Entropy Estimate for Degenerate SDEs with Applications to Nonlinear Kinetic Fokker-Planck Equations*

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

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

[b)] Mathematical Institute, Oxford University, Oxford, OX2 6GG, United Kingdom

[c)] Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Hong Kong, China

zhongmin.qian@maths.ox.ac.uk, panparen@gmail.com, wangfy@tju.edu.cn

April 21, 2024

Abstract

The relative entropy for two different degenerate diffusion processes is estimated by using the Wasserstein distance of initial distributions and the difference between coefficients. As applications, the entropy-cost inequality and exponential ergodicity in entropy are derived for distribution dependent stochastic Hamiltonian systems associated with nonlinear kinetic Fokker-Planck equations.

AMS subject Classification: 60J60, 60H30.
Keywords: Entropy estimate, degenerate diffusion process, stochastic Hamiltonian system, nonlinear kinetic Fokker-Planck equation.

1 Introduction

To characterize the stability of stochastic systems under perturbations, a natural way is to estimate the difference of distributions for two different processes, see [14] for a comparison theorem on transition densities (i.e. heat kernels) of diffusions with different drifts.

Recently, by using the entropy inequality established by Bogachev, Röckner and Shaposhnikov [1] for diffusion processes, and by developing a bi-coupling argument, the entropy and

*Supported in part by the National Key R&D Program of China (No. 2022YFA1006000, 2020YFA0712900) and NNSFC (11921001).

probability distances have been estimated in [17, 10] for different non-degenerate SDEs with distribution dependent noise. In this paper, we aim to establish entropy inequality for degenerate diffusion processes. As applications, we establish a log-Harnack inequality and study the exponential ergodicity in entropy for stochastic Hamiltonian systems with distribution dependent noise.

Let us start with a simple stochastic Hamiltonian system whose Hamiltonian function is given by

$$H(x) := V_1(x^{(1)}) + V_2(x^{(2)}) \ \text{ for } x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^d \times \mathbb{R}^d,$$

where $V_i \in C^2(\mathbb{R}^d)$ with $\|\nabla^2 V_i\|_\infty < \infty, i = 1, 2$. Then $X_t = (X_t^{(1)}, X_t^{(2)})$, the speed $X_t^{(1)}$ and the location $X_t^{(2)}$ of the stochastic particle, solves the following degenerate stochastic differential equation (SDE) on $\mathbb{R}^d \times \mathbb{R}^d$:

$$\begin{cases} \mathrm{d}X_t^{(1)} = \nabla V_2(X_t^{(2)})\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = \sqrt{2}\,\mathrm{d}W_t - \big(\nabla V_1(X_t^{(1)}) + \nabla V_2(X_t^{(2)})\big)\mathrm{d}t, \end{cases} \tag{1.1}$$

where W_t is the d-dimensional Brownian motion on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$. It is well known that the distribution density function of X_t solves the associated kinetic Fokker-Planck equation.

When for each $i = 1, 2$, $\mu^{(i)}(\mathrm{d}x^{(i)}) := \mathrm{e}^{-V_i(x^{(i)})}\mathrm{d}x^{(i)}$ is a probability measure on $\mathbb{R}^d$, SDE (1.1) has a unique invariant probability measure

$$\bar{\mu}(\mathrm{d}x) := \mu^{(1)}(\mathrm{d}x^{(1)})\mu^{(2)}(\mathrm{d}x^{(2)}), \ \text{ for } x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^d \times \mathbb{R}^d.$$

According to Villani [19], suppose that $\mu^{(i)}$ satisfies the Poincaré inequality

$$\mu^{(i)}(f^2) \leq \mu^{(i)}(f)^2 + C\mu^{(i)}(|\nabla f|^2), \ \ \forall f \in C_b^1(\mathbb{R}^d), i = 1, 2,$$

for some constant $C > 0$, where and in the sequel $\mu(f) := \int f \mathrm{d}\mu$ for a measure μ and a function f if the integral exists. Then the Markov semigroup P_t associated with (1.1) converges exponentially to $\bar{\mu}$ in $H^1(\bar{\mu})$, i.e. for some constants $c, \lambda > 0$,

$$\bar{\mu}\big(|P_t f - \bar{\mu}(f)|^2 + |\nabla P_t f|^2\big) \leq c\mathrm{e}^{-\lambda t}\bar{\mu}\big(|f - \bar{\mu}(f)|^2 + |\nabla f|^2\big)$$

for any $t \geq 0$ and $f \in C_b^1(\mathbb{R}^d)$. This property, known as "hypocoercivity " due to Villani [19], has been explored further by various authors in a series of papers for the exponential convergence of P_t in $L^2(\mu)$, such as [2] by Camrud, Herzog, Stoltz and Gordina, as well as [6] by Grothaus and Stilgenbauer, based on an abstract analytic framework built up by Dolbeaut, Mouhot and Schmeiser [4], see also the recent work [5] for the study of singular models. In case the Poincaré inequality fails, slower convergence rates are presented in [7, 11] using the weak Poincaré inequality developed by Röckner and the third named author [18].

On the other hand, the study of the exponential ergodicity in the relative entropy arising from information theory, which is stronger than that in L^2 (see [20]), becomes an important topic. Recall that if μ and ν are two probability measures, then the relative entropy of μ with respect to ν is defined by

$$\mathrm{Ent}(\mu|\nu) := \begin{cases} \mu\big(\log \frac{\mathrm{d}\mu}{\mathrm{d}\nu}\big), & \text{if } \mu \text{ is absolutely continuous w.r.t. } \nu, \\ \infty, & \text{otherwise.} \end{cases}$$

By Young's inequality, see for instance [**?**, Lemma 2.4], for any positive measurable function f such that $\nu(f)=1$, we have

$$\mu(\log f)=\nu\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log f\Big)\le \nu\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\log\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\Big)+\log\nu(f)=\mathrm{Ent}(\mu|\nu),$$

and the equality holds for $f=\frac{\mathrm{d}\mu}{\mathrm{d}\nu}$. Thus,

$$\mathrm{Ent}(\mu|\nu)=\sup_{f>0,\nu(f)=1}\mu(\log f)=\sup_{f>0,\nu(f)<\infty}\big[\mu(\log f)-\log\nu(f)\big], \tag{1.2}$$

since the right hand side is infinite if μ is not absolutely continuous with respect to ν.

By establishing a log-Harnack inequality, the exponential ergodicity in entropy has been been derived in [20] for stochastic Hamiltonian systems for linear ∇V_2, and has been further extended in [16, 9] to the case with distribution dependent drift. However, the log-Harnack inequality and the exponential ergodicity in entropy are still unknown for stochastic Hamiltonian systems with nonlinear ∇V_2.

To formulate distribution dependent SDEs, we introduce the Wasserstein space $\mathscr{P}_2(\mathbb{R}^d)$ for probability measures on $\mathbb{R}^d$ having finite second moment. It is a Polish space under the Wasserstein distance

$$\mathbb{W}_2(\mu,\nu):=\inf_{\pi\in\mathscr{C}(\mu,\nu)}\bigg(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\pi(\mathrm{d}x,\mathrm{d}y)\bigg)^{\frac{1}{2}},$$

where $\mathscr{C}(\mu,\nu)$ denotes the set of all couplings for μ and ν. Let $\mathscr{L}_\xi$ denote the distribution of the random variable ξ.

To illustrate our general results, we consider below the distribution dependent stochastic Hamiltonian system for $X_t:=(X_t^{(1)},X_t^{(2)})\in\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$:

$$\begin{cases}\mathrm{d}X_t^{(1)}=\big\{BX_t^{(2)}+b(X_t)\big\}\mathrm{d}t,\\ \mathrm{d}X_t^{(2)}=\sigma(\mathscr{L}_{X_t})\mathrm{d}W_t+Z(X_t^{(2)},\mathscr{L}_{X_t})\mathrm{d}t,\quad t\ge 0,\end{cases} \tag{1.3}$$

where B is a $d_1\times d_2$-matrix such that BB^* is invertible (i.e. Rank$(B)=d_1$), $b\in C_b^2(\mathbb{R}^{d_1+d_2})$ such that

$$\big\langle(\nabla^{(2)}b)B^*v,v\big\rangle\ge-\delta|B^*v|^2,\quad v\in\mathbb{R}^{d_1}$$

holds for some constant $\delta\in(0,1)$, where $\nabla^{(2)}$ is the gradient in $x^{(2)}\in\mathbb{R}^{d_2}$, and

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

are Lipschitz continuous. According to [21, Theorem 2.1], (1.3) is well-posed for distributions in $\mathscr{P}_2(\mathbb{R}^{d_1+d_2})$, i.e. for any $\mathscr{F}_0$-measurable initial value X_0 with $\mathscr{L}_{X_0}\in\mathscr{P}_2(\mathbb{R}^{d_1+d_2})$, (respectively, any initial distribution $\mu\in\mathscr{P}_2(\mathbb{R}^{d_1+d_2})$), the SDE has a unique strong (respectively, weak) solution with $\mathscr{L}_{X_t}\in\mathscr{P}_2(\mathbb{R}^{d_1+d_2})$ continuous in $t\ge0$. Let $P_t^*\mu:=\mathscr{L}_{X_t}$ where X_t is the solution of (1.3) with initial distribution $\mu\in\mathscr{P}_2$. If $\nabla Z(\cdot,\mu)$ is bounded and Lipschitz continuous uniformly in μ, then the following assertions are implied by Theorem 4.1.

- By (4.4) for $k = 0$, there exists a constant $c > 0$ such that

$$\mathrm{Ent}(P_t^*\mu|P_t^*\nu) \le \frac{c}{t^3}\mathbb{W}_2(\mu,\nu)^2, \quad t \in (0,1];\ \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}).$$

- If P_t^* is exponentially ergodic in $\mathbb{W}_2$, i.e. P_t^* has a unique invariant probability measure $\bar\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$ and there exist two positive constants c_1 and λ such that

$$\mathbb{W}_2(P_t^*\mu,\bar\mu)^2 \le c_1 \mathrm{e}^{-\lambda t}\mathbb{W}_2(\mu,\bar\mu)^2 \tag{1.4}$$

  holds for any $t \ge 0$ and $\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$, then the exponential ergodicity in entropy holds:

$$\mathrm{Ent}(P_t^*\mu|\bar\mu) \le cc_1 \mathrm{e}^{-\lambda(t-1)}\mathbb{W}_2(\mu,\bar\mu)^2$$

  holds for any $t \ge 0$ and $\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$. See Corollary 4.2 and Example 4.1 below for some concrete models satisfying (1.4).

The remainder of the paper is organized as follows. We establish an entropy inequality in Section 2 for some SDEs which applies also to the degenerate case, then apply the inequality to stochastic Hamiltonian systems and the distribution dependent model in Sections 3 and 4 respectively.

2 Entropy estimate between diffusion processes

Let $d, m \in \mathbb{N}, T \in (0,\infty)$, and $(W_t)_{t\in[0,T]}$ be an m-dimensional Brownian motion on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,T]}, \mathbb{P})$. Consider the following SDEs on $\mathbb{R}^d$:

$$\mathrm{d}X_t^{\langle i\rangle} = Z_i(t, X_t^{\langle i\rangle})\mathrm{d}t + \sigma_i(t, X_t^{\langle i\rangle})\mathrm{d}W_t \quad \text{for } t \in [0,T], \tag{2.1}$$

where

$$Z_i : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d \ \text{ and } \sigma_i : [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\otimes m}$$

are nice enough measurable maps such that the SDE is well-posed for $i = 1, 2$. Let $(P_{s,t}^{\langle i\rangle})_{0\le s\le t\le T}$ be the corresponding Markov semigroups, i.e.

$$P_{s,t}^{\langle i\rangle} f(x) := \mathbb{E}[f(X_{s,t}^{i,x})] \ \text{ for } f \in \mathscr{B}_b(\mathbb{R}^d) \text{ and } x \in \mathbb{R}^d,$$

where $(X_{s,t}^{i,x})_{t\in[s,T]}$ solves (2.1) for $t \in [s,T]$ with $X_{s,s}^{i,x} = x$. The corresponding generators are given by

$$L_t^{\langle i\rangle} := \mathrm{tr}\big\{a_i(t,\cdot)\nabla^2\big\} + Z_i(t,\cdot)\cdot\nabla \quad \text{for } t \in [0,T],$$

where $a_i := \frac{1}{2}\sigma_i\sigma_i^*$ which may be degenerate. If $v : [0,T] \mapsto \mathbb{R}^d$ is a path, then

$$\|v\|_{a_2}(t) := \sup_{x\in\mathbb{R}^d} \inf\big\{|w| : w \in \mathbb{R}^d, a_2(t,x)^{\frac{1}{2}}w = v(t)\big\} \quad \text{for } t \in [0,T],$$

where the convention that $\inf\emptyset = \infty$ is applied.

Let $\mathscr{P}(\mathbb{R}^d)$ denote the space of all probability measures on $\mathbb{R}^d$. For a given $\nu \in \mathscr{P}(\mathbb{R}^d)$, $X_t^{i,\nu}$ denotes the solution to (2.1) with $\mathscr{L}_{X_0^{i,\nu}} = \nu$, where and in the sequel, $\mathscr{L}_\xi$ stands for the law of a random variable ξ. Denote

$$P_t^{i,\nu} = \mathscr{L}_{X_t^{i,\nu}} \quad \text{for } t \in [0,T],\ \nu \in \mathscr{P}(\mathbb{R}^d) \text{ and } i = 1,2.$$

We shall make the following assumptions.

(A_1) For any $0 \le s \le t \le T$, $P_{s,t}^{\langle 2\rangle} C_b^2(\mathbb{R}^d) \subset C_b^2(\mathbb{R}^d)$ so that the Kolmogorov backward equation holds for any $f \in C_b^2(\mathbb{R}^d)$:

$$\partial_s P_{s,t}^{\langle 2\rangle} f = -L_s^{\langle 2\rangle} P_{s,t}^{\langle 2\rangle} f \quad \text{for } s \in [0,t] \text{ and } t \in (0,T].$$

(A_2) For any $t \in (0,T]$, $(a_1 - a_2)(t,\cdot)$ is differentiable on $\mathbb{R}^d$, and there exists a measurable function $H_{a_1-a_2}{}^{1,\nu} : (0,T] \mapsto (0,\infty)$ such that

$$\begin{aligned}&\big|\mathbb{E}\big[\mathrm{div}\{(a_1-a_2)(t,\cdot)\nabla f\}(X_t^{1,\nu})\big]\big|\\ &\le H_{a_1-a_2}^{1,\nu}(t)\big(\mathbb{E}[|a_2(t,\cdot)^{\frac12}\nabla f|^2(X_t^{1,\nu})]\big)^{\frac12}\end{aligned}$$

holds for any $t \in (0,T]$ and $f \in C_b^2(\mathbb{R}^d)$.

We remark that condition (A_1) is satisfied when the coefficients have bounded first and second order derivatives. For the non-degenerate case, it is satisfied for a class of Hölder continuous σ_2 and b_2, see for instance [12] and references within. According to [1], condition (A_2) is satisfied if a_2 is invertible and $X_t^{1,\nu}$ has a distribution density $\rho_t^{1,\nu}$ such that $\log \rho_t^{1,\nu}$ is in a Sobolev space. In this case, inequality (2.2) in the following theorem reduces to [1, Theorem 1.1]. In the next section, we shall verify these conditions for some important examples of degenerate SDEs.

We are now in a position to state and prove the main result.

Theorem 2.1. *Assume that (A_1) and (A_2) are satisfied. Then*

$$\mathrm{Ent}(P_t^{1,\nu}|P_t^{2,\nu}) \le \frac14 \int_0^t \Big\{\|Z_1 - Z_2 - \mathrm{div}(a_1-a_2)\|_{a_2}(s) + H_{a_1-a_2}^{1,\nu}(s)\Big\}^2 \mathrm{d}s \tag{2.2}$$

for any $t \in (0,T]$.

Proof. Let $X_t^{i,\nu}$ solve (2.1) with initial distribution ν, and let $X_0^{1,\nu} = X_0^{2,\nu}$. Let $C_{b,+}^2(\mathbb{R}^d)$ denote the space of all functions $f \in C_b^2(\mathbb{R}^d)$ such that $\inf f > 0$. By (1.2) and an approximation argument, we have

$$\begin{aligned}&\mathrm{Ent}(P_t^{1,\nu}|P_t^{2,\nu}) = \sup_{f\in C_{b,+}^2(\mathbb{R}^d)} I_t(f),\\ &I_t(f) := \mathbb{E}\log f(X_t^{1,\nu}) - \log \mathbb{E} f(X_t^{2,\nu}).\end{aligned} \tag{2.3}$$

Noting that $(X_t^{2,x} : x \in \mathbb{R}^d)_{t\in[0,T]}$ is a (time inhomogenous) Markov process, for any $f \in C_{b,+}^2(\mathbb{R}^d)$, we have

$$\mathbb{E}[f(X_t^{2,\nu})] = \int_{\mathbb{R}^d} (P_{0,t}^{\langle 2\rangle} f)\mathrm{d}\nu = \mathbb{E}[P_{0,t}^{\langle 2\rangle} f(X_0^{2,\nu})]. \tag{2.4}$$

So, by Jensen's inequality, we obtain

$$\begin{aligned} I_t(f) &= \mathbb{E}\log f(X_t^{1,\nu}) - \log \mathbb{E}(P_{0,t}^{\langle 2\rangle} f)(X_0^{2,\nu}) \\ &\le \mathbb{E}\log f(X_t^{1,\nu}) - \mathbb{E}\log(P_{0,t}^{\langle 2\rangle} f)(X_0^{2,\nu}) \\ &= \int_0^t \Big[\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\log(P_{s,t}^{\langle 2\rangle} f)(X_s^{1,\nu})\Big]\mathrm{d}s \end{aligned} \tag{2.5}$$

for every $t \in (0,T]$. By (A_1) and using Itô's formula for $X_s^{1,\nu}$, we derive that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\big(\log(P_{s,t}^{\langle 2\rangle} f)(X_s^{1,\nu})\big) = \mathbb{E}\bigg[\Big(L_s^{\langle 1\rangle}\log(P_{s,t}^{\langle 2\rangle} f) - \frac{L_s^{\langle 2\rangle}P_{s,t}^{\langle 2\rangle} f}{P_{s,t}^{\langle 2\rangle} f}\Big)(X_s^{1,\nu})\bigg] \\ &= \mathbb{E}\Big[(L_s^{\langle 1\rangle} - L_s^{\langle 2\rangle})\log(P_{s,t}^{\langle 2\rangle} f)(X_s^{1,\nu}) - \big|\{a_2(s,\cdot)^{\frac{1}{2}}\nabla\log P_{s,t}^{\langle 2\rangle} f\}\big|^2(X_s^{1,\nu})\Big] \\ &= \mathbb{E}\Big[\mathrm{div}\big\{(a_1-a_2)(s,\cdot)\nabla\log P_{s,t}^{\langle 2\rangle} f\big\}(X_s^{1,\nu}) - \big|\{a_2(s,\cdot)^{\frac{1}{2}}\nabla\log P_{s,t}^{\langle 2\rangle} f\big|^2(X_s^{1,\nu})\Big] \\ &\quad + \mathbb{E}\Big[\big\langle\{Z_1 - Z_2 - \mathrm{div}(a_1-a_2)\}(s,\cdot), \nabla\log P_{s,t}^{\langle 2\rangle} f\big\rangle(X_s^{1,\nu})\Big]. \end{aligned}$$

Combining this with (A_2) gives that

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}s}\mathbb{E}\big(\log(P_{s,t}^{\langle 2\rangle} f)(X_s^{1,\nu})\big) \\ &\le \Big[H_{a_1-a_2}^{1,\nu}(s) + \|Z_1 - Z_2 - \mathrm{div}(a_1-a_2)\|_{a_2}(s)\Big]\big(\mathbb{E}|a_2(s,\cdot)^{\frac{1}{2}}\nabla\log P_{s,t}^{\langle 2\rangle} f|^2(X_s^{1,\nu})\big)^{\frac{1}{2}} \\ &\quad - \mathbb{E}\Big[\big|a_2(s,\cdot)^{\frac{1}{2}}\nabla\log P_{s,t}^{\langle 2\rangle} f\big|^2(X_s^{1,\nu})\Big] \\ &\le \frac{1}{4}\Big[H_{a_1-a_2}^{1,\nu}(s) + \|Z_1 - Z_2 - \mathrm{div}(a_1-a_2)\|_{a_2}(s)\Big]^2 \end{aligned}$$

for every $s \in (0,t]$, which, together with (2.3) and (3.27), implies the desired estimate (2.2). □

As explained in [17] that $|H_{a_1-a_2}^{1,\nu}(s)|^2$ is normally singular for small s, such that the upper bound in (2.2) becomes infinite. To derive a finite upper bound of the relative entropy, we make use of the bi-coupling argument developed in [17], which leads to the following consequence where different initial distributions are also allowed.

Corollary 2.2. *Assume that (A_1) and (A_2) are satisfied, $H_{a_1-a_2}^{1,x}(s) := H_{a_1-a_2}^{1,\delta_x}(s)$ is measurable in $x \in \mathbb{R}^d$ such that*

$$H_{a_1-a_2}^{1,\nu} = \int_{\mathbb{R}^d} H_{a_1-a_2}^{1,x}(s)\nu(\mathrm{d}x).$$

Suppose that there exist a constant $p \in (1,\infty)$ and a decreasing function $\eta : (0,T] \mapsto (0,\infty)$ such that

$$|P_{s,t}^{\langle 2\rangle} f(x)|^p \le \big(P_{s,t}^{\langle 2\rangle}|f|^p(y)\big)\mathrm{e}^{\eta(t-s)|x-y|^2} \tag{2.6}$$

for any $0 \le s < t \le T$ and $f \in \mathscr{B}_b(\mathbb{R}^d)$. Then there exists a constant $c > 0$ such that

$$\begin{aligned}\mathrm{Ent}(P_t^{1,\mu}|P_t^{2,\nu}) \le \inf_{\pi\in\mathscr{C}(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \bigg(&\frac{p}{4}\int_{t_0}^t \Big\{\|b_1-b_2-\mathrm{div}(a_1-a_2)\|_{a_2}(s) + H_{a_1-a_2}^{1,x_1}(s)\Big\}^2 \mathrm{d}s \\ &+ (p-1)\log\mathbb{E}\Big\{\exp\big[c\eta(t-t_0)\big|X_{t_0}^{1,x_1} - X_{t_0}^{2,x_2}\big|^2\big]\Big\}\bigg)\pi(\mathrm{d}x_1,\mathrm{d}x_2)\end{aligned}$$

for any $0 < t_0 < t \le T$ and $x, y \in \mathbb{R}^d$.

Proof. For simplicity, denote $P_t^{i,x} = P_t^{i,\delta_x}$ where $i = 1, 2$, $x \in \mathbb{R}^d$, and δ_x is the Dirac measure at x. Let $X_t(x_1)$ be the diffusion process starting from the initial value x_1 with the infinitesimal generator given by

$$L_t := 1_{[0,t_0]}(t)L_t^{\langle 1\rangle} + 1_{(t_0,t]}(t)L_t^{\langle 2\rangle}.$$

Let $P_t^{\langle t_0\rangle x_1} = \mathscr{L}_{X_t(x_1)}$. By using (2.2) with $\nu = \delta_{x_1}$ and $P_t^{\langle t_0\rangle x_1}$ in place of P_t^{2,x_1}, and combining with [17, (2.4) and (2.9)], we deduce that

$$\begin{aligned}\mathrm{Ent}(P_t^{1,x_1}|P_t^{2,x_2}) \le &\frac{p}{4}\int_{t_0}^t \Big\{\|b_1-b_2-\mathrm{div}(a_1-a_2)\|_{a_2}(s) + H_s^{1,x_1}(a_1-a_2)\Big\}^2 \mathrm{d}s \\ &+ (p-1)\log\mathbb{E}\Big\{\exp\Big[c\eta(t-t_0)\big|X_{t_0}^{1,x_1} - X_{t_0}^{2,x_2}\big|^2\Big]\Big\}.\end{aligned} \tag{2.7}$$

On the other hand, if $\pi \in \mathscr{C}(\mu,\nu)$, then by using (2.3), (2.4) and Jensen's inequality, we obtain

$$\begin{aligned}\mathrm{Ent}(P_t^{1,\mu}|P_t^{2,\nu}) &= \sup_{f\in C_{b,+}^2(\mathbb{R}^d)} \big\{\mathbb{E}\log f(X_t^{1,\mu}) - \log\mathbb{E}f(X_t^{2,\nu})\big\} \\ &= \sup_{f\in C_{b,+}^2(\mathbb{R}^d)} \bigg\{\int_{\mathbb{R}^d} P_t^{\langle 1\rangle}(\log f)(x_1)\mu(\mathrm{d}x_1) - \log\int_{\mathbb{R}^d} P_t^{\langle 2\rangle} f(x_2)\nu(\mathrm{d}x_2)\bigg\} \\ &\le \sup_{f\in C_{b,+}^2(\mathbb{R}^d)} \bigg\{\int_{\mathbb{R}^d} P_t^{\langle 1\rangle}(\log f)(x_1)\mu(\mathrm{d}x_1) - \int_{\mathbb{R}^d} \log P_t^{\langle 2\rangle} f(x_2)\nu(\mathrm{d}x_2)\bigg\} \\ &= \sup_{f\in C_{b,+}^2(\mathbb{R}^d)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \big\{P_t^{\langle 1\rangle}(\log f)(x_1) - \log P_t^{\langle 2\rangle} f(x_2)\big\}\pi(\mathrm{d}x_1,\mathrm{d}x_2) \\ &\le \int_{\mathbb{R}^d\times\mathbb{R}^d} \sup_{f\in C_{b,+}^2(\mathbb{R}^d)} \big\{P_t^{\langle 1\rangle}(\log f)(x_1) - \log P_t^{\langle 2\rangle} f(x_2)\big\}\pi(\mathrm{d}x_1,\mathrm{d}x_2) \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{Ent}(P_t^{1,x_1}|P_t^{2,x_2})\pi(\mathrm{d}x_1,\mathrm{d}x_2),\end{aligned}$$

which, together with (2.7), yields the desired estimate. □

3 Stochastic Hamilton system

3.1 A general result

Let $d_1, d_2 \in \mathbb{N}$. For any initial distribution $\nu \in \mathscr{P}(\mathbb{R}^{d_1+d_2})$, consider the following degenerate SDEs for $X_t^{i,\nu} = (X_t^{i(1),\nu}, X_t^{i(2),\nu}) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ $(i = 1, 2)$:

$$(3.1) \qquad \begin{cases} \mathrm{d}X_t^{i(1),\nu} = \tilde{b}(t, X_t^{i,\nu})\mathrm{d}t, \\ \mathrm{d}X_t^{i(2),\nu} = Z_i(t, X_t^{i,\nu})\mathrm{d}t + \sigma_i(t, X_t^{i,\nu})\mathrm{d}W_t, \quad \mathscr{L}_{X_0^{i,\nu}} = \nu, \quad \text{for } t \in [0, T], \end{cases}$$

where W_t is a d_2-dimensional Brownian motion on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in[0,T]}, \mathbb{P})$, and

$$\tilde{b} : [0, T] \times \mathbb{R}^{d_1+d_2} \to \mathbb{R}^{d_1}, \quad Z_i : [0, T] \times \mathbb{R}^{d_1+d_2} \to \mathbb{R}^{d_2}, \quad \sigma_i : [0, T] \times \mathbb{R}^{d_1+d_2} \to \mathbb{R}^{d_2 \otimes d_2}$$

are measurable.

If $\nu = \delta_x$ where $x \in \mathbb{R}^{d_1+d_2}$, then the solution is simply denoted by $X_t^{i,x} = (X_t^{i(1),x}, X_t^{i(2),x})$. Let $\nabla^{(i)}$ be the gradient in $x^{(i)} \in \mathbb{R}^{d_i}$ for $i = 1, 2$.

Let us introduce the following technical conditions.

(B_1) The coefficients $\sigma_i(t, x), Z_i(t, x)$ (for $i = 1, 2$) and $\tilde{b}(t, x)$ are locally bounded in $(t, x) \in [0, T] \times \mathbb{R}^{d_1+d_2}$ and twice differentiable in the space variable x. The matrix valued function $a_2 := \frac{1}{2}\sigma_2\sigma_2^*$ is invertible. There exists a constant $K > 0$ such that

$$\|\nabla^j Z_i(t, x)\| + \|\nabla^j \tilde{b}(t, x)\| + \|\nabla^j \sigma_i(t, x)\| \le K$$

for $(t, x) \in [0, T] \times \mathbb{R}^{d_1+d_2}$ and $j = 1, 2$.

(B_2) There exists a function $\xi^\nu \in C((0, T]; (0, \infty))$ such that

$$\big|\mathbb{E}[(\nabla_v^{(2)} f)(X_t^{1,\nu})]\big| \le \xi_t^\nu \big(\mathbb{E}[f(X_t^{1,\nu})^2]\big)^{\frac{1}{2}}$$

for $t \in (0, T]$, $v^{(2)} \in \mathbb{R}^{d_2}$ with $|v^{(2)}| = 1$ and $f \in C_b^1(\mathbb{R}^{d_1+d_2})$.

It is well known that condition (B_1) implies the well-posededness of (3.1) and that condition (A_1) is satisfied. Let $P_t^{i,\nu}$ be the distribution of $X_t^{i,\nu}$.

To state our next result we recall that for a vector valued function g on $[0, T] \times \mathbb{R}^{d_1+d_2}$

$$\|g\|_{t,\infty} := \sup_{z\in\mathbb{R}^{d_1+d_2}} |g(t, z)|$$

for $t \in [0, T]$.

Theorem 3.1. *Assume that conditions (B_1) and (B_2) are satisfied. Let $(e_j)_{1\le j\le d_2}$ be the canonical basis on $\mathbb{R}^{d_2}$.*

1) The following equality holds:

$$\mathrm{Ent}(P_t^{1,\nu}|P_t^{2,\nu}) \le \frac{1}{4}\int_0^t \Big[\big\|a_2^{-\frac{1}{2}}\{Z_1 - Z_2 - \mathrm{div}(a_1 - a_2)\}\big\|_{s,\infty} + \xi_s^\nu \sum_{j=1}^{d_2} \big\|a_2^{-\frac{1}{2}}(a_1 - a_2)e_j\big\|_{s,\infty}\Big]^2 \mathrm{d}s.$$

2) Suppose (2.6) *holds, then there exists a constant* $c > 0$ *such that*

$$\mathrm{Ent}(P_t^{1,\mu}|P_t^{2,\nu}) \le \inf_{\pi\in\mathscr{C}(\mu,\nu)} \int_{\mathbb{R}^{d_1+d_2}\times\mathbb{R}^{d_1+d_2}} \Big(pI_{t_0,t}^{x_2} + (p-1)\log\mathbb{E}\Big[\mathrm{e}^{c\eta(t-t_0)|X_{t_0}^{1,x_1}-X_{t_0}^{2,x_2}|^2}\Big]\Big)\pi(\mathrm{d}x_1,\mathrm{d}x_2)$$

for any $0 < t_0 < t \le T$ *and* $\mu,\nu \in \mathscr{P}(\mathbb{R}^{d_1+d_2})$, *where*

$$I_{t_0,t}^x := \frac{1}{4}\int_{t_0}^t \Big[\big\|a_2^{-\frac{1}{2}}\{Z_1 - Z_2 - \mathrm{div}(a_1-a_2)\}\big\|_{s,\infty} + \xi_s^x \sum_{j=1}^{d_2} \big\|a_2^{-\frac{1}{2}}(a_1-a_2)e_j\big\|_{s,\infty}\Big]^2 \mathrm{d}s$$

and $\xi_s^x := \xi_s^{\delta_x}$ *for every* $x \in \mathbb{R}^{d_1+d_2}$ *and* $s \in [t_0,t]$.

Proof. As explained in the proof of Corollary 2.2, we only need to prove the first estimate. Since (B_2) is satisfied, we have

$$\begin{aligned}
&\big|\mathbb{E}\big[\mathrm{div}\{\mathrm{diag}\{\mathbf{0}_{d_1\times d_1},(a_1-a_2)(t,\cdot)\}\nabla f\}(X_t^{1,\nu})\big]\big|\\
&= \Big|\sum_{j=1}^{d_2}\mathbb{E}\big[\partial_{y_j}\{(a_1-a_2)(t,\cdot)\nabla^{(2)}f\}_j\big](X_t^{1,\nu})\Big|\\
&\le \xi_t^\nu \sum_{j=1}^{d_2}\big(\mathbb{E}\{(a_1-a_2)(t,\cdot)\nabla^{(2)}f\}_j(X_t^{1,\nu})^2\big)^{\frac{1}{2}}\\
&= \xi_t^\nu \sum_{j=1}^{d_2}\big(\mathbb{E}\langle a_2(t,\cdot)^{-\frac{1}{2}}(a_2-a_2)(t,\cdot)e_j, a_2(t,\cdot)^{\frac{1}{2}}\nabla^{(2)}f\rangle_{\mathbb{R}^{d_2}}(X_t^{1,\nu})^2\big)^{\frac{1}{2}}\\
&\le \xi_t^\nu \sum_{j=1}^{d} \||a_2^{-\frac{1}{2}}(a_1-a_2)e_j|\|_{t,\infty}\big(\mathbb{E}\big|a_2(t,\cdot)^{\frac{1}{2}}\nabla^{(2)}f\big|^2(X_t^{1,\nu})\big)^{\frac{1}{2}}.
\end{aligned}$$

Thus (A_2) is satisfied with

$$H_t^\nu(a_1-a_2) := \xi_t^\nu \sum_{j=1}^{d} \||a_2^{-\frac{1}{2}}(a_1-a_2)e_j|\|_{t,\infty}.$$

Since (B_1) implies (A_1), the desired estimate follows immediately from Theorem 2.1. □

3.2 A class of models

We next discuss a class of degenerate stochastic models for which condition (B_2) is satisfied and the dimension-free Harnack inequality (2.6) holds.

Consider the following SDE for $X_t^{i,\nu} = (X_t^{i(1),\nu}, X_t^{i(2),\nu}) \in \mathbb{R}^{d_1+d_2}$:

$$\begin{cases}\mathrm{d}X_t^{i(1),\nu} = \big\{AX_t^{i(1),\nu} + BX_t^{i(2),\nu} + b(X_t^{i,\nu})\big\}\mathrm{d}t,\\ \mathrm{d}X_t^{i(2),\nu} = \sigma_i(t)\mathrm{d}W_t + Z_i(t,X_t^{i,\nu})\mathrm{d}t, \quad \mathscr{L}_{X_0^{i,\nu}} = \nu \ \text{ for } i = 1,2,\end{cases} \tag{3.2}$$

where A, B, b, σ_i and Z_i satisfy the following assumption.

(B_3) 1) A is a $d_1 \times d_1$ matrix and B is a $d_1 \times d_2$ matrix, such that Kalman's condition

$$\text{Rank}\left[A^i B : 0 \le i \le k\right] = d_1 \tag{3.3}$$

holds for some $0 \le k \le d_1 - 1$.

2) $b \in C_b^1(\mathbb{R}^{d_1+d_2})$ with Lipschitz continuous ∇b, and there exists a constant $\delta \in (0,1)$ such that

$$\left\langle (\nabla^{(2)} b(x)) B^* v, v \right\rangle \ge -\delta |B^* v|^2, \quad v \in \mathbb{R}^{d_1}, x \in \mathbb{R}^{d_1+d_2}. \tag{3.4}$$

3) $\sigma_1(t)$ and $\sigma_2(t)$ are bounded, and $a_2(t) := \frac{1}{2}\sigma_2(t)\sigma_2(t)^*$ is invertible with bounded inverse.

4) $Z_i(t,x)$ (for $i = 1,2$) are locally bounded in $[0,T] \times \mathbb{R}^{d_1+d_2}$ and differentiable in x, such that

$$\sup_{t\in[0,T]} \left\{ \|\nabla Z_i(t,\cdot)\| + \frac{\|\nabla Z_i(t,x) - \nabla Z_i(t,y)\|}{|x-y|} \right\} \le K$$

holds for some constant $K > 0$.

We introduce ξ_t in two different cases:

$$\xi_t := \begin{cases} t^{-2k-\frac{1}{2}}, & \text{if } Z_1(t,x) = Z_1(t,x^{(2)}) \text{ is independent of } x^{(1)}, \\ t^{-2k-\frac{3}{2}}, & \text{otherwise.} \end{cases} \tag{3.5}$$

Corollary 3.2. *Assume that (B_3) is satisfied for either $k = 0$ or $k \ge 1$ but $b(x) = b(x^{(2)})$ only depends on $x^{(2)}$. Let $P_t^{i,\nu}$ be the distribution of $X_t^{i,\nu}$ solving (3.2). Then there exist constants $c > 0$ and $\varepsilon \in (0, \frac{1}{2}]$ such that for any $t \in (0,T]$ and $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$,*

$$\text{Ent}(P_t^{1,\nu}|P_t^{2,\mu}) \le \frac{c}{t^{4k+3}} \left(\mathbb{W}_2(\mu,\nu)^2 + \int_0^t \|Z_1 - Z_2\|_{s,\infty}^2 \right) + c \int_{\varepsilon(1\wedge t)^{4k+3}}^t \xi_s^2 \|a_1(s) - a_2(s)\|^2) \text{d}s.$$

Proof. Without loss of generality, we may and do assume that $\sigma_i = \sqrt{2a_i}$. Moreover, by a standard approximation argument, under (B_3) we may find a sequence $\{Z_i^{(n)}\}_{n\ge1}$ for each $i = 1,2$, such that

$$\sup_{n\ge1, k=1,2, t\in[0,T]} \|\nabla^k Z_i^{(n)}(t,\cdot)\| \le K,$$
$$\lim_{n\to\infty} \sup_{t\in[0,T]} \{\|(Z_i - Z_i^{(n)})(t,\cdot)\|_\infty + \|\nabla(Z_i - Z_i^{(n)})(t,\cdot)\|_\infty = 0.$$

Moreover, let $\{b^{(n)}\}_{n\ge1}$ be a bounded sequence in $C_b^2(\mathbb{R}^{d_1+d_2})$ such that $\|b^{(n)} - b\|_{C_b^1(\mathbb{R}^{d_1+b_2})} \to 0$ as $n \to \infty$. Let $P_t^{i,\nu,;n}$ be defined as $P_t^{i,\nu}$ for $(b^{(n)}, Z_i^{(n)})$ replacing (b, Z_i). It is well known that $P_t^{i,\nu;n} \to P_t^{i,\nu}$ weakly as $n \to \infty$, so that (2.3) implies that

$$\text{Ent}(P_t^{1,\nu}|P_t^{2,\mu}) \le \liminf_{n\to\infty} \text{Ent}(P_t^{1,\nu;n}|P_t^{2,\mu;n}).$$

Therefore, we may and do assume that $\|\nabla^k b\| + \|\nabla^k Z_i(t,\cdot)\|_\infty \le K$ holds for some constant $K>0$ and $i,k=1,2$, so that Theorem 3.1 applies.

(a) By (B_3), $\sigma_1 \ge 0, \sigma_2 \ge \lambda I_{d_2}$ for some constant $\lambda>0$, where I_{d_2} is the $d_2\times d_2$ identity matrix. So, according to the proof of [13, Lemma 3.3],

$$\|\sigma_1-\sigma_2\| = \bigg\|2\int_0^\infty \mathrm{e}^{-r\sigma_1}(a_1-a_2)\mathrm{e}^{-r\sigma_2}\mathrm{d}r\bigg\| \le \frac{2}{\lambda}\|a_1-a_2\|. \tag{3.6}$$

By Lemma 3.3 below, there exists a constant $c_1>0$ such that for any ν, condition (B_2) holds with

$$\xi_t^\nu = c_1\xi_t := \begin{cases} c_1 t^{-2k-\frac{1}{2}}, & \text{if } Z_1(t,x)=Z_1(t,x^{(2)}),\\ c_1 t^{-2k-\frac{3}{2}}, & \text{in general.}\end{cases} \tag{3.7}$$

Moreover, by Lemma 3.4 below, (2.6) holds for the following $\eta(s), s\in(0,T)$:

$$\eta(s) = c(p)s^{-4k-3},\quad s\in(0,T]. \tag{3.8}$$

Combining these with Theorem 3.1, and noting that a_2^{-1} is bounded and $\mathrm{div}(a_1-a_2)=0$, we can find a constant $c_2>0$ such that for any $0<t_0<t\le T$,

$$\begin{aligned} &\mathrm{Ent}(P_t^{1,\mu}|P_t^{2,\nu}) \le c_2\int_{t_0}^t \Big(\big\|Z_1-Z_2\big\|_{s,\infty}^2 + |\xi_s|^2\big\|a_1(s)-a_2(s)\big\|^2\Big)\mathrm{d}s\\ &+ c_2 \inf_{\pi\in\mathscr{C}(\mu,\nu)}\int_{\mathbb{R}^{d_1+d_2}\times\mathbb{R}^{d_1+d_2}} \log\mathbb{E}\Big[\mathrm{e}^{c_2(t-t_0)^{-4k-3}|X_{t_0}^{1,x_1}-X_{t_0}^{2,x_2}|^2}\Big]\pi(\mathrm{d}x_1,\mathrm{d}x_2).\end{aligned} \tag{3.9}$$

It remains to estimate the exponential expectation in the last term.

(b) By (B_3) and (3.6), there exists a constant $c_3\ge 1$ such that

$$\mathrm{d}|X_s^{1,x_1}-X_s^{2,x_2}|^2 \le c_3\big(|X_s^{1,x_1}-X_s^{2,x_2}|^2 + \|Z_1-Z_2\|_{s,\infty}^2 + \|a_1(s)-a_2(s)\|^2\big)\mathrm{d}s + \mathrm{d}M_s,$$

where

$$\mathrm{d}M_s := 2\langle X_s^{1,x_1}-X_s^{2,x_2}, \{\sigma_1(s)-\sigma_2(s)\}\mathrm{d}W_s\rangle$$

and therefore the following differential inequality holds:

$$\mathrm{d}\langle M\rangle_s \le c_3|X_s^{1,x_1}-X_s^{2,x_2}|^2\mathrm{d}s. \tag{3.10}$$

It follows that

$$\begin{aligned} &|X_s^{1,x_1}-X_s^{2,x_2}|^2 \le \mathrm{e}^{c_3 s}|x_1-x_2|^2\\ &+\int_0^s \mathrm{e}^{c_3(s-r)}\big(\|Z_1-Z_2\|_{r,\infty}^2 + \|a_1(r)-a_2(r)\|^2\big)\mathrm{d}r + \int_0^s \mathrm{e}^{c_3(s-r)}\mathrm{d}M_r.\end{aligned} \tag{3.11}$$

Let

$$\tau_n := \inf\big\{s\in[0,T]: |X_s^{1,x_1}-X_s^{2,x_2}|\ge n\big\},\quad \text{for } n=1,2,\cdots$$

with the convention that $\inf\emptyset := T$. Then $\tau_n\to T$ as $n\to\infty$. Let

$$\lambda := c_3(t-t_0)^{-4k-3},\quad c_4 := \mathrm{e}^{c_3 T}.$$

By (3.11) and the fact that

$$\mathbb{E}[\mathrm{e}^{\lambda \hat{N}_{t\wedge\tau_n}}] \le (\mathbb{E}\mathrm{e}^{2\lambda^2\langle \hat{N}\rangle_{t\wedge\tau_n}})^{\frac{1}{2}} \le (\mathbb{E}\mathrm{e}^{2\lambda^2 c_4^2\langle M\rangle_{t\wedge\tau_n}})^{\frac{1}{2}}, \quad \lambda \ge 0$$

holds for the continuous martingale

$$\hat{N}_t := \int_0^t \mathrm{e}^{c_3(s-r)}\mathrm{d}M_r, \quad t\ge 0,$$

we deduce that

$$\begin{aligned} &\mathbb{E}\big[\mathrm{e}^{\lambda|X^{1,x_1}_{s\wedge\tau_n}-X^{2,x_2}_{s\wedge\tau_n}|^2}\big] \\ &\le \mathrm{e}^{c_4\lambda|x_1-x_2|^2+c_4\lambda\int_0^s\left(\|Z_1-Z_2\|^2_{r,\infty}+\|a_1(r)-a_2(r)\|^2\right)\mathrm{d}r}\Big(\mathbb{E}\big[\mathrm{e}^{2\lambda^2c_4^2\langle M\rangle_{s\wedge\tau_n}}\big]\Big)^{\frac{1}{2}}. \end{aligned} \tag{3.12}$$

While by (3.10) and Jensen's inequality,

$$\begin{aligned} \mathbb{E}\big[\mathrm{e}^{2\lambda^2c_4^2\langle M\rangle_{s\wedge\tau_n}}\big] &\le \mathbb{E}\Big[\mathrm{e}^{2\lambda^2c_4^2c_3^2\int_0^s|X^{1,x_1}_{r\wedge\tau_n}-X^{2,x_2}_{r\wedge\tau_n}|^2\mathrm{d}r}\Big] \\ &\le \frac{1}{s}\int_0^s \mathbb{E}\big[\mathrm{e}^{2\lambda^2c_4^2c_3^2 s|X^{1,x_1}_{r\wedge\tau_n}-X^{2,x_2}_{r\wedge\tau_n}|^2}\big]\mathrm{d}r \\ &\le \sup_{r\in[0,t_0]}\mathbb{E}\big[\mathrm{e}^{2\lambda^2c_4^2c_3^2 t_0|X^{1,x_1}_{r\wedge\tau_n}-X^{2,x_2}_{r\wedge\tau_n}|^2}\big] \end{aligned} \tag{3.13}$$

for $s\in[0,t_0]$. Choosing

$$t_0 = \frac{1}{2c_4^2c_3^3}\Big(\frac{1\wedge t}{2}\Big)^{4k+3} =: \varepsilon(1\wedge t)^{4k+3} \tag{3.14}$$

such that

$$2\lambda c_4^2c_3^2t_0 = 2c_4^2c_3^3(t-t_0)^{-4k-3}t_0 \le 1,$$

we therefore conclude from (3.12) and (3.13) that

$$\begin{aligned} &\sup_{s\in[0,t_0]}\mathbb{E}\big[\mathrm{e}^{\lambda|X^{1,x_1}_{s\wedge\tau_n}-X^{2,x_2}_{s\wedge\tau_n}|^2}\big] \\ &\le \mathrm{e}^{c_4\lambda|x_1-x_2|^2+c_4\lambda\int_0^{t_0}\left(\|Z_1-Z_2\|^2_{r,\infty}+\|a_1(r)-a_2(r)\|^2\right)\mathrm{d}r}\Big(\sup_{s\in[0,t_0]}\mathbb{E}\big[\mathrm{e}^{\lambda|X^{1,x_1}_{s\wedge\tau_n}-X^{2,x_2}_{s\wedge\tau_n}|^2}\big]\Big)^{\frac{1}{2}}. \end{aligned}$$

This together with the definition of λ and Fatou's lemma yields

$$\begin{aligned} \mathbb{E}\big[\mathrm{e}^{c_3(t-t_0)^{-4k-3}|X^{1,x_1}_{t_0}-X^{2,x_2}_{t_0}|^2}\big] &\le \liminf_{n\to\infty}\mathbb{E}\big[\mathrm{e}^{\lambda|X^{1,x_1}_{t_0\wedge\tau_n}-X^{2,x_2}_{t_0\wedge\tau_n}|^2}\big] \\ &\le \mathrm{e}^{2c_4\lambda|x_1-x_2|^2+2c_4\lambda\int_0^{t_0}\left(\|Z_1-Z_2\|^2_{r,\infty}+\|a_1(r)-a_2(r)\|^2\right)\mathrm{d}r}. \end{aligned}$$

Combining (3.9) with (3.14), we can therefore find a constant $c_5>0$ such that

$$\begin{aligned} \mathrm{Ent}(P_t^{1,\mu}|P_t^{2,\nu}) \le& c_2\int_{\varepsilon(1\wedge t)^{4k+3}}^t\Big(\big\|Z_1-Z_2\big\|^2_{s,\infty}+|\xi_s|^2\big\|a_1(s)-a_2(s)\big\|^2\Big)\mathrm{d}s \\ &+\frac{c_5}{t^{4k+3}}\Big(\mathbb{W}_2(\mu,\nu)^2+\int_0^{\varepsilon(t\wedge1)^{4k+3}}\big(\|Z_1-Z_2\|^2_{r,\infty}+\|a_1(r)-a_2(r)\|^2\big)\mathrm{d}r\Big). \end{aligned}$$

The desired estimate now follow from (3.7) immediately. □

3.3 Verify conditions (B_2) and (2.6)

Let us consider $X_t = (X_t^{(1)}, X_t^{(2)})$ taking values in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, which solves the SDE:

$$\begin{cases} \mathrm{d}X_t^{(1)} = \big\{AX_t^{(1)} + BX_t^{(2)} + b(X_t)\big\}\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = Z(t, X_t)\mathrm{d}t + \sigma(t)\mathrm{d}W_t \quad \text{for } t \in [0, T]. \end{cases} \tag{3.15}$$

We have the following result which ensures condition (B_2).

Lemma 3.3. *Let A, B, b and $(Z_i, \sigma_i) := (Z, \sigma)$ satisfy conditions in (B_3), but b is not necessarily bounded. Let ξ_t be in (3.5). Then for any $p > 1$ there exists a constant $c(p) > 0$ such that for any solution X_t of (3.15),*

$$\sup_{v \in \mathbb{R}^{d_1+d_2}, |v|=1} \big|\mathbb{E}\big[(\nabla_v f)(X_t)\big]\big| \le c(p) t^{-2k-\frac{3}{2}} \big(\mathbb{E}|f(X_t)|^p\big)^{\frac{1}{p}}, \quad t \in (0, T], f \in C_b^1(\mathbb{R}^{d_1+d_2}). \tag{3.16}$$

If $Z(t,x) = Z(t, x^{(2)})$ does not depend on $x^{(1)}$, then

$$\sup_{v \in \mathbb{R}^{d_2}, |v|=1} \big|\mathbb{E}\big[(\nabla_v^{(2)} f)(X_t)\big]\big| \le c(p) t^{-2k-\frac{1}{2}} \big(\mathbb{E}|f(X_t)|^p\big)^{\frac{1}{p}}, \quad t \in (0, T], f \in C_b^1(\mathbb{R}^{d_1+d_2}). \tag{3.17}$$

Proof. We will follow the line of [22, Remark 2.1] to establish the integration by parts formula

$$\mathbb{E}\big[(\nabla_v f)(X_t)\big] = \mathbb{E}\big[f(X_t) M_t\big]$$

for some random variable $M_t \in L^{\frac{p}{p-1}}(\mathbb{P})$. To this end, we first estimate $D_h X_t$ and $D_h(\nabla X_t)^{-1}$, where D_h is the Malliavin derivative along an adapted process $(h_s)_{s\in[0,t]}$ on $\mathbb{R}^d$ with

$$\mathbb{E}\int_0^t |h_s'|^2 \mathrm{d}s < \infty.$$

(a) For any $s \in [0, T)$, let $\{K(t,s)\}_{t\in[s,T]}$ solve the following random ordinary differential equation on $\mathbb{R}^{d_1 \otimes d_1}$:

$$\partial_t K_{t,s} = \big\{AX_t^{(1)} + \nabla^{(1)} b(t, X_t)\big\} K_{t,s}, \quad K_{s,s} = I_{d_1} \quad \text{for } t \in [s, T].$$

Since ∇b is bounded, $K_{t,s}$ is bounded and invertible satisfying

$$\|K_{t,s}\| \vee \|K_{t,s}^{-1}\| \le \mathrm{e}^{K(t-s)} \quad \text{for } 0 \le s \le t \le T \tag{3.18}$$

for some constant $K > 0$.

Let

$$Q_{t,s} := \int_0^s \frac{r(t-r)}{t^2} K_{t,r} BB^* K_{t,r}^* \mathrm{d}r \quad \text{for } 0 \le s \le t \le T.$$

By [22, Theorem 4.2(1)] for (t, s) replacing (T, t), when $k \ge 1$ and $b(x) = b(x^{(2)})$, conditions (3.3) and (3.4) imply that

$$Q_{t,s} \ge \frac{c_0}{t} s^{2(k+1)} I_{d_1} =: \xi_{t,s} I_{d_1} \quad \text{for } 0 < s \le t \le T \tag{3.19}$$

holds for some constant $c_0 > 0$. It is easy to see that this estimate also holds for $k = 0$ and bounded $\nabla b(x)$ since in this case BB^* is invertible.

Let $X_t(x) = (X_t^j(x))_{1\le j\le d_1+d_2}$ be the solution to (3.15) with $X_0(x) = x$. Since ∇b and ∇Z are bounded, we see that

$$\nabla X_t(x) := (\partial_{x_i} X_t^j(x))_{1\le i,j\le d_1+d_2}$$

exists and is invertible, and the inverse $(\nabla X_t(x))^{-1} = \big((\nabla X_t(x))_{ki}^{-1}\big)_{1\le k,i\le d_1+d_2}$ satisfies

$$\big\|\{\nabla X_t(x)\}^{-1}\big\| \le c_1 \quad \text{for } t\in[0,T] \tag{3.20}$$

for some constant $c_1 > 0$.

(b) Since ∇b and ∇Z are bounded, $(D_h X_s)_{s\in[0,t]}$ is the unique solution of the random ODE

$$\begin{cases} \partial_s\big\{D_h X_s^{(1)}\big\} = AD_h X_s^{(1)} + BD_h X_s^{(2)} + \nabla_{D_h X_s} b(X_s), \\ \partial_s\big\{D_h X_s^{(2)}\big\} = \nabla_{D_h X_s} Z(s, X_s) + \sigma(s)h_s', \quad D_h X_0 = 0 \quad \text{for } s\in[0,t], \end{cases}$$

and there exists a constant $c_2 > 0$ such that

$$|D_h X_s| \le c_2 \int_0^s |h_r'|\mathrm{d}r \quad \text{for } s\in[0,t]. \tag{3.21}$$

Similarly, since $\nabla^2 b$ and $\nabla^2 Z$ are also bounded, for any $v\in\mathbb{R}^{d_1+d_2}$, $(D_h\nabla_v X_s)_{s\in[0,t]}$ solve the equations

$$\begin{cases} \partial_s\big\{D_h\nabla_v X_s^{(1)}\big\} = AD_h\nabla_v X_s^{(1)} + BD_h\nabla_v X_s^{(2)} + \nabla_{D_h\nabla_v X_s} b(X_s) \\ \qquad\qquad\qquad\qquad + \Big\{\nabla^2 b(X_s)\Big\}\Big(D_h X_s, \nabla_v X_s\Big) \\ \partial_s\big\{D_h\nabla_v X_s^{(2)}\big\} = \nabla_{D_h\nabla_v X_s} Z(s, X_s) + \big\{\nabla^2 Z(s, X_s)\big\}\big(D_h X_s, \nabla_v X_s\big) \end{cases}$$

for $D_h\nabla_v X_0 = 0$ and $s\in[0,t]$. Moreover, there exists a constant $c_3 > 0$ such that

$$\sup_{v\in\mathbb{R}^{d_1+d_2}, |v|\le 1} \big\|D_h\nabla_v X_t\big\| \le c_3\int_0^t \mathrm{d}s\int_0^s |h_r'|\mathrm{d}r \le c_3 t\int_0^t |h_s'|\mathrm{d}s. \tag{3.22}$$

(c) For any fixed $t\in(0,T]$, we may construct h by means of [22, (1.8) and (1.11)] for t replacing T with the specific choice $\phi(s) := \frac{s(t-s)}{t}$ satisfying $\phi(0) = \phi(t) = 0$ as required therein.

For any $v = (v^{(1)}, v^{(2)})\in\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$, let

$$\begin{aligned} \alpha_{t,s}(v) &:= \frac{t-s}{t}v^{(2)} - \frac{s(t-s)}{t^2}B^*K_{t,s}^*Q_{t,t}^{-1}\int_0^t \frac{t-r}{t}K_{t,r}Bv^{(2)}\mathrm{d}r \\ &\quad - \frac{s(t-s)B^*K_{t,s}^*}{\xi_{t,s}^2\mathrm{d}s}\int_0^t \xi_{t,s}^2 Q_{t,s}^{-1}K_{t,s}v^{(1)}\mathrm{d}s, \\ g_{t,s}(v) &:= K_{s,0}v^{(1)} + \int_0^s K_{s,r}B\alpha_{t,s}(v)\mathrm{d}s, \\ h_{t,s}(v) &:= \int_0^s \sigma(r)^{-1}\big\{\nabla_{(g_{t,r}(v),\alpha_{t,r}(v))}b(r, X_r) - \partial_r\alpha_{t,r}\big\}\mathrm{d}r \quad \text{for } s\in[0,t]. \end{aligned}$$

Let $\{e_i\}_{1\le i\le d_1+d_2}$ be the canonical ONB on $\mathbb{R}^{d_1+d_2}$. According to [22, Remark 2.1], we have

$$\begin{aligned}&\mathbb{E}\big[(\nabla_{e_i} f)(X_t\big] = \mathbb{E}\big[f(X_t)M_t(e_i)\big],\\ &M_t(e_i) := \sum_{j=1}^{d_1+d_2}\Big\{\delta(h_{t,\cdot}(e_j))(\nabla X_t)^{-1}_{ji} - D_{h_{t,\cdot}(e_j)}(\nabla X_t)^{-1}_{ji}\Big\}\Big],\end{aligned} \tag{3.23}$$

where

$$\delta(h_{t,\cdot}(e_j)) := \int_0^t \big\langle \partial_s h_{t,s}(e_j), \mathrm{d}W_s\big\rangle$$

is the Malliavin divergence of $h_{t,\cdot}(e_j)$. Consequently

$$\big|\mathbb{E}(\nabla_{e_i} f)(X_t)\big]\big| \le \big(\mathbb{E}|f(X_t)|^p\big)^{\frac{1}{p}}\big(\mathbb{E}[|M_t(e_i)|^{\frac{p}{p-1}}]\big)^{\frac{p-1}{p}} \tag{3.24}$$

for $t\in(0,T]$ and $1\le i\le d_1+d_2$.

By (3.20) and (3.22), there is a constant $c_4>0$ such that

$$\big(\mathbb{E}[|M_t(e_i)|^{\frac{p}{p-1}}]\big)^{\frac{p-1}{p}} \le c_4\sum_{j=1}^{d_1+d_2} 1_{\{\|(\nabla X_t)^{-1}_{ji}\|_\infty>0\}}\bigg\{\mathbb{E}\bigg(\int_0^t |\partial_s h_{t,s}(e_j)|^2\mathrm{d}s\bigg)^{\frac{p}{2(p-1)}}\bigg\}^{\frac{p-1}{p}} \tag{3.25}$$

for any $t\in(0,T]$ and $1\le i\le d_1+d_2$.

By (3.19), we have $\|Q_{t,s}^{-1}\|\le c_0^{-1}ts^{-2(k+1)}$. Combining this with (3.18), we may find a constant $c_5>0$ such that

$$\begin{aligned}&|\alpha_{t,s}(e_j)|\le c_5t^{-2k}+c_5 1_{\{j\le d_1\}}t^{-2k-1},\\ &|\partial_s\alpha_{t,s}(e_j)|\le c_5t^{-2k-1}+c_5 1_{\{j\le d_1\}}t^{-2k-2},\\ &|g_{t,s}(e_j)|\le c_5t+c_5 1_{\{j\le d_1\}}\quad \text{for } 0\le s<t\le T \text{ and } 1\le j\le d_1+d_2.\end{aligned}$$

Now noting that $\|\sigma(s)^{-1}\|\le K$, together with the previous estimates, we may conclude that there is a constant $c_6>0$ such that

$$\begin{aligned}\partial_s h_{t,s}(e_j)| &= \big|\sigma(s)^{-1}\{\nabla_{g_{t,s}(e_j),\alpha_{t,s}(e_j)}b(s,X_s)-\partial_s\alpha_{t,s}(e_j)\}\big|\\ &\le c_6t^{-2k-1}+c_6 1_{\{j\le d_1\}}t^{-2k-2}\end{aligned}$$

for any $0\le s<t\le T$ and for $1\le j\le d_1+d_2$. This together with (3.25) enables us to find a constant $c_7>0$ such that

$$\big(\mathbb{E}[|M_t(e_i)|^{\frac{p}{p-1}}]\big)^{\frac{p-1}{p}} \le c_7\begin{cases} t^{-2k-\frac{3}{2}}, & \text{if } \sup_{j\le d_1}\|(\nabla X_t)^{-1}_{ji}\|_\infty>0,\\ t^{-2k-\frac{1}{2}}, & \text{otherwise.}\end{cases}$$

Combining this with (3.24) we derive (3.16) for some constant $c(p)>0$.

(d) For the case where $Z(s,x)=(s,x^{(2)})$ is independent of $x^{(1)}$, we have $\nabla_j X_t^i=0$ for $i\ge d_1+1$ and $j\le d_1$, so that the previous estimate implies that

$$\big(\mathbb{E}[|M_t(e_i)|^{\frac{p}{p-1}}]\big)^{\frac{p-1}{p}} \le c_7t^{-2k-\frac{1}{2}}\quad \forall t\in(0,T],$$

where $d_1+1\le i\le d_1+d_2$. Combining this with (3.24) we derive we derive (3.17) with some constant $c(p)>0$ and $\xi_t:=t^{-2k-\frac{1}{2}}$. □

Lemma 3.4. *Let* (3.3) *and* (3.4) *hold, let* $b \in C_b^1$*, and let* Z *be locally bounded having bounded* ∇Z*. Then for any* $p > 1$ *there exists a constant* $c(p) > 0$ *such that the semigroup* P_t *associated with* (3.15) *satisfies the Harnack inequality*

$$(3.26) \qquad |P_t f(x)|^p(x) \le \big(P_t|f|^p(y)\big)\mathrm{e}^{\frac{c(p)|x-y|^2}{t^{4k+3}}}, \quad t \in (0,T], x, y \in \mathbb{R}^{d_1+d_2}, f \in \mathscr{B}_b(\mathbb{R}^{d_1+d_2}).$$

Proof. (a) Let $\bar{P}_t$ be the Markov semigroup associated with (3.15) for $b = 0$. By [22, Corollary 4.3(1)] for $l_1 = 0$, we find a constant $c_1(p) > 0$ such that

$$(3.27) \qquad \hat{P}_t|f|(x) \le (\hat{P}_t|f|^{p^{\frac{1}{3}}}(y))^{p^{-\frac{1}{3}}}\mathrm{e}^{\frac{c_1(p)|x-y|^2}{t^{4k+3}}}, \quad t \in (0,T], x, y \in \mathbb{R}^{d_1+d_2}$$

holds for all $f \in \mathscr{B}_b(\mathbb{R}^{d_1+d_2})$.

On the other hand, since b is bounded, there exists a constant $c_2(p) > 0$ such that

$$P_t|f| \le \mathrm{e}^{c_2(p)t}(\hat{P}_t|f|^{p^{\frac{1}{3}}})^{p^{-\frac{1}{3}}}), \quad \hat{P}_t|f| \le \mathrm{e}^{c_2(p)t}(P_t|f|^{p^{\frac{1}{3}}})^{p^{-\frac{1}{3}}}), \quad t \in [0,T].$$

Combining this with (3.27) we find a constant $c_3(p) > 0$ such that

$$(3.28) \qquad P_t|f|(x) \le (P_t|f|^p(y))^{\frac{1}{p}}\mathrm{e}^{c_3(p)+\frac{c_3(p)|x-y|^2}{t^{4k+3}}}, \quad t \in (0,T], x, y \in \mathbb{R}^{d_1+d_2}$$

holds for all $f \in \mathscr{B}_b(\mathbb{R}^{d_1+d_2})$.

Finally, since ∇b and ∇Z are bounded, $(\nabla X_t)_{t\in[0,T]}$ is bounded as well. So, there exists a constant $c_4 > 0$ such that

$$|\nabla P_t f| \le c_4 P_t|\nabla f|, \quad t \in [0,T], f \in C_b^1(\mathbb{R}^{d_1+d_2}).$$

According to the proof of [15, Theorem 2.2], this together with (3.28) implies (3.26) for some constant $c(p) > 0$. □

4 Distribution dependent stochastic Hamilton system

Consider the following distribution dependent SDEs

$$(4.1) \qquad \begin{cases} \mathrm{d}X_t^{(1)} = \big\{AX_t^{(1)} + BX_t^{(2)} + b(X_t, \mathscr{L}_{X_t})\big\}\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = Z(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, \mathscr{L}_{X_t})\mathrm{d}W_t \end{cases}$$

for $t \in [0,T]$, where $X_t = (X_t^{(1)}, X_t^{(2)})$ is $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ valued process. The coefficients A, B, b, Z and σ satisfy the following assumption.

(C_1) A, B and b satisfy conditions 1) and 2) in (B_3), $Z(t,x,\mu)$ is differentiable in $x \in \mathbb{R}^{d_1+d_2}$, and there exists a constant $K > 0$ such that

$$\|\nabla b(t,\cdot,\mu)(x) - \nabla b(t,\cdot,\mu)(y)\| \le K|x-y|,$$

$$|b(t,x,\mu) - b(t,y,\nu)| + \|\sigma(t,\mu) - \sigma(t,\nu)\| \le K\big\{|x-y| + \mathbb{W}_2(\mu,\nu)\big\}$$

$$\|Z(t,0,\delta_0)| + \|\sigma(t,\mu)\| + \|\sigma(t,\mu)^{-1}\| \le K$$

for $t \in [0,T]$, $x, y \in \mathbb{R}^{d_1+d_2}$ and $\mu, \nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$.

By, for instance, [21, Theorem 2.1], under this assumption the SDE (4.1) is well-posed for distributions in $\mathscr{P}_2(\mathbb{R}^{d_1+d_2})$, and $P_t^*\mu := \mathscr{L}_{X_t}$ for the solution X_t with initial distribution μ satisfies

$$\sup_{t\in[0,T]} \mathbb{W}_2(P_t^*\mu, P_t^*\nu) \le C\mathbb{W}_2(\mu,\nu), \quad \forall \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}) \tag{4.2}$$

for some constant $C > 0$.

Theorem 4.1. *Assume that condition (C_1) is satisfied.*

(1) *There exists a constant $c > 0$ such that*

$$\mathrm{Ent}(P_t^*\mu | P_t^*\nu) \le \frac{c}{t^{(4k+2)(4k+3)}} \mathbb{W}_2(\mu,\nu)^2, \qquad \forall t \in (0,T]. \tag{4.3}$$

If $Z(t,x,\mu) = Z(t,x^{(2)},\mu)$ does not dependent on $x^{(1)}$, then

$$\mathrm{Ent}(P_t^*\mu | P_t^*\nu) \le \frac{c}{t^{(4k+1)(4k+3)}} \mathbb{W}_2(\mu,\nu)^2, \qquad \forall t \in (0,T]. \tag{4.4}$$

(2) *If $Z(t,x,\mu) = Z(x,\mu)$ and $\sigma(t,\mu) = \sigma(\mu)$ do not depend on t, and there exist constants $c', \lambda > 0$ such that*

$$\mathbb{W}_2(P_t^*\mu, P_t^*\nu)^2 \le c'\mathrm{e}^{-\lambda t}\mathbb{W}_2(\mu,\nu)^2, \quad \forall t \ge 0 \quad \text{and} \quad \forall \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2}),$$

then P_t^* has a unique invariant probability measure $\bar\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$, and

$$\mathrm{Ent}(P_t^*\mu | \bar\mu) \le cc'\mathrm{e}^{-\lambda(t-1)}\mathbb{W}_2(\mu,\bar\mu)^2$$

for any $t \ge 0$ and for every $\mu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$.

Proof. It suffices to prove the first assertion. To this end, given $(\mu,\nu \in \mathscr{P}_2(\mathbb{R}^{d_1+d_2})$, let

$$\begin{aligned} &Z_1^{(2)}(t,x) := Z(t,x,P_t^*\mu), \quad Z_2^{(2)}(t,x) := Z(t,x,P_t^*\nu),\\ &\sigma_1(t) := \sigma(t,P_t^*\mu) \quad \sigma_2(t) := \sigma(t,P_t^*\nu), \quad t\in[0,T]. \end{aligned}$$

Then the desired estimates in Theorem 4.1(1) follow from Corollary 3.2 and (4.2). □

To illustrate this result, we consider the following typical example for $d_1 = d_2 = d$:

$$\begin{cases} \mathrm{d}X_t^{(1)} = \big\{BX_t^{(2)} + b(X_t)\big\}\mathrm{d}t, \\ \mathrm{d}X_t^{(2)} = \sigma(\mathscr{L}_{X_t})\mathrm{d}W_t - \Big(B^*\nabla V(\cdot,\mathscr{L}_{X_t})(X_t) + \beta B^*(BB^*)^{-1}X_t^{(1)} + X_t^{(2)}\Big)\,\mathrm{d}t, \end{cases} \tag{4.5}$$

where $\beta > 0$ is a constant, B is an invertible $d\times d$-matrix, and

$$V : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^{2d}) \to \mathbb{R}^d$$

is measurable and differentiable in $x^{(1)} \in \mathbb{R}^d$. Let

$$\begin{aligned} &\psi(x,y) := \sqrt{|x^{(1)} - y^{(1)}|^2 + |B(x^{(2)} - y^{(2)})|^2} \quad \text{for } x,y \in \mathbb{R}^{2d},\\ &\mathbb{W}_2^\psi(\mu,\nu) := \inf_{\pi\in\mathscr{C}(\mu,\nu)} \bigg(\int_{\mathbb{R}^{2d}\times\mathbb{R}^{2d}} \psi^2\mathrm{d}\pi\bigg)^{\frac12} \quad \text{for } \mu,\nu \in \mathscr{P}_2(\mathbb{R}^{2d}). \end{aligned}$$

We assume that the following technical condition is satisfied.

(C_2) $V(\cdot,\mu)$ is differentiable such that $\nabla V(\cdot,\mu)(x^{(1)})$ is Lipschitz continuous in $(x^{(1)},\mu)\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^{2d})$. Moreover, there exist constants $\theta_1,\theta_2\in\mathbb{R}$ with

$$\theta_1+\theta_2<\beta,$$

such that

$$\begin{aligned}&\big\langle BB^*\{\nabla V(\cdot,\mu)(x^{(1)})-\nabla V(\cdot,\nu)(y^{(1)})\},\ x^{(1)}-y^{(1)}+(1+\beta)B(x^{(2)}-y^{(2)})\big\rangle\\&-\frac{1+\beta}{2\beta}\|B\{\sigma(\mu)-\sigma(\nu)\}\|_{HS}^2\ge-\theta_1\psi(x,y)^2-\theta_2\mathbb{W}_2^{\psi}(\mu,\nu)^2\end{aligned}$$

for any $x,y\in\mathbb{R}^{2d}$ and $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^{2d})$.

Corollary 4.2. *Assume that condition (C_2) is satisfied. Let*

$$\kappa:=\frac{2(\beta-\theta_1-\theta_2)}{2+2\beta+\beta^2+\sqrt{\beta^4+4}}>0. \tag{4.6}$$

For any $\kappa'\in(0,\kappa)$, when $\|\nabla b\|_\infty$ is small enough, P_t^* has a unique invariant probability measure $\bar\mu\in\mathscr{P}_2(\mathbb{R}^{2d})$, and there exists a constant $c>0$ such that

$$\mathbb{W}_2(P_t^*\mu,\bar\mu)^2+\mathrm{Ent}(P_t^*\mu|\bar\mu)\le\frac{c\mathrm{e}^{-2\kappa' t}}{(1\wedge t)^3}\mathbb{W}_2(\mu,\bar\mu)^2 \tag{4.7}$$

for any $t>0$ and $\mu\in\mathscr{P}_2(\mathbb{R}^{2d})$.

Proof. The proof is completely similar to that of [16, Lemma 5.2] where $\sigma(\mu)=\sigma$ does not depend on μ. By Theorem 4.1, it suffices to find a constant $c'>9$ such that

$$\mathbb{W}_2(P_t^*\mu,P_t^*\nu)^2\le c'\mathrm{e}^{-2\kappa t}\mathbb{W}_2(\mu,\nu)^2 \tag{4.8}$$

for any $t>0$ and $\mu,\nu\in\mathscr{P}_2(\mathbb{R}^{2d})$.

a) Let

$$a:=\Big(\frac{1+\beta+\beta^2}{1+\beta}\Big)^{\frac12},\quad r:=a-\frac{\beta}{a}=\frac{1}{\sqrt{(1+\beta)(1+\beta+\beta^2)}}\in(0,1). \tag{4.9}$$

Define the distance

$$\bar\psi(x,y):=\sqrt{a^2|x^{(1)}-y^{(1)}|^2+|B(x^{(2)}-y^{(2)})|^2+2ra\langle x^{(1)}-y^{(1)},B(x^{(2)}-y^{(2)})\rangle}. \tag{4.10}$$

According to the proof of [16, Lemma 5.2], we have

$$\bar\psi(x,y)^2\le\frac{2+2\beta+\beta^2+\sqrt{\beta^4+4}}{2(1+\beta)}\psi(x,y)^2,\quad\forall x,y\in\mathbb{R}^{2d}, \tag{4.11}$$

and there exists a constant $C>1$ such that

$$C^{-1}|x-y|\le\bar\psi(x,y)\le C|x-y|,\quad\forall x,y\in\mathbb{R}^{2d}. \tag{4.12}$$

b) Let X_t and Y_t solve (4.5) with $\mathscr{L}_{X_0} = \mu, \mathscr{L}_{Y_0} = \nu$ such that

$$\mathbb{W}_2(\mu,\nu)^2 = \mathbb{E}\big[|X_0 - Y_0|^2\big]. \tag{4.13}$$

Let $\Xi_t = X_t - Y_t$, $\mu_t = P_t^*\mu := \mathscr{L}_{X_t}$ and $\nu_t := P_t^*\nu = \mathscr{L}_{Y_t}$. By using (C_2), Itô's formula, and noting that (4.9) implies

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

we obtain

$$\begin{aligned}
\frac{1}{2}\mathrm{d}\left(\bar{\psi}(X_t,Y_t)^2\right) &= \frac{1}{2}\|B\left(\sigma(\mu_t) - \sigma(\nu_t)\right)\|_{HS}^2 + \big\langle a^2\Xi_t^{(1)} + raB\Xi_t^{(2)}, B\Xi_t^{(2)} + b(X_t) - b(Y_t)\big\rangle \mathrm{d}t\\
&\quad - \big\langle B^*B\Xi_t^{(2)} + raB^*\Xi_t^{(1)},\ \beta B^*(BB^*)^{-1}\Xi_t^{(1)} + \Xi_t^{(2)}\big\rangle \mathrm{d}t\\
&\quad + \big\langle B^*B\Xi_t^{(2)} + raB^*\Xi_t^{(1)},\ B^*\{\nabla^{(1)}V(Y_t^{(1)},\nu_t) - \nabla^{(1)}V(X_t^{(1)},\mu_t)\}\big\rangle \mathrm{d}t\\
&\le \Big\{ -(1-ra)|B\Xi_t^{(2)} + (a^2-\beta-ra)\langle \Xi_t^{(1)}, B\Xi_t^{(2)}\rangle + \big[\|\nabla b\|_\infty(a^2+ra) - ra\beta\big]|\Xi_t^{(1)}|^2\\
&\quad + \big\langle B^*B\Xi_t^{(2)} + (1+\beta)^{-1}B^*\Xi_t^{(1)},\ B^*\{\nabla^{(1)}V(Y_t^{(1)},\nu_t) - \nabla^{(1)}V(X_t^{(1)},\mu_t)\}\big\rangle\Big\}\mathrm{d}t\\
&\le \Big\{\frac{\theta_2}{1+\beta}\mathbb{W}_2^{\psi}(\mu_t,\nu_t)^2 - \frac{\beta-\theta_1}{1+\beta}\psi(X_t,Y_t)^2 + \|\nabla b\|_\infty(a^2+ra)|\Xi_t^{(1)}|^2\Big\}\mathrm{d}t.
\end{aligned}$$

By (4.11) and the fact that

$$\mathbb{W}_2^{\psi}(\mu_t,\nu_t)^2 \le \mathbb{E}[\psi(X_t,Y_t)^2],$$

for $\kappa > 0$ in (4.6), when $\|\nabla b\|_\infty$ is small enough we find a constant $\kappa' \in (0,\kappa)$ such that we obtain

$$\begin{aligned}
&\frac{1}{2}\left(\mathbb{E}[\bar{\psi}(X_t,Y_t)^2] - \mathbb{E}[\bar{\psi}(X_s,Y_s)^2]\right)\\
&\le \|\nabla b\|_\infty(a^2+ra)\int_s^t \mathbb{E}[|\Xi_u^{(1)}|^2]\mathrm{d}u - \frac{\beta-\theta_1-\theta_2}{1+\beta}\int_s^t \mathbb{E}[\psi(X_u,Y_u)^2]\mathrm{d}u\\
&\le -\kappa'\int_s^t \mathbb{E}[\bar{\psi}(X_u,Y_u)^2]\mathrm{d}u, \quad t \ge s \ge 0.
\end{aligned}$$

By Gronwall's inequality, we then deduce that

$$\mathbb{E}[\bar{\psi}(X_t,Y_t)^2] \le \mathrm{e}^{-2\kappa' t}\mathbb{E}[\bar{\psi}(X_0,Y_0)^2]$$

for $t \ge 0$. Combining this with (4.12) and (4.13), we may conclude that there is a constant $c > 0$ such that (4.8) holds. □

To conclude this paper, we present the following example of degenerate nonlinear granular media equations, see [3] and [8] for the study of non-degenerate linear granular media equations.

Example 4.1 (Degenerate nonlinear granular media equation). Let $d \in \mathbb{N}$ and $W \in C^\infty(\mathbb{R}^d \times \mathbb{R}^{2d})$. Consider the following PDE for probability density functions $(\rho_t)_{t\geq 0}$ on $\mathbb{R}^{2d} = \mathbb{R}^d \times \mathbb{R}^d$:

$$\begin{aligned}\partial_t \rho_t(x) &= \frac{1}{2}\mathrm{tr}\big\{\sigma(\rho_t)\sigma(\rho_t)^*(\nabla^{(2)})^2\big\}\rho_t(x) - \langle \nabla^{(1)}\rho_t(x), x^{(2)} + b(x)\rangle \\ &+ \langle \nabla^{(2)}\rho_t(x), \nabla^{(1)}(W \circledast \rho_t)(x^{(1)}) + \beta x^{(1)} + x^{(2)}\rangle,\end{aligned} \tag{4.14}$$

where $x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^{2d}$, $t \geq 0$. $\beta > 0$ is a constant, and

$$(W \circledast \rho_t)(x^{(1)}) := \int_{\mathbb{R}^{2m}} W(x^{(1)}, z)\rho_t(z)\mathrm{d}z, \quad x^{(1)} \in \mathbb{R}^d$$

stands for the mean field interaction.

If there exist constants $\theta, \alpha > 0$ with

$$\theta\Big(\frac{1}{2} + \sqrt{2 + 2\beta + \beta^2}\Big) + \frac{\alpha(1+\beta)}{2\beta} < \beta,$$

such that

$$\begin{aligned}&|\nabla W(\cdot, z)(v) - \nabla W(\cdot, \bar{z})(\bar{v})| \leq \theta\big(|v - \bar{v}| + |z - \bar{z}|\big), \quad \forall v, \bar{v} \in \mathbb{R}^d, \text{ and } \forall z, \bar{z} \in \mathbb{R}^{2d}, \\ &\|\sigma(\mu) - \sigma(\nu)\|_{HS}^2 \leq \alpha \mathbb{W}_2(\mu, \nu)^2, \quad \forall \mu, \nu \in \mathscr{P}_2(\mathbb{R}^{2d}),\end{aligned} \tag{4.15}$$

then for any $\kappa' \in (0, \kappa)$, when $\|\nabla b\|_\infty$ is small enough there exists a unique probability measure $\bar{\mu} \in \mathscr{P}_2(\mathbb{R}^{2d})$ and a constant $c > 0$ such that for any probability density functions $(\rho_t)_{t\geq 0}$ solving (4.14), $\mu_t(\mathrm{d}x) := \rho_t(x)\mathrm{d}x$ satisfies

$$\mathbb{W}_2(\mu_t, \bar{\mu})^2 + \mathrm{Ent}(\mu_t|\bar{\mu}) \leq c\mathrm{e}^{-\kappa' t}\mathbb{W}_2(\mu_0, \bar{\mu})^2, \quad \forall t \geq 1 \tag{4.16}$$

where

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

To prove this claim, let (X_t, Y_t) solve (4.5) for

$$B := I_d, \quad \psi(x, y) = |x - y|, \quad \text{and } V(x, \mu) := \int_{\mathbb{R}^{2d}} W(x, z)\mu(\mathrm{d}z). \tag{4.17}$$

As shown in the proof of [16, Example 2.2], ρ_t solves (4.14) if and only if $\rho_t(x) = \frac{\mathrm{d}(P_t^*\mu)(\mathrm{d}x)}{\mathrm{d}x}$, where $P_t^*\mu := \mathscr{L}_{X_t}$.

By Corollary 4.2, we only need to verify (C_2) for B, V in (4.17) and

$$\theta_1 = \theta\Big(\frac{1}{2} + \sqrt{2 + 2\beta + \beta^2}\Big), \quad \theta_2 = \frac{\theta}{2}\sqrt{2 + 2\beta + \beta^2} + \frac{\alpha(\beta+1)}{2\beta}, \tag{4.18}$$

so that the desired assertion holds for

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

For simplicity, let ∇^v denote the gradient in v. By (4.15) and $V(x,\mu) := \mu(W(x,\cdot))$, for any constants $\alpha_1, \alpha_2, \alpha_3 > 0$ we have

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

Take

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

We have

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

Combining this with (4.15) and (4.18), we derive

$$
I - \frac{\beta + 1}{2\beta}\|\sigma(\mu) - \sigma(\nu)\|_{HS}^2 \ge -\theta_1|x - y)|^2 - \theta_2\mathbb{W}_2(\mu,\nu)^2,
$$

and therefore condition (C_2) is satisfied for B, ψ and V in (4.17) .

References

[1] V. I. Bogachev, M. Röckner, S. V. Shaposhnikov, *Distances between transition probabilities of diffusions and applications to nonlinear Fokker-Planck-Kolmogorov equations,* J. Funct. Anal. 271(2016), 1262-1300.

[2] E. Camrud, D. P. Herzog, G. Stoltz, M. Gordina *Weighted L^2-congractivity of Langevin dynamics with singular potentials,* Nonlinearity, 35(2022), 998-1035.

[3] 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.

[4] J. Dolbeault, C. Mouhot, C. Sschmeiser, *Hypocoercivity for linear kinetic equations conserving mass,* Trans. Amer. Math. Soc. 367(2015), 3807-3828.

[5] M. Grothaus, P. Ren, F.-Y. Wang, *Singular degenerate SDEs: well-posedness and exponential ergodicity,* arXiv:2305.00129.

[6] M. Grothaus, P. Stilgenbauer, *A hypocoercivity related ergodicity method for singularly distorted non-symmetric diffusions,* Integral Equations and Operator Theory 83(2015), 331-379.

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

[8] A. Guillin, W. Liu, L. Wu, *Uniform Poincaré and logarithmic Sobolev inequalities for mean field particle systems,* Ann. Appl. Probab. 32(2022),1590-1614.

[9] S. Hu, X. Wang, *Subexponential decay in kinetic Fokker-Planck equation: weak hypocoercivity,* Bernoulli 25(2019), 174-188.

[10] X. Huang, P. Ren, F.-Y. Wang, *Probability distance estimates between diffusion processes and applications to singular McKean-Vlasov SDEs,* arXiv:2304.07562.

[11] X. Huang, F.-Y. Wang, *Regularities and exponential ergodicity in entropy for SDEs driven by distribution dependent noise,* to appear in Bernoulli, arXiv:2209.14619.

[12] S. Menozzi, A. Pesce, X. Zhang, *Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift,* J. Diff. Equat. 272(2021), 330-369.

[13] E. Priola, F.-Y. Wang, *Gradient estimates for diffusion semigroups with singular coefficients,* J. Funct. Anal. 236(2006), 244-264.

[14] Z. Qian, F. Russo, W. Zheng, *Comparison theorem and estimates for transition probability densities of diffusion processes,* Probab. Theory Related Fields 127(2003), 388-406.

[15] P. Ren, *Singular McKean-Vlasov SDEs: well-posedness, regularities and Wang's Harnack inequality,* Stoch. Proc. Appl. 156(2023), 291-311.

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

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

[18] M. Röckner, F.-Y. Wang, *Weak Poincaré inequalities and convergence rates of Markov semigroups,* J. Funct. Anal. 185(2001), 564-603.

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

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

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

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