Exponential Convergence in Entropy and Wasserstein Distance for Distribution Dependent SDEs *

Panpan Ren[c], **Feng-Yu Wang**[a,b]

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

[b] Department of Mathematics, Swansea University, Bay Campus, SA1 8EN, United Kingdom

[c] Department of Mathematics, City University of HongKong, HongKong, China

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

October 24, 2020

Abstract

The following type exponential convergence is proved for (non-degenerate or degenerate) distribution dependent SDEs:

$$\mathbb{W}_2(\mu_t,\mu_\infty)^2+\mathrm{Ent}(\mu_t|\mu_\infty)\le c\mathrm{e}^{-\lambda t}\min\big\{\mathbb{W}_2(\mu_0,\mu_\infty)^2,\mathrm{Ent}(\mu_0|\mu_\infty)\big\},\quad t\ge 1,$$

where $c,\lambda>0$ are constants, μ_t is the distribution of the solution at time t, μ_∞ is the unique invariant probability measure, Ent is the relative entropy and $\mathbb{W}_2$ is the L^2-Wasserstein distance. As applications, this type exponential convergence is confirmed for non-degenerate/degenerate granular media type equations generalizing those studied in [7, 11] on the exponential convergence in a mean field entropy.

AMS subject Classification: 60B05, 60B10.
Keywords: Distribution dependent SDE, exponential convergence in entropy, stochastic Hamiltonian system, granular media equation.

1 Introduction

The convergence in entropy for stochastic systems is an important topic in both probability theory and mathematical physics, and has been well studied for Markov processes by using the

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

log-Sobolev inequality, see for instance [5] and references therein. However, the existing results derived in the literature do not apply to distribution dependent SDEs due to the nonlinearity of the associated Fokker-Planck equation. In 2003, Carrillo, McCann and Villani [7] proved the exponential convergence in a mean field entropy of the following granular media equation for probability density functions $(\rho_t)_{t\geq 0}$ on $\mathbb{R}^d$:

E0 (1.1) $$\partial_t \rho_t = \Delta \rho_t + \mathrm{div}\big\{\rho_t \nabla(V + W * \rho_t)\big\},$$

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

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

for $\mu, \nu \in \mathscr{P}$, the space of all probability measures on $\mathbb{R}^d$, and consider the free energy functional

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

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

ETP (1.2) $$\mathrm{Ent}^{V,W}(\mu) := E^{V,W}(\mu) - \inf_{\nu \in \mathscr{P}} E^{V,W}(\nu), \quad \mu \in \mathscr{P}.$$

According to [7], $E^{V,W}$ has a unique minimizer μ_∞, and $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ converges to μ_∞ exponentially in the mean field entropy:

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

Recently, this result was generalized in [11] by establishing the uniform log-Sobolev inequality for the associated mean field particle systems, such that $\mathrm{Ent}^{V,W}(\mu_t)$ decays exponentially for a class of non-convex $V \in C^2(\mathbb{R}^d)$ and $W \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$, where $W(x,y) = W(y,x)$ and $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ for ρ_t solving the nonlinear PDE

E00 (1.3) $$\partial_t \rho_t = \Delta \rho_t + \mathrm{div}\big\{\rho_t \nabla(V + W \circledast \rho_t)\big\},$$

where

A00 (1.4) $$W \circledast \rho_t := \int_{\mathbb{R}^d} W(\cdot, y)\rho_t(y)\mathrm{d}y.$$

In this case, $\mathrm{Ent}^{V,W}$ is defined in (1.2) for the free energy functional

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

To study (1.3) using probability methods, we consider the following McKean-Vlasov SDE with initial distribution μ_0:

E0' (1.5) $$\mathrm{d}X_t = \sqrt{2}\mathrm{d}B_t - \nabla\big\{V + W \circledast \mathscr{L}_{X_t}\big\}(X_t)\mathrm{d}t,$$

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

A01 (1.6) $$(W \circledast \mu)(x) := \int_{\mathbb{R}^d} W(x,y)\mu(\mathrm{d}y), \quad x \in \mathbb{R}^d, \mu \in \mathscr{P}$$

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

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

Therefore, ρ_t solves (1.3). On the other hand, by this fact and the uniqueness of (1.1) and (1.5), if ρ_t solves (1.1) with $\mu_0(\mathrm{d}x) := \rho_0(x)\mathrm{d}x$, then $\rho_t(x)\mathrm{d}x = \mathscr{L}_{X_t}(\mathrm{d}x)$ for X_t solving (1.5) with $\mathscr{L}_{X_0} = \mu_0$.

To extend the study of [7, 11], in this paper we investigate the exponential convergence in entropy for the following distribution dependent SDE on $\mathbb{R}^d$:

E1 (1.7) $$\mathrm{d}X_t = \sigma(X_t)\mathrm{d}W_t + b(X_t, \mathscr{L}_{X_t})\mathrm{d}t,$$

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

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

are measurable, and $\mathscr{P}_2$ is the class of probability measures on $\mathbb{R}^d$ with $\mu(|\cdot|^2) < \infty$. Since in this general setting the associated mean field entropy is not available, and it is less explicit as defined in (1.2) for the special model (1.5), we intend to study the exponential convergence of $\mathscr{L}_{X_t}$ in the classical relative entropy Ent and the Wasserstein distance. Recall that for any $p \geq 1$, the L^p-Wasserstein distance is defined by

$$\mathbb{W}_p(\mu_1,\mu_2) := \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)}\bigg(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\pi(\mathrm{d}x,\mathrm{d}y)\bigg)^{\frac{1}{p}}, \quad \mu_1,\mu_2 \in \mathscr{P},$$

where $\mathscr{C}(\mu_1,\mu_2)$ is the set of all couplings of μ_1 and μ_2. To this end, we will apply the log-Harnack and Talagrand inequalities, see Theorem 2.1 below. Since the log-Harnack inequality is not yet available when σ also depends on the distribution, see [25], in (1.7) we only consider distribution-free σ. In particular, for a class of granular media type equations generalizing the framework of [7, 11], we prove

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

for $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ and some constants $c, \lambda > 0$, see Theorem 2.2 below for details.

The remainder of the paper is organized as follows. In Section 2, we state our main results for non-degenerate and degenerate models respectively, where the first case includes the granular media type equations (1.3) or the corresponding Mckean-Vlasov SDE (1.5) as a special example, and the second case deals with the distribution dependent stochastic Hamiltonian system referring to the degenerate granular media equation. The main results are proved in Sections 3-5 respectively, where Section 4 establishes the log-Harnack inequality for distribution dependent stochastic Hamiltonian systems.

2 Main results and examples

We first present a criterion on the exponential convergence for distribution dependent SDEs by using the log-Harnack and Talagrand inequalities, and prove (2.15) for the granular media type equations (2.10) below which generalizes the framework of [11]. Then we state our results for the general distribution dependent SDE (1.7) with non-degenerate and degenerate noises respectively.

2.1 A criterion with application to Granular media type equations

In general, we consider the following distribution dependent SDE:

EP (2.1) $$\mathrm{d}X_t = \sigma(X_t, \mathscr{L}_{X_t})\mathrm{d}W_t + b(X_t, \mathscr{L}_{X_t})\mathrm{d}t,$$

where W_t is the m-dimensional Brownian motion and

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

are measurable. We assume that this SDE is strongly and weakly well-posed for square integrable initial values. It is in particular the case if there exists a constant $K > 0$ such that

KK (2.2) $$\langle b(x,\mu)-b(y,\nu), x-y\rangle^+ + \|\sigma(x,\mu)-\sigma(y,\nu)\|^2 \le K\big\{|x-y|^2 + \mathbb{W}_2(\mu,\nu)^2\big\}, \quad x,y\in\mathbb{R}^d, \mu,\nu\in\mathscr{P}_2,$$

see for instance [25]. See also [13, 27] and references therein for the well-posedness of distribution dependent SDEs with singular coefficients. For any $\mu \in \mathscr{P}_2$, let $P_t^*\mu = \mathscr{L}_{X_t}$ for the solution X_t with initial distribution $\mathscr{L}_{X_0} = \mu$. Let

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

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

T0 **Theorem 2.1.** *Assume that P_t^* has a unique invariant probability measure $\mu_\infty \in \mathscr{P}_2$ such that for some constants $t_0, c_0, C > 0$ we have the log-Harnack inequality*

LHI (2.3) $$P_{t_0}(\log f)(\nu) \le \log P_{t_0} f(\mu) + c_0 \mathbb{W}_2(\mu,\nu)^2, \quad \mu,\nu\in\mathscr{P}_2,$$

and the Talagrand inequality

TLI (2.4) $$\mathbb{W}_2(\mu,\mu_\infty)^2 \le C\mathrm{Ent}(\mu|\mu_\infty), \quad \mu\in\mathscr{P}_2.$$

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

EW (2.5) $$\mathbb{W}_2(P_t^*\mu, \mu_\infty)^2 \leq c_1 \mathrm{e}^{-\lambda t}\mathbb{W}_2(\mu, \mu_\infty)^2, \quad t \geq t_1, \mu \in \mathscr{P}_2,$$

then

ET (2.6) $$\begin{aligned}&\max\big\{c_0^{-1}\mathrm{Ent}(P_t^*\mu|\mu_\infty), \mathbb{W}_2(P_t^*\mu, \mu_\infty)^2\big\}\\ &\leq c_1\mathrm{e}^{-\lambda(t-t_0)}\min\big\{\mathbb{W}_2(\mu, \mu_\infty)^2, C\mathrm{Ent}(\mu|\mu_\infty)\big\}, \quad t \geq t_0 + t_1, \mu \in \mathscr{P}_2.\end{aligned}$$

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

EW' (2.7) $$\mathrm{Ent}(P_t^*\mu|\mu_\infty) \leq c_2\mathrm{e}^{-\lambda t}\mathrm{Ent}(\mu|\mu_\infty), \quad t \geq t_2, \nu \in \mathscr{P}_2,$$

then

ET' (2.8) $$\begin{aligned}&\max\big\{\mathrm{Ent}(P_t^*\mu, \mu_\infty), C^{-1}\mathbb{W}_2(P_t^*\mu, \mu_\infty)^2\big\}\\ &\leq c_2\mathrm{e}^{-\lambda(t-t_0)}\min\big\{c_0\mathbb{W}_2(\mu, \mu_\infty)^2, \mathrm{Ent}(\mu|\mu_\infty)\big\}, \quad t \geq t_0 + t_2, \mu \in \mathscr{P}_2.\end{aligned}$$

When σ does not depend on the distribution, the log-Harnack inequality (2.3) has been established in [25]. The Talagrand inequality was first found in [18] for μ_∞ being the Gaussian measure, and extended in [5] to μ_∞ satisfying the log-Sobolev inequality

LS0 (2.9) $$\mu_\infty(f^2\log f^2) \leq C\mu_\infty(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^d), \mu_\infty(f^2) = 1,$$

see [15] for an earlier result under a curvature condition, and see [19] for further extensions.

To illustrate this result, we consider the granular media type equation (2.10) for probability density functions $(\rho_t)_{t\geq 0}$ on $\mathbb{R}^d$:

E01 (2.10) $$\partial_t\rho_t = \mathrm{div}\big\{a\nabla\rho_t + \rho_t a\nabla(V + W \circledast \rho_t)\big\},$$

where $W \circledast \rho_t$ is in (1.4), and the functions

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

satisfy the following assumptions.

(H_1) $a := (a_{ij})_{1\leq i,j\leq d} \in C_b^2(\mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d)$, and $a \geq \lambda_a I_d$ for some constant $\lambda_a > 0$.

(H_2) $V \in C^2(\mathbb{R}^d), W \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$ with $W(x, y) = W(y, x)$, and there exist constants $\kappa_0 \in \mathbb{R}$ and $\kappa_1, \kappa_2, \kappa_0' > 0$ such that

ICC1 (2.11) $$\mathrm{Hess}_V \geq \kappa_0 I_d, \quad \kappa_0' I_{2d} \geq \mathrm{Hess}_W \geq \kappa_0 I_{2d},$$

ICC2 (2.12) $$\langle x, \nabla V(x)\rangle \geq \kappa_1|x|^2 - \kappa_2, \quad x \in \mathbb{R}^d$$

Moreover, for any $\lambda > 0$,

ICC3 (2.13) $$\int_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{e}^{-V(x)-V(y)-\lambda W(x,y)}\mathrm{d}x\mathrm{d}y < \infty.$$

(H_3) There exists a function $b_0 \in L^1_{loc}([0,\infty))$ with

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

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

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

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

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

and the corresponding finite-dimensional Gibbs measure

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

where $Z_N := \int_{\mathbb{R}^{dN}} \mathrm{e}^{-H_N(x)}\mathrm{d}x < \infty$ due to (2.13) in (H_2). For any $1 \le i \le N$, the conditional marginal of $\mu^{(N)}$ given $z \in \mathbb{R}^{d(N-1)}$ is given by

$$\mu_z^{(N)}(\mathrm{d}x) := \frac{1}{Z_N(z)}\mathrm{e}^{-H_N(x|z)}\mathrm{d}x, \quad Z_N(z) := \int_{\mathbb{R}^d} \mathrm{e}^{-H_N(x|z)}\mathrm{d}x,$$

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

We have the following result.

C2.0 **Theorem 2.2.** *Assume (H_1)-(H_3). If there is a constant $a > 0$ such that the uniform log-Sobolev inequality*

LSS (2.14) $$\mu_z^{(N)}(f^2 \log f^2) \le \frac{1}{\beta}\mu_z^{(N)}(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^d), \mu_z^{(N)}(f^2) = 1, N \ge 2, z \in \mathbb{R}^{d(N-1)}$$

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

WU (2.15) $$\mathbb{W}_2(\mu_t, \mu_\infty)^2 + \mathrm{Ent}(\mu_t|\mu_\infty) \le c\mathrm{e}^{-\lambda_a \beta(1-r_0)^2 t} \min\big\{\mathbb{W}_2(\mu_0, \mu_\infty)^2 + \mathrm{Ent}(\mu_0|\mu_\infty)\big\}, \quad t \ge 1$$

holds for any probability density functions $(\rho_t)_{t\ge 0}$ solving (2.10), where $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x, t \ge 0$.

This result allows V and W to be non-convex. For instance, let $V = V_1 + V_2 \in C^2(\mathbb{R}^d)$ such that $\|V_1\|_\infty \wedge \|\nabla V_1\|_\infty < \infty$, $\mathrm{Hess}_{V_2} \ge \lambda I_d$ for some $\lambda > 0$, and $W \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$ with $\|W\|_\infty \wedge \|\nabla W\|_\infty < \infty$. Then the uniform log-Sobolev inequality (2.14) holds for some constant $\beta > 0$. Indeed, by the Bakry-Emery criterion, $\mu_2(\mathrm{d}x) := \frac{1}{\int_{\mathbb{R}^d}\mathrm{e}^{-V(x)}\mathrm{d}x}\mathrm{e}^{-V_2(x)}\mathrm{d}x$ satisfies the log-Sobolev inequality

$$\mu_2(f^2 \log f^2) \le \frac{2}{\lambda}\mu_2(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^d), \mu_2(f^2) = 1.$$

then (2.14) with some constant $\beta > 0$ follows by the stability of the log-Sobolev inequality under bounded perturbations (see [9, 8]) as well as Lipschitz perturbations (see [1]) for the potential V_2. Moreover, assumption (H_1)-(H_3) hold provided $\|\mathrm{Hess}_W\|_\infty$ is small enough such that $r_0 < 1$. So, Theorem 2.2 applies. See [11] for more concrete examples satisfying (H_1)-(H_3) and (2.14).

2.2 The non-degenerate case

In this part, we make the following assumption:

(A_1) There exists a constant $K>0$ such that for any $x,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_2$,

$$\|\sigma(x)-\sigma(y)\|^2+\langle b(x,\mu)-b(y,\nu),x-y\rangle^+\le K\big\{|x-y|^2+\mathbb{W}_2(\mu,\nu)^2\big\},$$
$$|b(0,\mu)|\le K(1+\|\mu\|_2).$$

(A_2) $\sigma\sigma^*$ is invertible with $\lambda:=\|(\sigma\sigma^*)^{-1}\|_\infty<\infty$, and there exist constants $K_2>K_1\ge 0$ such that for any $x,y\in\mathbb{R}^d$ and $\mu,\nu\in\mathscr{P}_2$,

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

According to [25, Theorem 2.1], if (A_1) holds and $b(x,\mu)$ is continuous on $\mathbb{R}^d\times\mathscr{P}_2$, then for any initial value $X_0\in L^2(\Omega\to\mathbb{R}^d,\mathscr{F}_0,\mathbb{P})$, (1.7) has a unique solution which satisfies

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

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

T1 **Theorem 2.3.** *Assume (A_1) and (A_2). Then P_t^* has a unique invariant probability measure μ_∞ such that*

EX1 (2.16) $$\max\big\{\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2,\mathrm{Ent}(P_t^*\mu|\mu_\infty)\big\}\le\frac{c_1}{t\wedge 1}\mathrm{e}^{-(K_2-K_1)t}\mathbb{W}_2(\mu,\mu_\infty)^2,\quad t>0,\mu\in\mathscr{P}_2$$

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

EX2 (2.17) $$\max\big\{\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2,\mathrm{Ent}(P_t^*\mu|\mu_\infty)\big\}\le c_2\mathrm{e}^{-(K_2-K_1)t}\min\big\{\mathbb{W}_2(\mu,\mu_\infty)^2,\mathrm{Ent}(\mu|\mu_\infty)\big\}.$$

To illustrate this result, we consider the granular media equation (1.3), for which we take

SBB1 (2.18) $$\sigma=\sqrt{2}I_d,\qquad b(x,\mu)=-\nabla\big\{V+W\circledast\mu\big\}(x),\quad (x,\mu)\in\mathbb{R}^d\times\mathscr{P}_2.$$

The following example is not included by Theorem 2.2 since the function W may be non-symmetric.

Example 2.1 (Granular media equation). Consider (1.1) with $V\in C^2(\mathbb{R}^d)$ and $W\in C^2(\mathbb{R}^d\times\mathbb{R}^d)$ satisfying

CC1 (2.19) $$\mathrm{Hess}_V\ge\lambda I_d,\quad \mathrm{Hess}_W\ge\delta_1 I_d,\quad \|\mathrm{Hess}_W\|\le\delta_2$$

for some constants $\lambda_1,\delta_2>0$ and $\delta_1\in\mathbb{R}$. If $\lambda+\delta_1-\delta_2>0$, then there exists a unique $\mu_\infty\in\mathscr{P}_2$ and a constant $c>0$ such that for any probability density functions $(\rho_t)_{t\ge0}$ solving (1.3), $\mu_t(\mathrm{d}x):=\rho_t(x)\mathrm{d}x$ satisfies

ERR0 (2.20) $$\max\big\{\mathbb{W}_2(\mu_t,\mu_\infty),\mathrm{Ent}(\mu_t|\mu_\infty)\big\}\le c\mathrm{e}^{-(\lambda+\delta_1-\delta_2)t}\min\big\{\mathbb{W}_2(\mu_0,\mu_\infty),\mathrm{Ent}(\mu_0|\mu_\infty)\big\},\quad t\ge1.$$

Proof. Let σ and b be in (2.18). Then (2.19) implies (A_1) and

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

where we have used the formula

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

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

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

Therefore, if (2.19) holds for $\lambda+\delta_1-\delta_2>0$, Theorem 2.3 implies that P_t^* has a unique invariant probability measure $\mu_\infty\in\mathscr{P}_2$, such that (2.20) holds for $\mu_0\in\mathscr{P}_2$. When $\mu_0\notin\mathscr{P}_2$, we have $\mathbb{W}_2(\mu_0,\mu_\infty)^2=\infty$ since $\mu_\infty\in\mathscr{P}_2$. Combining this with the Talagrand inequality

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

for some constant $C>0$, see the proof of Theorem 2.3, we have $\mathrm{Ent}(\mu_0|\mu_\infty)=\infty$ for $\mu_0\notin\mathscr{P}_2$, so that (2.20) holds for all $\mu_0\in\mathscr{P}$. □

2.3 The degenerate case

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

E21 (2.21)
$$\begin{cases}\mathrm{d}X_t=BY_t\mathrm{d}t,\\ \mathrm{d}Y_t=\sqrt{2}\mathrm{d}W_t-\Big\{B^*\nabla V(\cdot,\mathscr{L}_{(X_t,Y_t)})(X_t)+\beta B^*(BB^*)^{-1}X_t+Y_t\Big\}\mathrm{d}t,\end{cases}$$

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

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

is measurable. Let

$$\begin{aligned}&\psi_B((x,y),(\bar x,\bar y)):=\sqrt{|x-\bar x|^2+|B(y-\bar y)|^2},\quad (x,y),(\bar x,\bar y)\in\mathbb{R}^{d_1+d_2},\\ &\mathbb{W}_2^{\psi_B}(\mu,\nu):=\inf_{\pi\in\mathscr{C}(\mu,\nu)}\bigg\{\int_{\mathbb{R}^{d_1+d_2}\times\mathbb{R}^{d_1+d_2}}{\psi_B}^2\mathrm{d}\pi\bigg\}^{\frac12},\quad \mu,\nu\in\mathscr{P}_2(\mathbb{R}^{d_1+d_2}).\end{aligned}$$

We assume

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

AH1 (2.22) $$\theta_1+\theta_2<\beta,$$

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

AH2 (2.23) $$\begin{aligned}&\big\langle BB^*\{\nabla V(\cdot,\mu)(x)-\nabla V(\cdot,\mu')(x')\},x-x'+(1+\beta)B(y-y')\big\rangle\\&\geq -\theta_1\psi_B((x,y),(x',y'))^2-\theta_2\mathbb{W}_2^{\psi_B}(\mu,\mu')^2.\end{aligned}$$

Obviously, **(C)** implies (A_1) for $d=m=d_1+d_2$, $\sigma=\mathrm{diag}\{0,\sqrt{2}I_{d_2}\}$, and

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

So, according to [25], (2.21) is well-posed for any initial value in $L^2(\Omega\to\mathbb{R}^{d_1+d_2},\mathscr{F}_0,\mathbb{P})$. Let $P_t^*\mu=\mathscr{L}_{(X_t,Y_t)}$ for the solution with initial distribution $\mu\in\mathscr{P}_2(\mathbb{R}^{d_1+d_2})$.

T2 **Theorem 2.4.** *Assume* **(C)**. *Then P_t^* has a unique invariant probability measure μ_∞ such that for any $t>0$ and $\mu\in\mathscr{P}_2(\mathbb{R}^{d_1+d_2})$,*

KK0 (2.24) $$\max\big\{\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2,\mathrm{Ent}(P_t^*\mu|\mu_\infty)\big\}\leq\frac{c\mathrm{e}^{-2\kappa t}}{(1\wedge t)^3}\min\big\{\mathrm{Ent}(\mu|\mu_\infty),\mathbb{W}_2(\mu,\mu_\infty)^2\big\}.$$

holds for some constant $c>0$ and

KK (2.25) $$\kappa:=\frac{2(\beta-\theta_1-\theta_2)}{2+2\beta+\beta^2+\sqrt{\beta^4+4}}.$$

Example 2.2 (Degenerate granular media equation). Let $m\in\mathbb{N}$ and $W\in C^2(\mathbb{R}^m\times\mathbb{R}^{2m})$. Consider the following PDE for probability density functions $(\rho_t)_{t\geq0}$ on $\mathbb{R}^{2m}$:

*EN2 (2.26) $$\partial_t\rho_t(x,y)=\Delta_y\rho_t(x,y)-\langle\nabla_x\rho_t(x,y),y\rangle+\langle\nabla_y\rho_t(x,y),\nabla_x(W\circledast\rho_t)(x)+\beta x+y\rangle,$$

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

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

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

*EN3 (2.27) $$|\nabla W(\cdot,z)(x)-\nabla W(\cdot,\bar z)(\bar x)|\leq\theta\big(|x-\bar x|+|z-\bar z|\big),\quad x,\bar x\in\mathbb{R}^m,z,\bar z\in\mathbb{R}^{2m},$$

then there exists a unique probability measure $\mu_\infty\in\mathscr{P}_2(\mathbb{R}^{2m})$ and a constant $c>0$ such that for any probability density functions $(\rho_t)_{t\geq0}$ solving (2.26), $\mu_t(\mathrm{d}x):=\rho_t(x)\mathrm{d}x$ satisfies

ERR (2.28) $$\max\big\{\mathbb{W}_2(\mu_t,\mu_\infty)^2,\mathrm{Ent}(\mu_t|\mu_\infty)\big\}\leq c\mathrm{e}^{-\kappa t}\min\big\{\mathbb{W}_2(\mu_0,\mu_\infty)^2,\mathrm{Ent}(\mu_0|\mu_\infty)\big\},\quad t\geq1$$

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

Proof. Let $d_1 = d_2 = m$ and (X_t, Y_t) solve (2.21) for

GPP (2.29) $$B := I_m, \quad V(x,\mu) := \int_{\mathbb{R}^{2m}} W(x,z)\mu(\mathrm{d}z).$$

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

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

Then ρ_t solves (2.26). On the other hand, by the uniqueness of of (2.21) and (2.26), for any solution ρ_t to (2.26) with $\mu_0(\mathrm{d}z) := \rho_0(z)\mathrm{d}z \in \mathscr{P}_2(\mathbb{R}^{2m})$ for $d = 2m$, $\rho_t(z)\mathrm{d}z = \mathscr{L}_{(X_t,Y_t)}\mathrm{d}z$ for the solution to (2.21) with initial distribution μ_0. So, as explained in the proof of Example 2.1, by Theorem 2.4 we only need to verify **(C)** for B, V in (2.29) and

TTH (2.30) $$\theta_1 = \theta\Big(\frac{1}{2} + \sqrt{2+2\beta+\beta^2}\Big), \quad \theta_2 = \frac{\theta}{2}\sqrt{2+2\beta+\beta^2},$$

so that the desired assertion holds for

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

By (2.27) and $V(x,\mu) := \mu(W(x,\cdot))$, for any constants $\alpha_1, \alpha_2, \alpha_3 > 0$ we have

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

Take

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

We have

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

$$\alpha_2+\alpha_3=\frac{1}{2}\sqrt{2+2\beta+\beta^2}.$$

Therefore,

$$I\geq -\frac{\theta}{2}\sqrt{2+2\beta+\beta^2}\mathbb{W}_2(\mu,\bar{\mu})^2-\theta\Big(\frac{1}{2}+\sqrt{2+2\beta+\beta^2}\Big)|(x,y)-(\bar{x},\bar{y})|^2,$$

i.e. **(C)** holds for B and V in (2.29) where $B=I_m$ implies that ψ_B is the Euclidean distance on $\mathbb{R}^{2m}$, and for θ_1,θ_2 in (2.30). □

3 Proof of Theorems 2.1 and 2.2

Proof of Theorem 2.1. (1) Since

$$\mathrm{Ent}(P_{t_0}^*\nu|P_{t_0}^*\mu)=\sup_{f\geq 0,(P_{t_0}f)(\mu)=1}P_{t_0}(\log f)(\nu),$$

(2.3) implies

$$\mathrm{Ent}(P_{t_0}^*\nu|P_{t_0}^*\mu)\leq c_0\mathbb{W}_2(\mu,\nu)^2.$$

This together with $P_{t_0}^*\mu_\infty=\mu_\infty$ gives

EWW (3.1) $$\mathrm{Ent}(P_{t_0}^*\mu|\mu_\infty)\leq c_0\mathbb{W}_2(\mu,\mu_\infty)^2,\quad \mu\in\mathscr{P}_2.$$

Combining (2.5) with (2.4) and (3.1), we obtain

$$\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2\leq c_1\mathrm{e}^{-\lambda t}\mathbb{W}_2(\mu,\mu_\infty)^2\leq c_1\mathrm{e}^{-\lambda t}\min\big\{\mathbb{W}_2(\mu,\mu_\infty)^2,C\mathrm{Ent}(\mu|\mu_\infty)\big\},\quad t\geq t_1$$

and

$$\begin{aligned}&\mathrm{Ent}(P_t^*\mu|\mu_\infty)=\mathrm{Ent}(P_{t_0}^*P_{t-t_0}^*\mu|\mu_\infty)\leq c_0\mathbb{W}_2(P_{t-t_0}^*\mu,\mu_\infty)^2\leq c_0c_1\mathrm{e}^{-\lambda(t-t_0)}\mathbb{W}_2(\mu,\mu_\infty)^2\\&=\{c_0c_1\mathrm{e}^{\lambda t_0}\}\mathrm{e}^{-\lambda t}\min\big\{\mathbb{W}_2(\mu,\mu_\infty)^2,C\mathrm{Ent}(\mu|\mu_\infty)\big\},\quad t\geq t_0+t_1.\end{aligned}$$

Therefore, (2.6) holds.

(2) Similarly, if (2.7) holds, then (2.4) and (3.1) imply

$$\begin{aligned}&\mathrm{Ent}(P_t^*\mu|\mu_\infty)\leq c_2\mathrm{e}^{-\lambda(t-t_0)}\min\big\{\mathrm{Ent}(P_{t_0}^*\mu|\mu_\infty),\mathrm{Ent}(\mu|\mu_\infty)\big\}\\&\leq c_2\mathrm{e}^{-\lambda(t-t_0)}\min\big\{c_0\mathbb{W}_2(\mu,\mu_\infty)^2,\mathrm{Ent}(\mu|\mu_\infty)\big\},\quad t\geq t_0+t_2\end{aligned}$$

and

$$\begin{aligned}&C^{-1}\mathbb{W}_2(P_t^*\mu,\mu_\infty)^2\leq\mathrm{Ent}(P_{t-t_0}^*P_{t_0}^*\mu|\mu_\infty)\\&\leq c_2\min\big\{\mathrm{e}^{-\lambda t}\mathrm{Ent}(\mu|\mu_\infty),\mathrm{e}^{-\lambda(t-t_0)}\mathrm{Ent}(P_{t_0}^*\mu|\mu_\infty)\big\}\\&\leq c_2\mathrm{e}^{-\lambda(t-t_0)}\min\big\{\mathrm{Ent}(\mu|\mu_\infty),c_0\mathbb{W}_2(\mu,\mu_\infty)^2\big\},\quad t\geq t_0+t_2.\end{aligned}$$

Then (2.8) holds, and the proof is finished. □

Proof of Theorem 2.2. By [11, Theorem 10], there exits a unique $\mu_\infty \in \mathscr{P}_2$ such that

MIN (3.2) $$\mathrm{Ent}^{V,W}(\mu_\infty) = 0.$$

Let $\mu_0 = \rho_0 \mathrm{d}x \in \mathscr{P}_2$. We first note that $\mu_t = P_t^*\mu_0 := \mathscr{L}_{X_t}$ for X_t solving the distribution dependent SDE (2.1) with

SBB (3.3) $$\sigma(x,\mu) = \sqrt{2a(x)}, \quad b(x,\mu) = \sum_{j=1}^d \partial_j a_{\cdot,j}(x) - a\nabla\{V + W \circledast \mu\}(x), \quad x \in \mathbb{R}^d, \mu \in \mathscr{P}_2.$$

Obviously, for this choice of (σ, b), assumptions (H_1) and (H_2) imply condition (2.25) for some constant $K > 0$, so that the distribution dependent SDE is weakly and strongly well-posed. For any $N \geq 2$, let $\mu_t^{(N)} = \mathscr{L}_{X_t^{(N)}}$ for the mean field particle system $X_t^{(N)} = (X_t^{N,k})_{1\leq i\leq N}$:

MF (3.4) $$\mathrm{d}X_t^{N,k} = \sqrt{2}\sigma(X_t^{N,k})\mathrm{d}B_t^k + \Big\{\sum_{j=1}^d \partial_j a_{\cdot,j}(X_t^{N,k}) - a(X_t^{N,k})\nabla_k H_N(X_t^{(N)})\Big\}\mathrm{d}t, \quad t \geq 0,$$

where ∇_k denotes the gradient in the k-th component, and $\{X_0^{N,k}\}_{1\leq i\leq N}$ are i.i.d. with distribution $\mu_0 \in \mathscr{P}_2$. According to the propagation of chaos, see [17], (H_1)-(H_3) imply

SNN (3.5) $$\lim_{N\to\infty} \mathbb{W}_2(\mathscr{L}_{X_t^{N,1}}, P_t^*\mu_0) = 0.$$

Next, our conditions imply (25) and (26) in [11] for $\rho_{LS} = \beta(1-r_0)^2$. So, by [11, Theorem 8(2)], we have the log-Sobolev inequality

LSN (3.6) $$\mu^{(N)}(f^2 \log f^2) \leq \frac{2}{\beta(1-r_0)^2}\mu^{(N)}(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^{dN}), \mu^{(N)}(f^2) = 1.$$

By [5], this implies the Talagrand inequality

TLN (3.7) $$\mathbb{W}_2(\nu^{(N)}, \mu^{(N)})^2 \leq \frac{2}{\beta(1-r_0)^2}\mathrm{Ent}(\nu^{(N)}|\mu^{(N)}), \quad t \geq 0, N \geq 2, \nu^{(N)} \in \mathscr{P}(\mathbb{R}^{dN}).$$

On the other hand, by Itô's formula we see that the generator of the diffusion process $X_t^{(N)}$ is

$$L^{(N)}(x^{(N)}) = \sum_{i,j,k=1}^d \Big\{a_{ij}(x^{N,k})\partial_{x_i^{N,k}}\partial_{x_j^{N,k}} + \partial_j a_{ij}(x^{N,k})\partial_{x_i^{N,k}} - a_{ij}(x^{N,k})\big[\partial_{x_i^{N,k}} H_N(x^{(N)})\big]\partial_{x_i^{N,k}}\Big\},$$

for $x^{(N)} = (x^{N,1}, \cdots, x^{N,N}) \in \mathbb{R}^{dN}$, where $x_i^{N,k}$ is the i-th component of $x^{N,k} \in \mathbb{R}^d$. Using the integration by parts formula, we see that this operator is symmetric in $L^2(\mu^{(N)})$:

$$\mathscr{E}^{(N)}(f,g) := \int_{\mathbb{R}^{dN}} \langle a^{(N)}\nabla f, \nabla g\rangle \mathrm{d}\mu^{(N)} = -\int_{\mathbb{R}^{dN}} f L^{(N)} g \mathrm{d}\mu^{(N)}, \quad f, g \in C_0^\infty(\mathbb{R}^{dN}),$$

where $a^{(N)}(x^{(N)}) := \text{diag}\{a(x^{N,1}), \cdots, a(x^{N,N})\}, x^{(N)} = (x^{N,1}, \cdots, x^{N,N}) \in \mathbb{R}^{dN}$. So, the closure of the pre-Dirichelt form $(\mathscr{E}^{(N)}, C_0^\infty(\mathbb{R}^{dN}))$ in $L^2(\mu^{(N)})$ is the Dieichlet form for the Markov semigroup $P_t^{(N)}$ of $X_t^{(N)}$. By (H_1) we have $a^{(N)} \geq \lambda_a I_{dN}$, so that (3.6) implies

$$\mu^{(N)}(f^2 \log f^2) \leq \frac{2}{\beta\lambda_a(1-r_0)^2}\mathscr{E}^{(N)}(f,f), \quad f \in C_b^1(\mathbb{R}^{dN}), \mu^{(N)}(f^2) = 1.$$

It is well known that this log-Sobolev inequality implies the exponential convergence

EXN (3.8)
$$\begin{aligned}&\text{Ent}(\mu_t^{(N)}|\mu^{(N)}) \leq \text{e}^{-\lambda_a\beta(1-r_0)^2t}\text{Ent}(\mu_0^{(N)}|\mu^{(N)})\\ &= \text{e}^{-\lambda_a\beta(1-r_0)^2t}\text{Ent}(\mu^{\otimes N}|\mu^{(N)}), \quad t \geq 0, N \geq 2,\end{aligned}$$

see for instance [5, Theorem 5.2.1]. Moreover, since Hess_V and Hess_W are bounded from below, (H_1) implies that the Bakry-Emery curvature of the generator of $X_t^{(N)}$ is bounded by a constant. Then according to [20], there exists a constant $K \geq 0$ such that the Markov semigroup $P_t^{(N)}$ of $X_t^{(N)}$ satisfies the log-Harnack inequality

PLM (3.9)
$$P_t^{(N)} \log f(x) \leq \log P_t^{(N)} f(y) + \frac{K\rho^{(N)}(x,y)^2}{2(1-\text{e}^{-2Kt})}, \quad 0 < f \in \mathscr{B}_b(\mathbb{R}^{dN}), t > 0, x, y \in \mathbb{R}^{dN},$$

where $\rho^{(N)}$ is the intrinsic distance induced by the Dirichlet form $\mathscr{E}^{(N)}$. Since $a^{(N)} \geq \lambda_a I_{dN}$, we have $\rho^{(N)}(x,y)^2 \leq \lambda_a^{-1}|x-y|^2$. So, (3.9) implies (2.3) for $P_t^{(N)}$ replacing P_{t_0} and $c_0 = \frac{K}{2\lambda_a(1-\text{e}^{-2Kt})}$:

$$P_t^{(N)}(\log f)(\nu) \leq \log P_t^{(N)} f(\mu) + \frac{K|\mathbb{W}_2(\mu,\nu)^2}{2\lambda_a(1-\text{e}^{-2Kt})}, \quad 0 < f \in \mathscr{B}_b(\mathbb{R}^{dN}), t > 0, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^{dN}).$$

Thus, by Theorem 2.1, (3.8) implies

EXN2 (3.10)
$$\mathbb{W}_2(\mu_t^{(N)}|\mu^{(N)})^2 \leq \frac{c_1\text{e}^{-\lambda_a\beta(1-r_0)^2t}}{1\wedge t}\mathbb{W}_2(\mu^{\otimes N}, \mu^{(N)})^2, \quad t > 0, N \geq 2$$

for some constant $c_1 > 0$. Moreover, (3.7), (3.2) and [11, Lemma 17] yield

EXN3 (3.11)
$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\mathbb{W}_2(\mu_\infty^{\otimes N}, \mu^{(N)})^2 &\leq \limsup_{N\to\infty}\frac{2}{\beta(1-r_0)^2N}\text{Ent}(\mu_\infty^{\otimes N}|\mu^{(N)})^2\\ &= \frac{2}{\beta(1-r_0)^2}\text{Ent}^{V,W}(\mu_\infty) = 0.\end{aligned}$$

Combining this with (3.10) we derive

EXN5 (3.12)
$$\begin{aligned}&\limsup_{N\to\infty}\frac{1}{N}\mathbb{W}_2(\mu_t^{(N)}, \mu_\infty^{\otimes N})^2 = \limsup_{N\to\infty}\frac{1}{N}\mathbb{W}_2(\mu_t^{(N)}, \mu^{(N)})^2\\ &\leq \frac{c_1\text{e}^{-\lambda_a\beta(1-r_0)^2t}}{1\wedge t}\limsup_{N\to\infty}\frac{1}{N}\mathbb{W}_2(\mu_0^{\otimes N}, \mu^{(N)})^2\\ &= \frac{c_1\text{e}^{-\lambda_a\beta(1-r_0)^2t}}{1\wedge t}\limsup_{N\to\infty}\frac{1}{N}\mathbb{W}_2(\mu_0^{\otimes N}, \mu_\infty^{\otimes N})^2 = \frac{c_1\text{e}^{-\lambda_a\beta(1-r_0)^2t}}{1\wedge t}\mathbb{W}_2(\mu_0,\mu_\infty)^2\ , \quad t > 0.\end{aligned}$$

Now, let $\xi = (\xi_i)_{1\le i\le N}$ and $\eta = (\eta_i)_{1\le i\le N}$ be random variables on $\mathbb{R}^{dN}$ such that $\mathscr{L}_\xi = \mu_t^{(N)}, \mathscr{L}_\eta = \mu_\infty^{\otimes N}$ and

$$\sum_{i=1}^N \mathbb{E}|\xi_i - \eta_i|^2 = \mathbb{E}|\xi - \eta|^2 = \mathbb{W}_2(\mu_t^{(N)}, \mu_\infty^{\otimes N})^2.$$

We have $\mathscr{L}_{\xi_i} = \mathscr{L}_{X_t^{N,1}}, \mathscr{L}_{\eta_i} = \mu_\infty$ for any $1 \le i \le N$, so that

EXX (3.13) $$N\mathbb{W}_2(\mathscr{L}_{X_t^{N,1}}, \mu_\infty)^2 \le \sum_{i=1}^N \mathbb{E}|\xi_i - \eta_i|^2 = \mathbb{W}_2(\mu_t^{(N)}, \mu_\infty^{\otimes N})^2.$$

Substituting this into (3.12), we arrive at

$$\limsup_{N\to\infty} \mathbb{W}_2(\mathscr{L}_{X_t^{N,1}}, \mu_\infty)^2 \le \frac{c_1 \mathrm{e}^{-\lambda_a \beta (1-r_0)^2 t}}{1\wedge t} \mathbb{W}_2(\mu, \mu_\infty)^2 \ , \quad t > 0.$$

This and (3.5) imply

EXN6 (3.14) $$\mathbb{W}_2(P_t^*\mu, \mu_\infty)^2 \le \frac{c_1 \mathrm{e}^{-\lambda_a \beta (1-r_0)^2 t}}{1\wedge t} \mathbb{W}_2(\mu, \mu_\infty)^2 \ , \quad t > 0.$$

By [25, Theorem 4.1], (H_1)-(H_3) imply the log-Harnack inequality

LHK (3.15) $$P_t(\log f)(\nu) \le \log P_t f(\mu) + \frac{c_2}{1\wedge t} \mathbb{W}_2(\mu, \nu)^2, \quad \mu, \nu \in \mathscr{P}_2, t > 0$$

for some constant $c_2 > 0$. Similarly to the proof of (3.13) we have

$$N\mathbb{W}_2(\mu_\infty, \mu^{(N,1)})^2 \le \mathbb{W}_2(\mu_\infty^{\otimes N}, \mu^{(N)})^2,$$

where $\mu^{(N,1)} := \mu^{(N)}(\cdot \times \mathbb{R}^{d(N-1)})$ is the first marginal distribution of $\mu^{(N)}$. This together with (3.11) implies

$$\lim_{N\to\infty} \mathbb{W}_2(\mu^{(N,1)}, \mu_\infty)^2 = 0.$$

Therefore, applying (3.6) to $f(x)$ depending only on the firs component x_1, and letting $N \to \infty$, we derive the log-Sobolev inequality

$$\mu_\infty(f^2 \log f^2) \le \frac{2}{\beta(1-r_0)^2} \mu_\infty(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^d), \mu_\infty(f^2) = 1.$$

By [5], this implies (2.4) for $C = \frac{2}{\beta(1-r_0)^2}$. Combining this with the log-Harnack inequality and (3.14), by Theorem 2.1 we prove (2.15) for some constant $c > 0$ and $\mu_t = \mathscr{L}_{X_t} = P_t^*\mu_0$ for solutions to (2.1) with b, σ in (3.3).

Similarly to the link of (1.3) and (1.5) shown in Introduction, for any probability density functions ρ_t solving (2.10), we have $\rho_t \mathrm{d}x = P_t^*\mu_0$ for $\mu_0 = \rho_0 \mathrm{d}x \in \mathscr{P}_2$. So, we have proved (2.15) for ρ_t solving (2.10) with $\mu_0 \in \mathscr{P}_2$. As explained in the proof of Example 2.1 that $\mathrm{Ent}(\mu_0, \mu_\infty) = \mathbb{W}_2(\mu, \mu_\infty) = \infty$ for $\mu_0 \notin \mathscr{P}_2$, so that the desired inequality (2.15) tribally true. Then the proof is finished. □

4 Proof of Theorems 2.3

According to [25, Theorem 3.1], (A_1) and (A_2) imply that P_t^* has a unique invariant probability measure μ_∞ and

001 (4.1) $$\mathbb{W}_2(P_t^*\mu,\mu_\infty)\le \mathrm{e}^{-\frac{1}{2}(K_2-K_1)t}\mathbb{W}_2(\mu,\mu_\infty),\quad t\ge 0,\mu\in\mathscr{P}_2,$$

while [25, Corollary 4.3] implies

$$\mathrm{Ent}(P_t^*\mu|\mu_\infty)\le \frac{c_0}{1\wedge t}\mathbb{W}_2(\mu,\mu_\infty)^2,\quad t>0,\mu\in\mathscr{P}_2$$

for some constant $c_0>0$. Then for any $p>1$, combining these with $P_t^*=P_{1\wedge t}^*P_{(t-1)^+}^*$, we obtain

$$\begin{aligned}&\mathrm{Ent}(P_t^*\mu|\mu_\infty)=\mathrm{Ent}(P_{1\wedge t}^*P_{(t-1)^+}^*\mu|\mu_\infty)\le \frac{c_0}{1\wedge t}\mathbb{W}_2(P_{(t-1)^+}^*\mu,\mu_\infty)^2\\ &\le \frac{c_0\mathrm{e}^{-(K_2-K_1)(t-1)^+}}{1\wedge t}\mathbb{W}_2(\mu,\mu_\infty)^2=\frac{c_0\mathrm{e}^{K_2-K_1}}{1\wedge t}\mathrm{e}^{-(K_2-K_1)t}\mathbb{W}_2(\mu,\mu_\infty)^2.\end{aligned}$$

This together with (4.1) implies (2.16) for some constant $c_1>0$.

Now, let $\sigma\in C_b^2(\mathbb{R}^d\to\mathbb{R}^d\otimes\mathbb{R}^m)$. To deduce (2.17) from (2.16), it remains to find a constant $c>0$ such that the following Talagrand inequality holds:

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

According to [5], this inequality follows from the log-Sobolev inequality

LSII (4.2) $$\mu_\infty(f^2\log f^2)\le c\mu_\infty(|\nabla f|^2),\quad f\in C_b^1(\mathbb{R}^d),\mu_\infty(f^2)=1.$$

To prove this inequality, we consider the diffusion process $\bar{X}_t$ on $\mathbb{R}^d$ generated by

$$\bar{L}:=\frac{1}{2}\sum_{i,j=1}^d(\sigma\sigma^*)_{ij}\partial_i\partial_j+\sum_{i=1}^\infty b_i(\cdot,\mu_\infty)\partial_i,$$

which can be constructed by solving the SDE

BSDE (4.3) $$\mathrm{d}\bar{X}_t=\sigma(\bar{X}_t)\mathrm{d}W_t+b(\bar{X}_t,\mu_\infty)\mathrm{d}t.$$

Let $\bar{P}_t$ be the associated Markov semigroup. Since $P_t^*\mu_\infty=\mu_\infty$, when $\mathscr{L}_{X_0}=\mu_\infty$ the SDE (4.3) coincides with (1.7) so that by the uniqueness, we see that μ_∞ is an invariant probability measure of $\bar{P}_t$. Combining this with (A_2) and Itô's formula, we obtain

ECC (4.4) $$\mathbb{W}_2(\mathscr{L}_{\bar{X}_t},\mu_\infty)^2\le \mathrm{e}^{-K_2t}\mathbb{W}_2(\mathscr{L}_{\bar{X}_0},\mu_\infty)^2,\quad t>0.$$

To prove the log-Sobolev inequality (4.2), we first verify the hyperboundedness of $\bar{P}_t$, i.e. for large $t>0$ we have

HP (4.5) $$\|\bar{P}_t\|_{L^2(\mu_\infty)\to L^4(\mu_\infty)}<\infty.$$

It is easy to see that conditions (A_1) and (A_2) in Theorem 2.3 imply that σ and $b(\cdot,\mu_\infty)$ satisfy conditions (A1)-(A3) in [21] for $K=-(K_2-K_1), \lambda_t^2=\lambda$ and $\delta_t=\|\sigma\|_\infty$. So, by [21, Theorem 1.1(3)], we find a constant $C>0$ such that the following Harnack inequality holds:

$$(\bar{P}_t f(x))^2 \le \bar{P}_t f^2(y)\exp\Big[\frac{C|x-y|^2}{\mathrm{e}^{(K_2-K_1)t}-1}\Big],\quad t>0.$$

Then for any f with $\mu_\infty(f^2)\le 1$, we have

$$\begin{aligned}&\big(\bar{P}_t f(x)\big)^2\int_{\mathbb{R}^d}\exp\Big[-\frac{C|x-y|^2}{\mathrm{e}^{(K_2-K_1)t}-1}\Big]\mu_\infty(\mathrm{d}y)\\ &\le \mu_\infty(\bar{P}_t f^2)=\mu_\infty(f^2)\le 1.\end{aligned}$$

So,

UPP
$$\begin{aligned}\sup_{\mu_\infty(f^2)\le 1}|\bar{P}_t f(x)|^4 &\le \frac{1}{\big(\int_{\mathbb{R}^d}\mathrm{e}^{-\frac{C|x-y|^2}{\mathrm{e}^{(K_2-K_1)t}-1}}\mu_\infty(\mathrm{d}y)\big)^2}\\ &\le \frac{1}{\big(\int_{B(0,1)}\mathrm{e}^{-\frac{C|x-y|^2}{\mathrm{e}^{(K_2-K_1)t}-1}}\mu_\infty(\mathrm{d}y)\big)^2}\le C_1\exp\big[C_1\mathrm{e}^{-(K_2-K_1)t}|x|^2\big],\quad t\ge 1, x\in\mathbb{R}^d.\end{aligned}\tag{4.6}$$

Next, by $\|\sigma\|_\infty<\infty$, (A_2) and Itô's formula, for any $k\in(0,K_2)$ there exists a constant $c_k>0$ such that

$$\mathrm{d}|\bar{X}_t|^2\le 2\langle\bar{X}_t,\sigma(\bar{X}_t)\mathrm{d}W_t\rangle+\big\{c_k-k|\bar{X}_t|^2\big\}\mathrm{d}t.$$

Then for any $\varepsilon>0$,

PW
$$\mathrm{d}\mathrm{e}^{\varepsilon|\bar{X}_t|^2}\le 2\varepsilon\mathrm{e}^{\varepsilon|\bar{X}_t|^2}\langle\bar{X}_t,\sigma(\bar{X}_t)\mathrm{d}W_t\rangle+\varepsilon\mathrm{e}^{\varepsilon|\bar{X}_t|^2}\big\{c_k+2\varepsilon\|\sigma\|_\infty^2|\bar{X}_t|^2-k|\bar{X}_t|^2\big\}\mathrm{d}t.\tag{4.7}$$

When $\varepsilon>0$ is small enough such that $2\varepsilon\|\sigma\|_\infty^2<K_2$, there exist constants $c_1(\varepsilon),c_2(\varepsilon)>0$ such that

$$\varepsilon\mathrm{e}^{\varepsilon|\bar{X}_t|^2}\big\{c_k+2\varepsilon\|\sigma\|_\infty^2|\bar{X}_t|^2-k|\bar{X}_t|^2\big\}\le c_1(\varepsilon)-c_2(\varepsilon)\mathrm{e}^{\varepsilon|\bar{X}_t|^2}.$$

Combining this with (4.7) we obtain

$$\mathrm{d}\mathrm{e}^{\varepsilon|\bar{X}_t|^2}\le\big\{c_1(\varepsilon)-c_2(\varepsilon)\big\}\mathrm{e}^{\varepsilon|\bar{X}_t|^2}\mathrm{d}t+2\varepsilon\mathrm{e}^{\varepsilon|\bar{X}_t|^2}\langle\bar{X}_t,\sigma(\bar{X}_t)\mathrm{d}W_t\rangle.$$

So, letting for instance $\bar{X}_0=0$ we get

$$\frac{c_2(\varepsilon)}{t}\int_0^t\mathbb{E}\mathrm{e}^{\varepsilon|\bar{X}_s|^2}\mathrm{d}s\le\frac{1+c_1(\varepsilon)t}{t},\quad t>0.$$

This together with (4.4) yields

$$\mu_\infty(\mathrm{e}^{\varepsilon(|\cdot|^2\wedge N)})=\lim_{t\to\infty}\frac{1}{t}\int_0^t\mathbb{E}\mathrm{e}^{\varepsilon(|\bar{X}_s|^2\wedge N)}\mathrm{d}s\le\frac{c_1(\varepsilon)}{c_2(\varepsilon)},\quad N>0.$$

By letting $N \to \infty$ we derive $\mu_\infty(\mathrm{e}^{\varepsilon|\cdot|^2}) < \infty$. Obviously, this and (4.6) imply (4.5) for large $t > 0$. Moreover, since $\|(\sigma\sigma^*)^{-1}\|_\infty < \infty$, $\sigma \in C_b^2(\mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^m)$ and noting that (A_2) gives

$$\langle v, \nabla_v b(\cdot, \mu_\infty)\rangle \le -K_2|v|^2, \quad v \in \mathbb{R}^d,$$

we find a constant $K_0 \in \mathbb{R}$ such that for any $f \in C^\infty(\mathbb{R}^d)$,

$$\Gamma_2(f) := \frac{1}{2}\bar{L}|\sigma^*\nabla f|^2 - \langle \sigma^*\nabla f, \sigma^*\nabla \bar{L} f\rangle \ge K_0|\sigma^*\nabla f|^2,$$

i.e. the Bakry-Emery curvature of $\bar{L}$ is bounded below by a constant K_0. According to [16, Theorem 2.1], this and the hyperboundedness (4.5) imply the defective log-Sobolev inequality

DFL (4.8)
$$\begin{aligned} &\mu_\infty(f^2 \log f^2) \le C_1\mu_\infty(|\sigma^*\nabla f|^2) + C_2 \\ &\le C_1\|\sigma\|_\infty^2\mu_\infty(|\nabla f|^2) + C_2, \quad f \in C_b^1(\mathbb{R}^d), \mu_\infty(f^2) = 1 \end{aligned}$$

for some constants $c_1, c_2 > 0$. Since $\bar{L}$ is elliptic, the invariant probability measure μ_∞ is equivalent to the Lebesgue measure, see for instance [6, Theorem 1.1(ii)], so that the Dirichlet form

$$\mathscr{E}(f, g) := \mu_\infty(\langle \nabla f, \nabla g\rangle), \quad f, g \in W^{1,2}(\mu)$$

is irreducible, i.e. $f \in W^{1,2}(\mu)$ and $\mathscr{E}(f, f) = 0$ imply that f is constant. So, by [23, Corollary 1.3], see also [14], the defective log-Sobolev inequality (4.8) implies the desired log-Sobolev inequality (4.2) for some constant $c > 0$. Hence, the proof is finished.

5 Proof of Theorem 2.4

We first establish the log-Harnack inequality for a more general model, which extends existing results derived in [12, 4] to the distribution dependent setting.

5.1 Log-Harnack inequality

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

E' (5.1)
$$\begin{cases} \mathrm{d}X_t = \big(AX_t + BY_t\big)\mathrm{d}t, \\ \mathrm{d}Y_t = Z((X_t, Y_t), \mathscr{L}_{(X_t,Y_t)})\mathrm{d}t + \sigma\mathrm{d}W_t, \end{cases}$$

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

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

is measurable. We assume

(C) σ is invertible, Z is Lipschitz continuous, and A, B satisfy the following Kalman's rank condition for some $k \ge 1$:

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

Obviously, this assumption implies (A_1), so that (5.1) has a unique solution (X_t, Y_t) for any initial value (X_0, Y_0) with $\mu := \mathscr{L}_{(X_0,Y_0)} \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$. Let $P_t^*\mu := \mathscr{L}_{(X_t,Y_t)}$ and

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

By [25, Theorem 3.1], the Lipschitz continuity of Z implies

ESO (5.2) $$\mathbb{W}_2(P_t^*\mu, P_t^*\nu) \leq \mathrm{e}^{Kt}\mathbb{W}_2(\mu,\nu), \quad t \geq 0, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$$

for some constant $K > 0$. We have the following result.

P1 **Proposition 5.1.** *Assume* **(C)**. *Then there exists a constant $c > 0$ such that*

LH (5.3) $$(P_T \log f)(\nu) \leq \log(P_T f)(\mu) + \frac{c\mathrm{e}^{cT}}{T^{4k-1} \wedge 1}\mathbb{W}_2(\mu,\nu)^2, \quad T > 0, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}).$$

Consequently,

LH' (5.4) $$\mathrm{Ent}(P_T^*\nu | P_T^*\mu) \leq \frac{c\mathrm{e}^{cT}}{T^{4k-1} \wedge 1}\mathbb{W}_2(\mu,\nu)^2, \quad T > 0, \mu, \nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}).$$

Proof. According to [25, Corollary 4.3], (5.3) implies (5.4). Below we prove (5.3) by using the coupling by change of measures summarized in [22, Section 1.1]. By the Kalman rank condition in **(C)**,

$$Q_T := \int_0^T t(T-t)\mathrm{e}^{(T-t)A}BB^*\mathrm{e}^{(T-t)A^*}\mathrm{d}t$$

is invertible and there exists a constant $c_1 > 0$ such that

EQ (5.5) $$\|Q_T^{-1}\| \leq \frac{c_1\mathrm{e}^{c_1 T}}{(T \wedge 1)^{2k+1}}, \quad T > 0,$$

see for instance [26, Theorem 4.2(1)].

Let $(X_0, Y_0), (\bar{X}_0, \bar{Y}_0) \in L^2(\Omega \to \mathbb{R}^{d_1+d_2}, \mathscr{F}_0, \mathbb{P})$ such that $\mathscr{L}_{(X_0,Y_0)} = \mu, \mathscr{L}_{(\bar{X}_0,\bar{Y}_0)} = \nu$ and

E'2 (5.6) $$\mathbb{E}\big(|X_0 - \bar{X}_0|^2 + |Y_0 - \bar{Y}_0|^2\big) = \mathbb{W}_2(\mu,\nu)^2.$$

Next, let (X_t, Y_t) solve (5.1). Then $\mathscr{L}_{(X_t,Y_t)} = P_t^*\mu$. Consider the the modified equation with initial value $(\bar{X}_0, \bar{Y}_0)$:

E'3 (5.7) $$\begin{cases} \mathrm{d}\bar{X}_t = \big(A\bar{X}_t + B\bar{Y}_t\big)\mathrm{d}t, \\ \mathrm{d}\bar{Y}_t = \Big\{Z((X_t,Y_t), P_t^*\mu) + \frac{Y_0 - \bar{Y}_0}{T} + \frac{\mathrm{d}}{\mathrm{d}t}\big[t(T-t)B^*\mathrm{e}^{(T-t)A^*}v\big]\Big\}\mathrm{d}t + \sigma\mathrm{d}W_t, \end{cases}$$

where

EV (5.8) $$v := Q_T^{-1}\bigg\{\mathrm{e}^{TA}(X_0 - \bar{X}_0) + \int_0^T \frac{t-T}{T}\mathrm{e}^{(T-t)A}B(\bar{Y}_0 - Y_0)\mathrm{d}t\bigg\}.$$

Then

ES1 (5.9)
$$\begin{aligned}\bar{Y}_t - Y_t &= \bar{Y}_0 - Y_0 + \int_0^t \Big\{\frac{Y_0 - \bar{Y}_0}{T} + \frac{\mathrm{d}}{\mathrm{d}r}\big[r(T-r)B^*\mathrm{e}^{(T-r)A^*}v\big]\Big\}\mathrm{d}r \\ &= \frac{T-t}{T}(\bar{Y}_0 - Y_0) + t(T-t)B^*\mathrm{e}^{(T-t)A^*}v, \quad t \in [0,T].\end{aligned}$$

Consequently, $Y_T = \bar{Y}_T$, and combining with Duhamel's formula, we obtain

ES2 (5.10)
$$\bar{X}_t - X_t = \mathrm{e}^{tA}(\bar{X}_0 - X_0) + \int_0^t \mathrm{e}^{(t-r)A}B\Big\{\frac{T-r}{T}(\bar{Y}_0 - Y_0) + r(T-r)B^*\mathrm{e}^{(T-r)A^*}v\Big\}\mathrm{d}r$$

for $t \in [0,T]$. This and (5.8) imply

$$\bar{X}_T - X_T = \mathrm{e}^{TA}(\bar{X}_0 - X_0) + \int_0^T \frac{T-r}{T}\mathrm{e}^{(T-r)A}B(\bar{Y}_0 - Y_0)\mathrm{d}r + Q_T v = 0,$$

which together with $Y_T = \bar{Y}_T$ observed above yields

CP (5.11)
$$(X_T, Y_T) = (\bar{X}_T, \bar{Y}_T).$$

On the other hand, let

$$\xi_t = \sigma^{-1}\Big\{\frac{1}{T}(Y_0 - \bar{Y}_0) + \frac{\mathrm{d}}{\mathrm{d}t}\Big[t(T-t)B^*\mathrm{e}^{(T-t)A^*}v\Big] + Z\big((X_t, Y_t), P_t^*\mu\big) - Z\big((\bar{X}_t, \bar{Y}_t), P_t^*\nu\big)\Big\}, \ \ t \in [0,T].$$

By **(C)**, (5.2), (5.5), (5.8), (5.9), and (5.10), we find a constant $c_2 > 0$ such that

E'4 (5.12)
$$|\xi_t|^2 \le \frac{c_2}{(T \wedge 1)^{4k}}\mathrm{e}^{c_2 T}\big\{|X_0 - \bar{X}_0|^2 + |Y_0 - \bar{Y}_0|^2 + \mathbb{W}_2(\mu,\nu)^2\big\}, \quad t \in [0,T].$$

So, the Girsanov theorem implies that

$$\tilde{W}_t := W_t + \int_0^t \xi_s ds, \quad t \in [0,T]$$

is a d_2-dimensional Brownian motion under the probability measure $\mathbb{Q} := R\mathbb{P}$, where

RR (5.13)
$$R := \mathrm{e}^{-\int_0^T \langle \xi_t, \mathrm{d}W_t\rangle - \frac{1}{2}\int_0^T |\xi_t|^2 \mathrm{d}t}.$$

Reformulating (5.7) as

$$\begin{cases} \mathrm{d}\bar{X}_t = \big(A\bar{X}_t + B\bar{Y}_t\big)\mathrm{d}t, \\ \mathrm{d}\bar{Y}_t = Z((\bar{X}_t, \bar{Y}_t), P_t^*\nu)\mathrm{d}t + \sigma \mathrm{d}\tilde{W}_t, \quad t \in [0,T], \end{cases}$$

by the weak uniqueness of (5.1) and that the distribution of $(\bar{X}_0, \bar{Y}_0)$ under $\mathbb{Q}$ coincides with $\mathscr{L}_{(\bar{X}_0,\bar{Y}_0)} = \nu$, we obtain $\mathscr{L}_{(\bar{X}_t,\bar{Y}_t)|\mathbb{Q}} = P_t^*\nu$ for $t \in [0,T]$. Combining this with (5.11) and using the Young inequality, for any $f \in \mathscr{B}_b^+(\mathbb{R}^{d_1+d_2})$ we have

LH3 (5.14)
$$\begin{aligned}&(P_T \log f)(\nu) = \mathbb{E}[R \log f(\bar{X}_T, \bar{Y}_T)] = \mathbb{E}[R \log f(X_T, Y_T)] \\ &\le \log \mathbb{E}[f(X_T, Y_T)] + \mathbb{E}[R \log R] = \log(P_T f)(\mu) + \mathbb{E}_{\mathbb{Q}}[\log R].\end{aligned}$$

By (5.12), and (5.13), $\tilde{W}_t$ is a Brownian motion under $\mathbb{Q}$, and noting that $\mathbb{Q}|_{\mathscr{F}_0} = \mathbb{P}|_{\mathscr{F}_0}$ and (5.6) imply

$$\mathbb{E}_{\mathbb{Q}}\big(|X_0 - \bar{X}_0|^2 + |Y_0 - \bar{Y}_0|^2\big) = \mathbb{W}_2(\mu,\nu)^2,$$

we find a constant $c > 0$ such that

$$\mathbb{E}_{\mathbb{Q}}[\log R] = \frac{1}{2}\mathbb{E}_{\mathbb{Q}}\int_0^T |\xi_t|^2 \mathrm{d}t \le \frac{c\mathrm{e}^{cT}}{(T\wedge 1)^{4k-1}}\mathbb{W}_2(\mu,\nu)^2.$$

Therefore, (5.3) follows from (5.14). □

5.2 Proof of Theorem 2.4

We first prove the exponential convergence of P_t^* in $\mathbb{W}_2$.

LN1 **Lemma 5.2.** *Assume* **(C)**. *Then there exists a constant $c_1 > 0$ such that*

AC0 (5.15) $$\mathbb{W}_2(P_t^*\mu, P_t^*\nu)^2 \le c_1\mathrm{e}^{-\kappa t}\mathbb{W}_2(\mu,\nu)^2, \quad t \ge 0, \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}).$$

Consequently, P_t^* has a unique invariant probability measure $\mu_\infty \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$.

Proof. As shown in the proof of [25, Theorem 3.1(2)] that the second assertion follows from the first. So, it suffices to prove (5.15). For

AC1 (5.16) $$a := \Big(\frac{1+\beta+\beta^2}{1+\beta}\Big)^{\frac{1}{2}}, \quad r := a - \frac{\beta}{a} = \frac{1}{\sqrt{(1+\beta)(1+\beta+\beta^2)}} \in (0,1),$$

we define the distance

AC2 (5.17) $$\bar{\psi}_B((x,y),(\bar{x},\bar{y})) := \sqrt{a^2|x-\bar{x}|^2 + |B(y-\bar{y})|^2 + 2ra\langle x-\bar{x}, B(y-\bar{y})\rangle}$$

for $(x,y),(\bar{x},\bar{y}) \in \mathbb{R}^{d_1+d_2}$. Then there exists a constant $C > 1$ such that

ACC (5.18) $$C^{-1}|(x-\bar{x},y-\bar{y})| \le \bar{\psi}_B((x,y),(\bar{x},\bar{y})) \le C|(x-\bar{x},y-\bar{y})|.$$

Moreover, we claim that

AC3 (5.19) $$\bar{\psi}_B((x,y),(\bar{x},\bar{y}))^2 \le \frac{2+2\beta+\beta^2+\sqrt{\beta^4+4}}{2(1+\beta)}\psi_B((x,y),(\bar{x},\bar{y}))^2.$$

Indeed, by (5.16) and (5.17), for any $\varepsilon > 0$ we have

AC (5.20) $$\bar{\psi}_B((x,y),(\bar{x},\bar{y}))^2 \le a^2(1+\varepsilon)|x-\bar{x}|^2 + \Big(1 + \frac{1}{\varepsilon(1+\beta)(1+\beta+\beta^2)}\Big)|B(y-\bar{y})|^2.$$

Obviously, by (5.16),

$$\varepsilon := \frac{1-a^2+\sqrt{(a^2-1)^2 + 4a^2(1+\beta)^{-1}(1+\beta+\beta^2)^{-1}}}{2a^2} = \frac{\sqrt{\beta^4+4}-\beta^2}{2(1+\beta+\beta^2)}$$

satisfies

$$a^2(1+\varepsilon) = 1 + \frac{1}{\varepsilon(1+\beta)(1+\beta+\beta^2)} = \frac{2+2\beta+\beta^2+\sqrt{\beta^4+4}}{2(1+\beta)}.$$

Thus, (5.19) follows from (5.20).

Now, let (X_t, Y_t) and $(\bar{X}_t, \bar{Y}_t)$ solve (2.21) with $\mathscr{L}_{(X_0,Y_0)} = \mu, \mathscr{L}_{(\bar{X}_0,\bar{Y}_0)} = \nu$ such that

AC4 (5.21) $$\mathbb{W}_2(\mu,\nu)^2 = \mathbb{E}|(X_0 - \bar{X}_0, Y_0 - \bar{Y}_0)|^2.$$

Simply denote $\mu_t = \mathscr{L}_{(X_t,Y_t)}, \bar{\mu}_t = \mathscr{L}_{(\bar{X}_t,\bar{Y}_t)}$. By **(C)** and Itô's formula, and noting that (5.16) implies

$$a^2 - \beta - ra = 0, \quad 1 - ra = ra\beta = \frac{\beta}{1+\beta},$$

we obtain

$$\begin{aligned}
&\frac{1}{2}\mathrm{d}\big\{\bar{\psi}_B((X_t,Y_t),(\bar{X}_t,\bar{Y}_t))^2\big\} = \big\langle a^2(X_t-\bar{X}_t) + raB(Y_t-\bar{Y}_t), B(Y_t-\bar{Y}_t)\big\rangle\mathrm{d}t \\
&+ \big\langle B^*B(Y_t-\bar{Y}_t) + raB^*(X_t-\bar{X}_t),\ \beta B^*(BB^*)^{-1}(\bar{X}_t-X_t) + \bar{Y}_t - Y_t\big\rangle\mathrm{d}t \\
&+ \big\langle B^*B(Y_t-\bar{Y}_t) + raB^*(X_t-\bar{X}_t),\ B^*\{\nabla V(\bar{X}_t,\bar{\mu}_t) - \nabla V(X_t,\mu_t)\}\big\rangle\mathrm{d}t \\
&\le \Big\{ -(1-ra)|B(Y_t-\bar{Y}_t)|^2 + (a^2-\beta-ra)\langle X_t-\bar{X}_t, B(Y_t-\bar{Y}_t)\rangle - ra\beta|X_t-\bar{X}_t|^2 \\
&\qquad + \big\langle B^*B(Y_t-\bar{Y}_t) + (1+\beta)^{-1}B^*(X_t-\bar{X}_t),\ B^*\{\nabla V(\bar{X}_t,\bar{\mu}_t) - \nabla V(X_t,\mu_t)\}\big\rangle\Big\}\mathrm{d}t \\
&\le \Big\{\frac{\theta_2}{1+\beta}\mathbb{W}_2^{\psi_B}(\mu_t,\bar{\mu}_t)^2 - \frac{\beta-\theta_1}{1+\beta}\psi_B((X_t,Y_t),(\bar{X}_t,\bar{Y}_t))^2\Big\}\mathrm{d}t.
\end{aligned}$$

By (5.19) and the fact that

$$\mathbb{W}_2^{\psi_B}(\mu_t,\bar{\mu}_t)^2 \le \mathbb{E}[\psi_B((X_t,Y_t),(\bar{X}_t,\bar{Y}_t))^2],$$

for $\kappa > 0$ in (2.25) we obtain

$$\begin{aligned}
&\frac{1}{2}\big\{\mathbb{E}[\bar{\psi}_B((X_t,Y_t),(\bar{X}_t,\bar{Y}_t))^2] - \mathbb{E}[\bar{\psi}_B((X_s,Y_s),(\bar{X}_s,\bar{Y}_s))^2]\big\} \\
&\le -\frac{\beta-\theta_1-\theta_2}{1+\beta}\int_s^t \mathbb{E}[\psi_B((X_r,Y_r),(\bar{X}_r,\bar{Y}_r))^2]\mathrm{d}r \\
&\le -\kappa\int_s^t \mathbb{E}[\bar{\psi}_B((X_r,Y_r),(\bar{X}_r,\bar{Y}_r))^2]\mathrm{d}r, \quad t \ge s \ge 0.
\end{aligned}$$

Therefore, the Gronwall's inequality implies

$$\mathbb{E}[\bar{\psi}_B((X_t,Y_t),(\bar{X}_t,\bar{Y}_t))^2] \le \mathrm{e}^{-2\kappa t}\mathbb{E}[\bar{\psi}_B((X_0,Y_0),(\bar{X}_0,\bar{Y}_0))^2], \quad t \ge 0.$$

Combining this with (5.18) and (5.21), we prove (5.15) for some constant $c > 0$. □

Proof of Theorem 2.4. By Proposition 5.1 with $k = 1$, Lemma 5.2 and Theorem 2.1, we only need to verify the Talagrand inequality. As shown in the beginning of [10, Section 3] that μ_∞ has the representation

$$\mu_\infty(\mathrm{d}x, \mathrm{d}y) = Z^{-1}\mathrm{e}^{\bar{V}(x,y)}\mathrm{d}x\mathrm{d}y, \quad \bar{V}(x,y) := V(x,\mu_\infty) + \frac{\beta}{2}|(BB^*)^{-\frac{1}{2}}x|^2 + \frac{1}{2}|y|^2,$$

where $Z := \int_{\mathbb{R}^{d_1+d_2}} \mathrm{e}^{-\bar{V}(x,y)}\mathrm{d}x\mathrm{d}y$ is the normalization constant. Since (2.23) implies

$$BB^*\mathrm{Hess}_{V(\cdot,\mu_\infty)} \geq -\theta_1 I_{d_1},$$

we deduce from (2.22) that

$$\mathrm{Hess}_{\bar{V}} \geq \gamma I_{d_1+d_2}, \quad \gamma := 1 \wedge \frac{\beta - \theta_1}{\|B\|^2} > 0.$$

So, by the Bakry-Emery criterion [2], we have the log-Sobolev inequality

$$\mu_\infty(f^2 \log f^2) \leq \frac{2}{\gamma}\mu_\infty(|\nabla f|^2), \quad f \in C_b^1(\mathbb{R}^{d_1+d_2}), \mu_\infty(f^2) = 1.$$

According to [5], this implies the Talagrand inequality

$$\mathbb{W}_2(\mu, \mu_\infty)^2 \leq \frac{2}{\gamma}\mathrm{Ent}(\mu|\mu_\infty).$$

Then the proof if finished. □

References

Aida [1] S. Aida, *Uniformly positivity improving property, Sobolev inequalities and spectral gap,* J. Funct. Anal. 158(1998), 152–185.

BE84 [2] D. Bakry, M. Emery, *Hypercontractivité de semi-groupes de diffusion*, C. R. Acad. Sci. Paris. Sér. I Math. 299(1984), 775–778.

BGL [3] D. Bakry, I. Gentil, M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators,* Springer, Berlin, 2014.

BWY15 [4] J. Bao, F.-Y. Wang, C. Yuan, *Hypercontractivity for functional stochastic partial differential equations,* Comm. Elect. Probab. 20(2015), 1–15.

BGL [5] S. G. Bobkov, I. Gentil, M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations,* J. Math. Pures Appl. 80(2001), 669–696.

BRW01 [6] V. I. Bogachev, M. Röckner, F.-Y. Wang, *Elliptic equations for invariant measures on finite and infinite dimensional manifolds,* J. Math. Pure Appl. 80(2001), 177–221.

CMV [7] J. A. Carrillo, R. J. McCann, C. Villani, *Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates,* Rev. Mat. Iberoam. 19(2003), 971–1018.

CW97 [8] M. F. Chen, F.-Y. Wang, *Estimates of logarithmic Sobolev constant: an improvement of Bakry-Emery criterion,* J. Funct. Anal. 144(1997), 287–300.

DS [9] J.-D. Deuschel, D. W. Stroock, *Hypercontractivity and spectral gap of symmetric diffusion with applications to the stochastic Ising models,* J. Funct. Anal. 92(1990), 30–48.

GW19 [10] M. Grothaus, F.-Y. Wang, *Weak poincaré inequalities for convergence rate of degenerate diffusion processes,* Ann. Probab. 47(2019), 2930–2952.

GLW [11] A. Guillin, W. Liu, L. Wu, *Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems,* arXiv:1909.07051v1

GW12 [12] A. Guillin, F.-Y. Wang, *Degenerate Fokker-Planck equations: Bismut formula, gradient estimate and Harnack inequality,* J. Diff. Equat. 253(2012), 20–40.

HW20 [13] X. Huang, F.-Y. Wang, *McKean-Vlasov SDEs with Drifts Discontinuous under Wasserstein Distance,* arXiv:2002.06877, to appear in Discrete Contin. Dyn. Syst. A.

Miclo [14] L. Miclo, *On hyperboundedness and spectrum of Markov operators,* Invent. math. 200(2015), 311–343.

OV [15] F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality,* J. Funct. Anal. 173(2000), 361–400.

RW03 [16] M. Röckner, F.-Y. Wang, *Supercontractivity and ultracontractivity for (non-symmetric) diffusion semigroups on manifolds,* Forum Math. 15 (2003), 893–921.

SN [17] A.-S. Sznitman, *Topics in propagations of chaos,* Lecture notes in Math. Vol. 1464, pp. 165–251, Springer, Berlin, 1991.

TAL [18] M. Talagrand, *Transportation cost for Gaussian and other product measures,* Geom. Funct. Anal. 6(1996), 587–600.

W04 [19] F.-Y. Wang, *Probability distance inequalities on Riemannian manifolds and path spaces,* J. Funct. Anal. 206(2004), 167–190.

W10 [20] F.-Y. Wang, *Harnack inequalities on manifolds with boundary and applications,* J. Math. Pures Appl. 94(2010), 304–321.

W11 [21] F.-Y. Wang, *Harnack inequality for SDE with multiplicative noise and extension to Neumann semigroup on non-convex manifolds,* Ann. Probab. 39(2011), 1449–1467.

W13 [22] F.-Y. Wang, *Harnack Inequalities and Applications for Stochastic Partial Differential Equations*, Springer, 2013, Berlin.

W14 [23] F.-Y. Wang, *Criteria on spectral gap of Markov operators,* J. Funct. Anal. 266(2014), 2137–2152.

W17 [24] F.-Y. Wang, *Hypercontractivity and applications for stochastic Hamiltonian systems,* J. Funct. Anal. 272(2017), 5360–5383.

W18 [25] F.-Y. Wang, *Distribution dependent SDEs for Landau type equations,* Stoch. Proc. Appl. 128(2018), 595–621.

WZ13 [26] F.-Y. Wang, X. Zhang, *Derivative formula and applications for degenerate diffusion semigroups,* J. Math. Pures Appl. 99(2013), 726–740.

Zhao20 [27] G. Zhao, *On Distribution depend SDEs with singular drifts,* arXiv:2003.04829.