2.\end{aligned} \tag{2.34}$$

Proof. (1) Let $\mathbb{P}^B, \mathbb{E}^B$ be in (2.21). It is standard to derive from (2.31) that

$$\mathbb{E}^B[\sup_{t\in[0,T]}|X_t^1|^q] < c_1\left(1 + \mathbb{E}|X_0^1|^q + \sup_{s\in[0,T]}\left|\int_0^s \tilde{\sigma}_t \mathrm{d}B_t\right|^q\right) \tag{2.35}$$

for some constant $c_1 > 0$ depending on q, T. Denote $\mu_t^i = \mathscr{L}^{\mathbb{P}}_{X_t^i|\mathscr{F}_t^B}, i \geq 1$. Since (2.29) is well-posed under (2.31) due to Lemma 1.1, μ_t^i does not depend on i and we write $\mu_t = \mu_t^i, 1 \leq i \leq N$. Letting $\tilde{\mu}_t^N = \frac{1}{N}\sum_{j=1}^N \delta_{X_t^j}$, we obtain from the triangle inequality that

$$\begin{aligned}\mathbb{W}_2(\hat{\mu}_s^N, \mu_s) &\leq \mathbb{W}_2(\hat{\mu}_s^N, \tilde{\mu}_s^N) + \mathbb{W}_2(\tilde{\mu}_s^N, \mu_s)\\ &\leq \left(\frac{1}{N}\sum_{i=1}^N |X_s^{i,N} - X_s^i|^2\right)^{\frac{1}{2}} + \mathbb{W}_2(\tilde{\mu}_s^N, \mu_s).\end{aligned} \tag{2.36}$$

By BDG's inequality, (2.31) and (2.36), there exists a constant $c_2 > 0$ such that

$$\sum_{i=1}^{N}\mathbb{E}^B\sup_{s\in[0,t]}|X_s^{i,N}-X_s^i|^2 \le c_2\sum_{i=1}^{N}\mathbb{E}|X_0^{i,N}-X_0^i|^2 + c_2\int_0^t \mathbb{E}^B\sum_{i=1}^{N}|X_s^{i,N}-X_s^i|^2\mathrm{d}s + c_2N\mathbb{E}^B\int_0^t \mathbb{W}_2(\tilde{\mu}_s^N,\mu_s)^2\mathrm{d}s.$$

By Gronwall's inequality, we can find a constant $c_3 > 0$ such that

$$\sum_{i=1}^{N}\mathbb{E}^B\sup_{s\in[0,t]}|X_s^{i,N}-X_s^i|^2 \le c_3\sum_{i=1}^{N}\mathbb{E}|X_0^{i,N}-X_0^i|^2 + c_3N\mathbb{E}^B\int_0^t \mathbb{W}_2(\tilde{\mu}_s^N,\mu_s)^2\mathrm{d}s. \tag{2.37}$$

By [11, Theorem 1] for $p = 2$ and (2.35), there exists a constant $C_0 > 0$ depending only on q, d such that

$$\begin{aligned}\mathbb{E}^B\mathbb{W}_2(\tilde{\mu}_s^N,\mu_s)^2 &\le C_0\left(\mathbb{E}^B[\sup_{t\in[0,T]}|X_t^1|^q]\right)^{\frac{2}{q}} R_{d,q}(N)\\ &\le C_0c_1^{\frac{2}{q}}\left(1+\mathbb{E}|X_0^1|^q+\sup_{s\in[0,T]}\left|\int_0^s\tilde{\sigma}_t\mathrm{d}B_t\right|^q\right)^{\frac{2}{q}} R_{d,q}(N),\quad s\in[0,T].\end{aligned} \tag{2.38}$$

So, we derive (2.33) by combining with (2.37) and the fact

$$\begin{aligned}&\mathbb{W}_2(\mathscr{L}_{(X_t^1,X_t^2,\cdots,X_t^k)|\mathscr{F}_t^B},\mathscr{L}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{k,N})|\mathscr{F}_t^B})^2\\ &\le \frac{k}{N}\mathbb{W}_2(\mathscr{L}_{(X_t^1,X_t^2,\cdots,X_t^N)|\mathscr{F}_t^B},\mathscr{L}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{N,N})|\mathscr{F}_t^B})^2.\end{aligned}$$

Finally, (2.38) together with (2.37) and (2.36) derives

$$\begin{aligned}\mathbb{E}^B\mathbb{W}_2(\hat{\mu}_s^N,\mu_s)^2 &\le C_1\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}|X_0^{i,N}-X_0^i|^2\\ &\quad + C_1\left(1+\mathbb{E}|X_0^1|^q+\sup_{s\in[0,T]}\left|\int_0^s\tilde{\sigma}_t\mathrm{d}B_t\right|^q\right)^{\frac{2}{q}} R_{d,q}(N),\quad s\in[0,T],\end{aligned}$$

for some constant $C_1 > 0$ depending on d, T, which yields (2.32) by taking expectation with respect to $\mathbb{P}$.

(2) **(Step (i))** We first assume that b is bounded. Define

$$\mathbb{P}^{B,0} := \mathbb{P}(\,\cdot\,|\mathscr{F}_T^B\bigvee\mathscr{F}_0),\quad \mathbb{E}^{B,0} := \mathbb{E}(\,\cdot\,|\mathscr{F}_T^B\bigvee\mathscr{F}_0). \tag{2.39}$$

Consider

$$\mathrm{d}\bar{X}_t^i = b_t(\bar{X}_t^i,\mu_t)\mathrm{d}t+\sigma_t(\bar{X}_t^i)\mathrm{d}W_t^i+\tilde{\sigma}_t\mathrm{d}B_t,\quad t\in[0,T], \bar{X}_0^i = X_0^{i,N}, 1\le i\le N. \tag{2.40}$$

Rewrite (2.40) as

$$\mathrm{d}\bar{X}_t^i = b_t(\bar{X}_t^i, \frac{1}{N}\sum_{i=1}^N \delta_{\bar{X}_t^i})\mathrm{d}t + \sigma_t(\bar{X}_t^i)\mathrm{d}\hat{W}_t^i + \tilde{\sigma}_t \mathrm{d}B_t, \quad t \in [0,T], 1 \le i \le N.$$

where for $1 \le i \le N$,

$$\begin{aligned}
&\hat{W}_t^i := W_t^i - \int_0^t \gamma_s^i \mathrm{d}s,\\
&\gamma_t^i := [\sigma_t^*(\sigma_t\sigma_t^*)^{-1}](\bar{X}_t^i)\left(b_t(\bar{X}_t^i, \frac{1}{N}\sum_{i=1}^N \delta_{\bar{X}_t^i}) - b_t(\bar{X}_t^i, \mu_t)\right), \quad t \in [0,T].
\end{aligned}$$

Fix $t_0 \in [0,T]$. Let

$$\begin{aligned}
&\gamma_t = (\gamma_t^1, \gamma_t^2, \cdots, \gamma_t^N), \quad \hat{W}_t = (\hat{W}_t^1, \hat{W}_t^2, \cdots, \hat{W}_t^N),\\
&R_t := \mathrm{e}^{\int_0^t \sum_{i=1}^N \langle \gamma_r^i, \mathrm{d}W_r^i\rangle - \frac{1}{2}\int_0^t \sum_{i=1}^N |\gamma_r^i|^2 \mathrm{d}r},\\
&\mathbb{Q}_t^{B,0} := R_t \mathbb{P}^{B,0}, \quad t \in [0,t_0].
\end{aligned}$$

Since b is bounded, we can apply Girsanov's theorem to conclude that $\{\hat{W}_t\}_{t\in[0,t_0]}$ is an $(N \times d_W)$-dimensional Brownian motion under the weighted conditional probability $\mathbb{Q}_{t_0}^{B,0}$. So, we have

$$\mathscr{L}^{\mathbb{Q}_{t_0}^{B,0}}_{(\{\bar{X}_t^i\}_{1\le i\le N}, B_t)_{t\in[0,t_0]}} = \mathscr{L}^{\mathbb{P}^{B,0}}_{(\{X_t^{i,N}\}_{1\le i\le N}, B_t)_{t\in[0,t_0]}}.$$

This gives

$$\mathbb{E}_{\mathbb{Q}_{t_0}^{B,0}} \int_0^{t_0} \mathbb{W}_2(\mu_t, \frac{1}{N}\sum_{i=1}^N \delta_{\bar{X}_t^i})^2 \mathrm{d}t = \mathbb{E}^{B,0}\int_0^{t_0} \mathbb{W}_2(\mu_t, \hat{\mu}_t^N)^2 \mathrm{d}t,$$

which together with (2.31), $\lambda \le \sigma\sigma^* \le \lambda^{-1}$ and Young's inequality implies that

$$\begin{aligned}
&\mathbb{E}^{B,0} \log F(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{N,N})\\
&\le \log \mathbb{E}^{B,0}[F(\bar{X}_{t_0}^1, \bar{X}_{t_0}^2, \cdots, \bar{X}_{t_0}^N)] + \mathbb{E}^{B,0}(R_{t_0}\log R_{t_0})\\
&\le \log \mathbb{E}^{B,0}[F(\bar{X}_{t_0}^1, \bar{X}_{t_0}^2, \cdots, \bar{X}_{t_0}^N)] + \frac{1}{2}\sum_{i=1}^N \mathbb{E}_{\mathbb{Q}_{t_0}^{B,0}} \int_0^{t_0} |\gamma_t^i|^2 \mathrm{d}t\\
&\le \log \mathbb{E}^{B,0}[F(\bar{X}_{t_0}^1, \bar{X}_{t_0}^2, \cdots, \bar{X}_{t_0}^N)]\\
&+ c_1 N \mathbb{E}^{B,0}\int_0^{t_0} \mathbb{W}_2(\mu_t, \hat{\mu}_t^N)^2 \mathrm{d}t, \quad 0 < F \in \mathscr{B}_b((\mathbb{R}^d)^N).
\end{aligned} \tag{2.41}$$

for some constant $c_1 > 0$. On the other hand, let

$$b_t^{\mu,B}(x) = b_t\left(x + \int_0^t \tilde{\sigma}_s \mathrm{d}B_s, \mu_t\right), \quad \sigma_t^B(x) = \sigma_t\left(x + \int_0^t \tilde{\sigma}_s \mathrm{d}B_s\right), \quad t \in [0,T], x \in \mathbb{R}^d.$$

Consider

$$\mathrm{d}Y_t^i = b_t^{\mu,B}(Y_t^i) + \sigma_t(Y_t^i)\mathrm{d}W_t^i, \quad Y_0^i = X_0^i, 1 \le i \le N,$$

and

$$\mathrm{d}\bar{Y}_t^i = b_t^{\mu,B}(\bar{Y}_t^i) + \sigma_t(\bar{Y}_t^i)\mathrm{d}W_t^i, \quad \bar{Y}_0^i = X_0^{i,N}, 1 \le i \le N.$$

By [32, Theorem 3.4.1] and (2.31), for large enough $p > 1$, we get Wang's Harnack inequality with power p for some constant $c(p) > 0$:

$$\begin{aligned}\left(\mathbb{E}^{B,0}[\bar{F}(\bar{Y}_t^1, \bar{Y}_t^2, \cdots, \bar{Y}_t^N)]\right)^p &\le \mathbb{E}^{B,0}[\bar{F}(Y_t^1, Y_t^2, \cdots, Y_t^N)^p] \\ &\times \exp\left\{\frac{c(p)\sum_{i=1}^N |X_0^{i,N} - X_0^i|^2}{1 - \mathrm{e}^{-(K^2+2K)t}}\right\}, \quad \bar{F} \in \mathscr{B}_b^+((\mathbb{R}^d)^N), t \in (0, T].\end{aligned}$$

In view of $\bar{X}_t^i = \bar{Y}_t^i + \int_0^t \tilde{\sigma}_s \mathrm{d}B_s$ and $X_t^i = Y_t^i + \int_0^t \tilde{\sigma}_s \mathrm{d}B_s$, we conclude that

$$\begin{aligned}\left(\mathbb{E}^{B,0}[F(\bar{X}_t^1, \bar{X}_t^2, \cdots, \bar{X}_t^N)]\right)^p &\le \mathbb{E}^{B,0}[F(X_t^1, X_t^2, \cdots, X_t^N)^p] \\ &\times \exp\left\{\frac{c(p)\sum_{i=1}^N |X_0^{i,N} - X_0^i|^2}{1 - \mathrm{e}^{-(K^2+2K)t}}\right\}, \quad F \in \mathscr{B}_b^+((\mathbb{R}^d)^N), t \in (0, T].\end{aligned}$$

This together with (2.41), [32, Theorem 1.4.2(1)-(2)] and [25, Lemma 2.1] implies that

$$\begin{aligned}&\mathbb{E}^{B,0} \log F(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{N,N}) \\ &\le \log \mathbb{E}^{B,0}[F(X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^N)] + c_1 pN\mathbb{E}^{B,0} \int_0^{t_0} \mathbb{W}_2(\mu_t, \hat{\mu}_t^N)^2 \mathrm{d}t \\ &+ \frac{c(p)\sum_{i=1}^N |X_0^{i,N} - X_0^i|^2}{1 - \mathrm{e}^{-(K^2+2K)t_0}}, \quad 0 < F \in \mathscr{B}_b((\mathbb{R}^d)^N).\end{aligned} \tag{2.42}$$

Taking expectation with respect to $\mathbb{E}^B$ on both sides and using Jensen's inequality, we derive

$$\begin{aligned}&\mathbb{E}^{B} \log F(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{N,N}) \\ &\le \log \mathbb{E}^{B}[F(X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^N)] + c_1 pN\mathbb{E}^{B} \int_0^{t_0} \mathbb{W}_2(\mu_t, \hat{\mu}_t^N)^2 \mathrm{d}t \\ &+ \frac{c(p)\sum_{i=1}^N \mathbb{E}|X_0^{i,N} - X_0^i|^2}{1 - \mathrm{e}^{-(K^2+2K)t_0}}, \quad 0 < F \in \mathscr{B}_b((\mathbb{R}^d)^N).\end{aligned} \tag{2.43}$$

For any $1 \le k \le N$ and $0 < f \in \mathscr{B}_b((\mathbb{R}^d)^k)$, take

$$F_f(x_1, x_2, \cdots, x_{\lfloor \frac{N}{k} \rfloor k}) = \prod_{i=0}^{\lfloor \frac{N}{k} \rfloor - 1} f(x_{ik+1}, x_{ik+2}, \cdots, x_{ik+k}),$$

where $\lfloor \frac{N}{k} \rfloor$ stands for the integer part of $\frac{N}{k}$. Since $(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{N,N})$ is exchangeable and $X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^N$ are i.i.d. under $\mathbb{P}^B$ and $\lfloor \frac{N}{k} \rfloor^{-1} \leq \frac{2k}{N}, 1 \leq k \leq N$, we derive from (2.43) for $F = F_f$ that

$$\begin{aligned}
&\mathbb{E}^B \log f(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{k,N}) \\
&\leq \log \mathbb{E}^B[f(X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^k)] + \frac{2c(p)k}{1 - \mathrm{e}^{-(K^2+2K)t_0}} \frac{1}{N} \sum_{i=1}^N \mathbb{E}|X_0^{i,N} - X_0^i|^2 \\
&\quad + 2c_1 pk \mathbb{E}^B \int_0^{t_0} \mathbb{W}_2(\mu_t, \hat{\mu}_t^N)^2 \mathrm{d}t, \quad 0 < f \in \mathscr{B}_b((\mathbb{R}^d)^k).
\end{aligned} \tag{2.44}$$

Again using [32, Theorem 1.4.2(2)], we derive (2.34) from (2.32) and (2.44).

(Step (ii)) In general, let $b^{(n)} = (-n \vee b^i \wedge n)_{1 \leq i \leq d}, n \geq 1$. Noting that (2.31) holds for $b^{(n)}$ in place of b, (2.34) follows from **Step (i)** and an approximation technique. □

Remark 2.4. *(1) Note that in the present case, the coefficients are only assumed to be Lipschitz continuous in $\mathbb{W}_2$-distance with respect to the measure variable so that [11, Theorem 1] for $p = 2$ is used to estimate the convergence rate of conditional propagation of chaos, which depends on the dimension d and seems a little complicated. One can also refer to [5, Theorem 2.12] for the case $q > 4$ and $X_0^{i,N} = X_0^i, 1 \leq i \leq N$. However, if we only consider the special case:*

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

for some Lipschitz continuous functions $\bar{b}, \bar{\sigma}$ uniformly in time variable t, the convergence rate in Theorem 2.3 and Theorem 3.2 below can be improved to be $R_{d,q}(N) = \frac{1}{N}$ and $q = 2$.

(2) In Theorem 2.3(2), the coefficients before the private noise can depend on the spatial variable and the initial distribution of interacting particle system (2.30) *is allowed to be singular with that of the conditional McKean-Vlasov SDEs* (2.29) *since* (2.34) *only involves in $\frac{1}{N} \mathbb{W}_2(\mathscr{L}_{(X_0^1, X_0^2, \cdots, X_0^N)}, \mathscr{L}_{(X_0^{1,N}, X_0^{2,N}, \cdots, X_0^{N,N})})^2$. See also [2, 15, 16] for the quantitative propagation of chaos in relative entropy by the entropy method in the additive noise case and under the assumption*

$$\lim_{N \to \infty} \frac{\mathrm{Ent}(\mathscr{L}_{(X_0^{1,N}, X_0^{2,N}, \cdots, X_0^{N,N})} | \mathscr{L}_{(X_0^1, X_0^2, \cdots, X_0^N)})}{N} = 0.$$

3 Conditional distribution dependent stochastic Hamiltonian system

In this part, we consider conditional distribution dependent stochastic Hamiltonian system with additive noise, which is a type of degenerate model. More precisely, we consider

$$\begin{cases} \mathrm{d}X_t^{(1)} = \{AX_t^{(1)} + MX_t^{(2)}\} \mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = b_t(X_t, \mathscr{L}_{X_t | \mathscr{F}_t^B}) \mathrm{d}t + \sigma_t \mathrm{d}W_t + \tilde{\sigma}_t(\mathscr{L}_{X_t | \mathscr{F}_t^B}) \mathrm{d}B_t, \quad t \in [0, T], \end{cases} \tag{3.1}$$

where $X_t = (X_t^{(1)}, X_t^{(2)})$, W_t, B_t are given in Section 1, $b : [0,T] \times \mathbb{R}^{m+d} \times \mathscr{P}(\mathbb{R}^{m+d}) \to \mathbb{R}^d$, $\sigma : [0,T] \to \mathbb{R}^d \otimes \mathbb{R}^{d_W}$, $\tilde\sigma : [0,T] \times \mathscr{P}(\mathbb{R}^{m+d}) \to \mathbb{R}^d \otimes \mathbb{R}^{d_B}$ are measurable and bounded on bounded sets, A is an $m \times m$ matrix and M is an $m \times d$ matrix.

3.1 Log-Haranck inequality

To establish the log-Harnack inequality, we make the following assumption.

(C) $(\sigma_t\sigma_t^*)^{-1}$ is bounded in $t \in [0,T]$, and there exist constants $K, \tilde K > 0$ such that

$$|b_t(x,\mu) - b_t(y,\nu)| \le K(|x-y| + \mathbb{W}_2(\mu,\nu)),$$
$$\|\tilde\sigma_t(\mu) - \tilde\sigma_t(\nu)\|_{HS} \le \tilde K \mathbb{W}_2(\mu,\nu), \quad t \in [0,T],\ x,y \in \mathbb{R}^{m+d},\ \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{m+d}).$$

Moreover, the following Kalman's rank condition holds for some integer $1 \le l \le m$:

$$\mathrm{Rank}[A^i M, 0 \le i \le l-1] = m.$$

By Lemma 1.1, **(C)** implies that (3.1) is well-posed. As in [14], for any $t > 0$, we consider the modified distance

$$\rho_t(x,y) := \sqrt{t^{-2}|x^{(1)} - y^{(1)}|^2 + |x^{(2)} - y^{(2)}|^2}, \quad x = (x^{(1)}, x^{(2)}), y = (y^{(1)}, y^{(2)}) \in \mathbb{R}^m \times \mathbb{R}^d,$$

and define the associated L^2-Wasserstein distance

$$\mathbb{W}_{2,t}(\mu,\nu) := \inf_{\pi \in \mathscr{C}(\mu,\nu)} \left(\int_{\mathbb{R}^{m+d} \times \mathbb{R}^{m+d}} \rho_t(x,y)^2 \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac12}, \quad \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{m+d}).$$

The next theorem characterizes the log-Harnack inequality and the proof is similar to that of [14, Theorem 3.1] since $\int_0^s \tilde\sigma_t(\mathscr{L}_{X_t|\mathscr{F}_t^B})\mathrm{d}B_t$ is deterministic given B. Hence, we will give an outline of the procedure in the following.

Theorem 3.1. *Assume* **(C)**. *Then there exists a constant $c > 0$ such that for any $0 < f \in \mathscr{B}_b(\mathbb{R}^{m+d}), \mu_0, \nu_0 \in \mathscr{P}_2(\mathbb{R}^{m+d}), t \in (0,T]$ and $\xi, \tilde\xi \in L^2(\Omega^1 \to \mathbb{R}^{m+d}, \mathscr{F}_0^1, \mathbb{P}^1)$ with $\mathscr{L}_\xi = \mu_0, \mathscr{L}_{\tilde\xi} = \nu_0$,*

$$\mathbb{E}\{\mathrm{Ent}(\mathscr{L}_{X_t^\xi|\mathscr{F}_t^B} | \mathscr{L}_{X_t^{\tilde\xi}|\mathscr{F}_t^B})\} \le \frac{c}{t^{4l-3}} \mathbb{W}_{2,t}(\mu_0,\nu_0)^2 \le \frac{c(1 \vee T^2)}{t^{4l-1}} \mathbb{W}_2(\mu_0,\nu_0)^2, \tag{3.2}$$

and consequently,

$$P_t \log f(\nu_0) - \log P_t f(\mu_0) \le \frac{c}{t^{4l-3}} \mathbb{W}_{2,t}(\mu_0,\nu_0)^2 \le \frac{c(1 \vee T^2)}{t^{4l-1}} \mathbb{W}_2(\mu_0,\nu_0)^2. \tag{3.3}$$

Proof. For any $t_0\in(0,T]$ and $\mu_0,\nu_0\in\mathscr{P}_2(\mathbb{R}^{m+d})$, let $X_0^{\mu_0},X_0^{\nu_0}$ be $(\Omega^1,\mathscr{F}_0^1)$-measurable such that

$$\mathscr{L}_{X_0^{\mu_0}}=\mu_0,\quad \mathscr{L}_{X_0^{\nu_0}}=\nu_0,\quad \mathbb{E}[\rho_{t_0}(X_0^{\mu_0},X_0^{\nu_0})^2]=\mathbb{W}_{2,t_0}(\mu_0,\nu_0)^2. \tag{3.4}$$

Let $X_t^{\mu_0}$ and $X_t^{\nu_0}$ solve (3.1) with initial value $X_0^{\mu_0}$ and $X_0^{\nu_0}$ respectively and let μ_t,ν_t,η^μ be in (2.3) and (2.15). Then (2.16) and (2.17) still hold. For fixed $t_0\in(0,T]$, let

$$\begin{aligned}
&Q_t:=\int_0^t\frac{s(t-s)}{t^2}\mathrm{e}^{-sA}MM^*\mathrm{e}^{-sA^*}\mathrm{d}s,\quad t\in[0,t_0]\\
&v=X_0^{\nu_0}-X_0^{\mu_0},\\
&V_{t_0}^{\mu,\nu}:=\int_0^{t_0}\mathrm{e}^{-rA}M\Big\{\frac{t_0-r}{t_0}v^{(2)}+\frac{r}{t_0}\big(\eta_{t_0}^\mu-\eta_{t_0}^\nu\big)+\eta_r^\nu-\eta_r^\mu\Big\}\mathrm{d}r,\\
&\alpha_{t_0}(s):=\frac{s}{t_0}\big(\eta_{t_0}^\mu-\eta_{t_0}^\nu-v^{(2)}\big)\\
&\qquad-\frac{s(t_0-s)}{t_0^2}M^*\mathrm{e}^{-sA^*}Q_{t_0}^{-1}\big(v^{(1)}+V_{t_0}^{\mu,\nu}\big),\quad s\in[0,t_0].
\end{aligned}\tag{3.5}$$

Denote $Y_t=(Y_t^{(1)},Y_t^{(2)})$ the solution to the SDE:

$$\begin{cases}\mathrm{d}Y_t^{(1)}=\big\{AY_t^{(1)}+MY_t^{(2)}\big\}\mathrm{d}t,\\ \mathrm{d}Y_t^{(2)}=\big\{b_t(X_t^{\mu_0},\mu_t)+\alpha_{t_0}'(t)\big\}\mathrm{d}t+\sigma_t\mathrm{d}W_t+\tilde\sigma_t(\nu_t)\mathrm{d}B_t,\quad Y_0=X_0^{\nu_0}.\end{cases}\tag{3.6}$$

which combined with (3.5) yields $Y_{t_0}=X_{t_0}^{\mu_0}$. Let

$$\gamma_s:=\sigma_s^*(\sigma_s\sigma_s^*)^{-1}\big\{b_s(Y_s,\nu_s)-b_s(X_s^{\mu_0},\mu_s)-\alpha_{t_0}'(s)\big\},\quad s\in[0,t_0].$$

By **(C)** and [14, (3.17)], there exists a constant $c_1>0$ uniformly in $t_0\in(0,T]$ such that

$$\begin{aligned}|\gamma_s|^2\le& c_1\Big\{\mathbb{W}_2(\mu_s,\nu_s)^2+t_0^{4(1-l)}\rho_{t_0}(X_0^{\mu_0},X_0^{\nu_0})^2+t_0^{4(1-l)}\sup_{t\in[0,t_0]}|\eta_t^\mu-\eta_t^\nu|^2\Big\}\\
&+c_1t_0^{2-4l}\Big(\rho_{t_0}(X_0^{\mu_0},X_0^{\nu_0})^2+\sup_{t\in[0,t_0]}|\eta_t^\nu-\eta_t^\mu|^2\Big),\quad s\in[0,t_0].\end{aligned}\tag{3.7}$$

Recall that $\mathbb{P}^{B,0}$ is defined in (2.39). By Girsanov's theorem,

$$\hat W_t:=W_t-\int_0^t\gamma_s\mathrm{d}s,\quad t\in[0,t_0]$$

is a d_W-dimensional Brownian motion under the weighted conditional probability measure $\mathrm{d}\mathbb{Q}^{B,0}:=R\mathrm{d}\mathbb{P}^{B,0}$, where

$$R:=\mathrm{e}^{\int_0^{t_0}\langle\gamma_s,\mathrm{d}W_s\rangle-\frac12\int_0^{t_0}|\gamma_s|^2\mathrm{d}s}.$$

By (3.6), $\hat Y_t:=Y_t-(0,\eta_t^\nu)$ solves the SDE

$$\begin{cases}\mathrm{d}\hat Y_t^{(1)}=\big\{A\hat Y_t^{(1)}+M\hat Y_t^{(2)}+M\eta_t^\nu\big\}\mathrm{d}t,\\ \mathrm{d}\hat Y_t^{(2)}=b_t(\hat Y_t+(0,\eta_t^\nu),\nu_t)\mathrm{d}t+\sigma_t\mathrm{d}\hat W_t,\quad t\in[0,t_0],\ \hat Y_0=Y_0.\end{cases}$$

Observe that $\hat{X}_t^\nu := X_t^\nu - (0, \eta_t^\nu)$ solves the same equation as $\hat{Y}_t$ for W_t replacing $\hat{W}_t$. By the weak uniqueness and the fact that η_t^ν is $\mathscr{F}_T^B$-measurable, we get

$$\mathscr{L}_{Y_{t_0}}^{\mathbb{Q}^{B,0}} = \mathscr{L}_{\hat{Y}_{t_0}+(0,\eta_{t_0}^\nu)}^{\mathbb{Q}^{B,0}} = \mathscr{L}_{\hat{X}_{t_0}^{\nu_0}+(0,\eta_{t_0}^\nu)}^{\mathbb{P}^{B,0}} = \mathscr{L}_{X_{t_0}^{\nu_0}}^{\mathbb{P}^{B,0}}.$$

This together with $Y_{t_0} = X_{t_0}^{\mu_0}$, Young's inequality and (3.7) yields that we find some constant $c_2 > 0$ such that for any $0 < f \in \mathscr{B}_b(\mathbb{R}^{m+d})$,

$$\begin{aligned}
&\mathbb{E}^{B,0}[\log f(X_{t_0}^{\nu_0})] = \mathbb{E}^{B,0}[R \log f(Y_{t_0})] = \mathbb{E}^{B,0}[R \log f(X_{t_0}^{\mu_0})] \\
&\le \log \mathbb{E}^{B,0}[f(X_{t_0}^{\mu_0})] + \mathbb{E}^{B,0}[R \log R] \\
&= \log \mathbb{E}^{B,0}[f(X_{t_0}^{\mu_0})] + \frac{1}{2}\mathbb{E}_{\mathbb{Q}^{B,0}} \int_0^{t_0} |\gamma_t|^2 \mathrm{d}t \\
&\le \log \mathbb{E}^{B,0}[f(X_{t_0}^{\mu_0})] \\
&\quad + c_2\Big\{t_0 \sup_{s\in[0,t_0]} \mathbb{W}_2(\mu_s, \nu_s)^2 + t_0^{3-4l}\rho_{t_0}(X_0^{\mu_0}, X_0^{\nu_0})^2 + t_0^{3-4l} \sup_{t\in[0,t_0]} |\eta_t^\nu - \eta_t^\mu|^2\Big\}.
\end{aligned}$$

By taking expectation with respect to $\mathbb{P}^B$, using Jensen's inequality, we obtain

$$\begin{aligned}
&\mathbb{E}^B[\log f(X_{t_0}^{\nu_0})] \le \log \mathbb{E}^B[f(X_{t_0}^{\mu_0})] \\
&\quad + c_2\Big\{t_0 \sup_{s\in[0,t_0]} \mathbb{W}_2(\mu_s, \nu_s)^2 + t_0^{3-4l}\mathbb{E}\rho_{t_0}(X_0^{\mu_0}, X_0^{\nu_0})^2 + t_0^{3-4l} \sup_{t\in[0,t_0]} |\eta_t^\nu - \eta_t^\mu|^2\Big\}.
\end{aligned}$$

Then by Lemma 1.1, (2.16), (2.17), [14, (3.5)], [32, Theorem 1.4.2(2)] and (3.4), we prove (3.2), which gives (3.3) by (2.28) and [32, Theorem 1.4.2(2)]. □

3.2 Conditional propagation of chaos

Let N be a positive integer and $(X_0^i, W_t^i)_{1\le i\le N}$, $(X_0^{i,N})_{1\le i\le N}$ and B_t be defined in the same way as in Section 2.2 with $m+d$ replacing d. $b : [0,T] \times \mathbb{R}^{m+d} \times \mathscr{P}(\mathbb{R}^{m+d}) \to \mathbb{R}^d$, $\sigma : [0,T] \to \mathbb{R}^d \otimes \mathbb{R}^{d_W}$, $\tilde{\sigma} : [0,T] \to \mathbb{R}^d \otimes \mathbb{R}^{d_B}$ are measurable and bounded on bounded sets, A is an $m\times m$ matrix and M is an $m \times d$ matrix. Let $X_t^i = (X_t^{i,(1)}, X_t^{i,(2)})$ solve the conditional distribution dependent stochastic Hamiltonian system:

$$\tag{3.8} \begin{cases} \mathrm{d}X_t^{i,(1)} = \{AX_t^{i,(1)} + MX_t^{i,(2)}\}\mathrm{d}t, \\ \mathrm{d}X_t^{i,(2)} = b_t(X_t^i, \mathscr{L}_{X_t^i|\mathscr{F}_t^B})\mathrm{d}t + \sigma_t \mathrm{d}W_t^i + \tilde{\sigma}_t \mathrm{d}B_t, \quad t \in [0,T], \end{cases}$$

and consider the mean field interacting particle system with common noise:

$$\tag{3.9} \begin{cases} \mathrm{d}X_t^{i,N,(1)} = \{AX_t^{i,N,(1)} + MX_t^{i,N,(2)}\}\mathrm{d}t, \\ \mathrm{d}X_t^{i,N,(2)} = b_t(X_t^{i,N}, \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}})\mathrm{d}t + \sigma_t \mathrm{d}W_t^i + \tilde{\sigma}_t \mathrm{d}B_t, \quad t \in [0,T], \end{cases}$$

here $X_t^{i,N} = (X_t^{i,N,(1)}, X_t^{i,N,(2)})$. Recall that $R_{d,q}(N)$ is defined in Theorem 2.3.

Theorem 3.2. *Assume* **(C)**, $\mathbb{E}|X_0^{1,N}|^2<\infty$ *and* $\mathbb{E}|X_0^1|^q<\infty$ *for some* $q>2$. *Then there exists a constant* $C>0$ *depending only on* d,T *and* $\mathbb{E}|X_0^1|^q$ *such that for any* $k\geq 1$ *and* $k\leq N$, *it holds*

$$\begin{aligned}(3.10)\qquad &\mathbb{E}\mathrm{Ent}(\mathscr{L}^{\mathbb{P}}_{(X_t^{1,N},X_t^{2,N},\cdots,X_t^{k,N})|\mathscr{F}_t^B}|\mathscr{L}^{\mathbb{P}}_{(X_t^1,X_t^2,\cdots,X_t^k)|\mathscr{F}_t^B})\\ &\leq CkR_{d,q}(N)+\frac{C}{t^{4l-1}}\frac{k}{N}\mathbb{W}_2(\mathscr{L}_{(X_0^1,X_0^2,\cdots,X_0^N)},\mathscr{L}_{(X_0^{1,N},X_0^{2,N},\cdots,X_0^{N,N})})^2,\quad t\in(0,T].\end{aligned}$$

Proof. Since **(C)** implies (2.31), (2.32) holds for (3.8)-(3.9) in place of (2.29)-(2.30). We first assume that b is bounded. Fix $t_0\in(0,T]$. Let Q_t be defined in (3.5) and for $s\in[0,t_0]$,

$$\begin{aligned}\alpha^i_{t_0}(s):=&-\frac{s}{t_0}(X_0^{i,N,(2)}-X_0^{i,(2)})\\ &-\frac{s(t_0-s)}{t_0^2}M^*\mathrm{e}^{-sA^*}Q_{t_0}^{-1}\bigg((X_0^{i,N,(1)}-X_0^{i,(1)})\\ &\qquad+\int_0^{t_0}\mathrm{e}^{-rA}M\frac{t_0-r}{t_0}(X_0^{i,N,(2)}-X_0^{i,(2)})\mathrm{d}r\bigg),\quad 1\leq i\leq N.\end{aligned}$$

Again set $\mu_t=\mu_t^i=\mathscr{L}_{X_t^i|\mathscr{F}_t^B}$. Construct $Y_t^{i,N}=(Y_t^{i,N,(1)},Y_t^{i,N,(2)})$ as

$$(3.11)\qquad \begin{cases}\mathrm{d}Y_t^{i,N,(1)}=\big\{AY_t^{i,N,(1)}+MY_t^{i,N,(2)}\big\}\mathrm{d}t,\\ \mathrm{d}Y_t^{i,N,(2)}=[b_t(X_t^i,\mu_t)+(\alpha^i_{t_0})'(t)]\mathrm{d}t+\sigma_t\mathrm{d}W_t^i+\tilde{\sigma}_t\mathrm{d}B_t,\quad Y_0^{i,N}=X_0^{i,N}.\end{cases}$$

Then we have

$$\begin{aligned}Y_t^{i,N,(2)}&=X_t^{i,(2)}+(X_0^{i,N,(2)}-X_0^{i,(2)})+\alpha_{t_0}(t),\\ Y_t^{i,N,(1)}&=X_t^{i,(1)}+\mathrm{e}^{At}(X_0^{i,N,(1)}-X_0^{i,(1)})\\ &\quad+\int_0^t\mathrm{e}^{A(t-s)}M[(X_0^{i,N,(2)}-X_0^{i,(2)})+\alpha_{t_0}(s)]\mathrm{d}s,\quad 1\leq i\leq N.\end{aligned}$$

In particular, it holds

$$(3.12)\qquad Y_{t_0}^{i,N}=X_{t_0}^i,\quad 1\leq i\leq N.$$

Let

$$\begin{aligned}&\mathrm{d}\hat{W}_t^i=\mathrm{d}W_t^i-\gamma_t^{i,N}\mathrm{d}t,\\ &\gamma_t^{i,N}=\sigma_t^*(\sigma_t\sigma_t^*)^{-1}\left[b_t(Y_t^{i,N},\frac{1}{N}\sum_{i=1}^N\delta_{Y_t^{i,N}})-b_t(X_t^i,\mu_t)-\alpha'_{t_0}(t)\right],\quad 1\leq i\leq N,\end{aligned}$$

and

$$R_t=\exp\left\{\int_0^t\sum_{i-1}^N\langle\gamma_s^{i,N},\mathrm{d}W_s^i\rangle-\frac{1}{2}\int_0^t\sum_{i=1}^N|\gamma_s^{i,N}|^2\mathrm{d}s\right\}.$$

Similarly to (3.7) and using the boundedness of b, there exists a constant $c_0 > 0$ such that

$$\int_0^{t_0} |\gamma_s^{i,N}|^2 \mathrm{d}s \le c_0 \int_0^{t_0} \left(1 \wedge \mathbb{W}_2\Big(\frac{1}{N}\sum_{i=1}^N \delta_{Y_t^{i,N}}, \mu_t\Big)^2\right) \mathrm{d}t + c_0 \frac{|X_0^i - X_0^{i,N}|^2}{t_0^{4l-1}}. \tag{3.13}$$

By Girsanov's theorem, $((\hat{W}_t^i)_{t\in[0,t_0]})_{1\le i\le N}$ is an $(N \times d_W)$-dimensional Brownian motion under the conditional probability measure $\mathbb{Q}^{B,0} = R_{t_0}\mathbb{P}^{B,0}$. Moreover, (3.11) can be rewritten as

$$\begin{cases} \mathrm{d}Y_t^{i,N,(1)} = \big\{AY_t^{i,N,(1)} + MY_t^{i,N,(2)}\big\}\mathrm{d}t, \\ \mathrm{d}Y_t^{i,N,(2)} = b_t(Y_t^{i,N}, \frac{1}{N}\sum_{i=1}^N \delta_{Y_t^{i,N}})\mathrm{d}t + \sigma_t \mathrm{d}\hat{W}_t^i + \tilde{\sigma}_t \mathrm{d}B_t, \quad Y_0^{i,N} = X_0^{i,N}, t \in [0,t_0]. \end{cases}$$

By the weak uniqueness, it holds

$$\mathscr{L}^{\mathbb{Q}^{B,0}}_{\{Y_t^{i,N}\}_{1\le i\le N}} = \mathscr{L}^{\mathbb{P}^{B,0}}_{\{X_t^{i,N}\}_{1\le i\le N}}, \quad t \in [0,t_0].$$

This together with (3.12) implies

$$\begin{aligned} \mathbb{E}^{B,0} f(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{N,N}) &= \mathbb{E}^{B,0}[R_{t_0} f(Y_{t_0}^{1,N}, Y_{t_0}^{2,N}, \cdots, Y_{t_0}^{N,N})] \\ &= \mathbb{E}^{B,0}[R_{t_0}(X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^N)], \quad f \in \mathscr{B}_b((\mathbb{R}^{m+d})^N). \end{aligned}$$

Note that

$$\mathbb{E}_{\mathbb{Q}^{B,0}} \mathbb{W}_2\Big(\frac{1}{N}\sum_{i=1}^N \delta_{Y_t^{i,N}}, \mu_t\Big)^2 = \mathbb{E}^{B,0}\mathbb{W}_2\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}, \mu_t\Big)^2.$$

Combining this with Young's inequality and (3.13), we can find a constant $c > 0$ such that

$$\begin{aligned} &\mathbb{E}^{B,0} \log f(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{N,N}) \\ &\le \log \mathbb{E}^{B,0}[f(X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^N)] + \frac{1}{2}\mathbb{E}_{\mathbb{Q}^{B,0}} \int_0^{t_0} \sum_{i=1}^N |\gamma_s^{i,N}|^2 \mathrm{d}s \\ &\le \log \mathbb{E}^{B,0}[f(X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^N)] \\ &+ c\sum_{i=1}^N \mathbb{E}^{B,0}\int_0^{t_0} \mathbb{W}_2\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}, \mu_t\Big)^2 \mathrm{d}t + c\sum_{i=1}^N \frac{|X_0^i - X_0^{i,N}|^2}{t_0^{4l-1}}, \quad 0 < f \in \mathscr{B}_b((\mathbb{R}^{m+d})^N). \end{aligned}$$

Repeating the same argument to derive (2.44) from (2.42), we obtain

$$\begin{aligned} &\mathbb{E}^B \log f(X_{t_0}^{1,N}, X_{t_0}^{2,N}, \cdots, X_{t_0}^{k,N}) \\ &\le \log \mathbb{E}^B[f(X_{t_0}^1, X_{t_0}^2, \cdots, X_{t_0}^k)] \\ &+ 2ck\left\{\mathbb{E}^B \int_0^{t_0} \mathbb{W}_2\Big(\frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}}, \mu_t\Big)^2 \mathrm{d}t + \frac{\sum_{i=1}^N \mathbb{E}|X_0^i - X_0^{i,N}|^2}{N t_0^{4l-1}}\right\}, \quad 0 < f \in \mathscr{B}_b((\mathbb{R}^{m+d})^k). \end{aligned}$$

Again using [32, Theorem 1.4.2(2)] and (2.32), we derive (3.10) when b is bounded. Finally, by the same approximation technique in **(Step (ii))** in the proof of Theorem 2.3(2), we obtain the desired result for general b. □

Data Availability Statement Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

[1] J. Bao, J. Shao, D. Wei, Wellposedness of conditional McKean-Vlasov equations with singular drifts and regime-switching, *Discrete Contin. Dyn. Syst. Ser. B* 28(2023), 2911-2926.

[2] D. Bresch, P.-E. Jabin, Z. Wang, Mean-field limit and quantitative estimates with singular attractive kernels, *Duke Math. J.* 172(13), 2591-2641.

[3] P. Briand, P. Cardaliaguet, P.-E. Chaudru de Raynal, Y. Hu, Forward and Backward Stochastic Differential Equations with normal constraint in law, *Stochastic Process. Appl.* 130(2020),7021-7097.

[4] L. Bo, T. Li, X. Yu, Centralized systemic risk control in the interbank system: weak formulation and Gamma-convergence, *Stochastic Process. Appl.* 150(2022), 622-654.

[5] R. Carmona, F. Delarue, Probabilistic Theory of Mean Field Games with Applications II, Mean-field Games with Common Noise and Master Equations, *Springer*, 2018.

[6] M. Coghi, F. Flandoli, propagation of chaos for interacting particles subject to environmental noise, *Ann. Appl. Probab.* 26(2016), 1407-1442.

[7] M. Coghi, B. Gess, Stochastic nonlinear Fokker-Planck equations, *Nonlinear Anal.* 187(2019), 259-278.

[8] D. A. Dawson, Stochastic evolution equations and related measure processes, *J. Multivariate Anal.* 5 (1975) 1-52.

[9] D. A. Dawson, J. Vaillancourt, Stochastic McKean-Vlasov equations, *Nonlinear Differential Equations. Appl.* 2(1995),199-229.

[10] X. Erny, E. Löcherbach, D. Loukianova, Conditional propagation of chaos for mean field systems of interacting neurons, *Electron. J. Probab.* 26(2021), 1-25.

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

[12] S. Guo, D. Luo, Scaling limit of moderately interacting particle systems with singular interaction and environmental noise, *Ann. Appl. Probab.* 33(2023), 2066-2102.

[13] W. R. P. Hammersley, D. Šiška, L. Szpruch, Weak existence and uniqueness for McKean-Vlasov SDEs with common noise, *Ann. Probab.* 49(2021), 527-555.

[14] X. Huang, F.-Y. Wang, Regularities and Exponential Ergodicity in Entropy for SDEs Driven by Distribution Dependent Noise, *arXiv:2209.14619, to appear in Bernoulli.*

[15] P.-E. Jabin, Z. Wang, Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels, *Invent. Math.* 214 (2018), 523-591.

[16] P.-E. Jabin, Z. Wang, Mean field limit and propagation of chaos for Vlasov systems with bounded forces, *J. Funct. Anal.* 271(2016), 3588-3627.

[17] V. N. Kolokoltsov, M. Troeva, On the mean field games with common noise and the McKean-Vlasov SPDEs, *Stoch. Anal. Appl.* 37(2019), 522-549.

[18] C. Kumar, Neelima, C. Reisinger, W. Stockinger, Well-posedness and tamed schemes for McKean-Vlasov equations with common noise, *Ann. Appl. Probab.* 32(2022), 3283-3330.

[19] T. G. Kurtz, J. Xiong, Particle representations for a class of nonlinear SPDEs, *Stochastic Process. Appl.* 83(1999), 103-126.

[20] D. Lacker, Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions, *Probab. Math. Phys.* 4(2023), 377-432.

[21] D. Lacker, M. Shkolnikov, J. Zhang, Superposition and mimicking theorems for conditional McKean-Vlasov equations, *J. Eur. Math. Soc.* 25(2023), 3229-3288.

[22] J.-M. Lasry, P.-L. Lions, Mean field games, *Japan. J. Math.* 2(2007), 229-260.

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

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

[25] P. Ren, F.-Y. Wang, Entropy estimate between diffusion processes and application to McKean-Vlasov SDEs, *arXiv:2302.13500.*

[26] J. Shao, D. Wei, Propagation of chaos and conditional McKean-Vlasov SDEs with regime-switching. *Front. Math. China* 17(2022), 731-746.

[27] J. Shao, T. Tian, S. Wang, Conditional McKean-Vlasov SDEs with jumps and Markovian regime-switching: wellposedness, propagation of chaos, averaging principle, *J. Math. Anal. Appl.* 534(2024), Paper No. 128080.

[28] Y. Shao, X. Zhao, Quantitative particle approximations of stochastic 2D Navier-Stokes equation, *arXiv:2402.02336.*

[29] 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.

[30] Y. V. Vuong, M. Hauray, E. Pardoux, Conditional propagation of chaos in a spatial stochastic epidemic model with common noise, *Stoch. Partial Differ. Equ. Anal. Comput.* 10(2022), 1180-1210.

[31] F.-Y. Wang, Harnack inequality for SDE with multiplicative noise and extension to Neumann semigroup on nonconvex manifolds, *Ann. Probab.* 39(2011), 1449-1467.

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

[33] F.-Y. Wang, Image-dependent conditional McKean-Vlasov SDEs for measure-valued diffusion processes, *J. Evol. Equ.* 21(2021), 2009-2045.