Entropy Estimate Between Diffusion Processes and Application to McKean-Vlasov SDEs*

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

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

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

wangfy@tju.edu.cn

March 17, 2023

Abstract

By developing a new technique called the bi-coupling argument, we estimate the relative entropy between different diffusion processes in terms of the distances of initial distributions and drift-diffusion coefficients. As an application, the log-Harnack inequality is established for McKean-Vlasov SDEs with multiplicative distribution dependent noise, which appears for the first time in the literature.

AMS subject Classification: 35A01, 35D35.
Keywords: Entropy, bi-coupling, diffusion process, McKean-Vlasov SDE, log-Harnack inequality.

1 Introduction

In this paper, we introduce the bi-coupling argument to estimate the relative entropy between two diffusion processes. The relative entropy, also called the Kullback-Leibler divergence or the information divergence, is a physical quantity measuring the chaos of one distribution with respect to another. As an application, we establish the log-Harnack inequality for McKean-Vlasov SDEs with multiplicative distribution dependent noise, which is unknown so far.

As a member in the family of dimension-free Hanranck inequalities (see [18, 19, 21]), the log-Harnack inequality bounds the entropy by the quadratic Wasserstein distance, hence can be regarded as an inverse of the Talagrand inequality [17]. The log-Harnack inequality has crucial applications in optimal transport, curvature on Riemennian manifolds or metric measure spaces,

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

and exponential ergodicity in entropy, see for instance [1, 15, 19]. See [20] for more applications of this type inequalities.

Let $T > 0$, and let Γ be the space of (a, b), where

$$b : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d, \quad a : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$$

are measurable, and for any $(t, x) \in [0, T] \times \mathbb{R}^d$, $a(t, x)$ is positive definite. For any $(a, b) \in \Gamma$, consider the time dependent second order differential operators on $\mathbb{R}^d$:

$$L_t^{a,b} := \mathrm{tr}\{a(t, \cdot)\nabla^2\} + b(t, \cdot) \cdot \nabla, \quad t \in [0, T].$$

Let $(a_i, b_i) \in \Gamma, i = 1, 2$, such that for any $s \in [0, T)$, each $(L_t^{a_i,b_i})_{t\in[s,T]}$ generates a unique diffusion process $(X_{s,t}^{i,x})_{(t,x)\in[s,T]\times\mathbb{R}^d}$ with $X_{s,s}^{i,x} = x$, and for any $t \in (s, T]$, the distribution $P_{s,t}^{i,x}$ of $X_{s,t}^{i,x}$ has positive density function $p_{s,t}^{i,x}$ with respect to the Lebesgue measure. When $s = 0$, we simply denote

$$X_{0,t}^{i,x} = X_t^{i,x}, \quad P_{0,t}^{i,x} = P_t^{i,x}.$$

The associated Markov semigroup $(P_{s,t}^i)_{0\le s\le t\le T}$ is given by

$$P_{s,t}^{(i)} f(x) := \mathbb{E}[f(X_{s,t}^{i,x})], \quad 0 \le s \le t \le T, x \in \mathbb{R}^d, f \in \mathscr{B}_b(\mathbb{R}^d).$$

If the initial value is random with distributions $\nu \in \mathscr{P}$, where $\mathscr{P}$ is the set of all probability measures on $\mathbb{R}^d$, we denote the diffusion process by $X_t^{i,\nu}$, which has distribution

$$P_t^{i,\nu} = \int_{\mathbb{R}^d} P_t^{i,x} \nu(\mathrm{d}x), \quad i = 1, 2, \ t \in (0, T].$$

Let $p_t^{i,\nu}$ be the density function of $P_t^{i,\nu}$ with respect to the Lebesgue measure.

We estimate the relative entropy

$$\mathrm{Ent}(P_t^{1,\nu_1}|P_t^{2,\nu_2}) := \int_{\mathbb{R}^d} \Big(\log \frac{\mathrm{d}P_t^{1,\nu_1}}{\mathrm{d}P_t^{2,\nu_2}}\Big)\mathrm{d}P_t^{1,\nu_1} = \mathbb{E}\Big[\Big(\log \frac{p_t^{1,\nu_1}}{p_t^{2,\nu_2}}\Big)(X_t^{1,\nu_1})\Big], \quad t \in (0, T].$$

Before moving on, let us recall a nice entropy inequality derived in [5]. For a $d \times d$-matrix valued function $a = (a^{kl})_{1\le k,l\le d}$, the divergence is an $\mathbb{R}^d$-valued function defined by

$$\mathrm{div} a := \Big(\sum_{l=1}^d \partial_l a^{kl}\Big)_{1\le k\le d},$$

where $\partial_l := \frac{\partial}{\partial x^l}$ for $x = (x^l)_{1\le l\le d} \in \mathbb{R}^d$. Let

$$\begin{aligned}\Phi^\nu(s, y) :=& (a_1(s, y) - a_2(s, y))\nabla \log p_s^{1,\nu}(y) + \mathrm{div}\{a_1(s, \cdot) - a_2(s, \cdot)\}(y) \\ &+ b_2(s, y) - b_1(s, y), \quad s \in (0, T], y \in \mathbb{R}^d, \nu \in \mathscr{P},\end{aligned}$$

where ∇ is the gradient operator for weakly differentiable functions on $\mathbb{R}^d$. In particular, $\|\nabla f\|_\infty$ is the Lipschitz constant of f.

By [5, Theorem 1.1], the entropy inequality

$$\mathrm{Ent}(P_t^{1,\nu}|P_t^{2,\nu}) \le \frac{1}{2} \int_0^t \mathbb{E}\big[\big|a_2(s, X_s^{1,\nu})^{-\frac{1}{2}}\Phi^\nu(s, X_s^{1,\nu})\big|^2\big]\mathrm{d}s, \quad t \in (0, T] \tag{1.1}$$

holds under the following assumption (H).

(H) For each $i=1,2$, b_i is locally bounded, and there exists a constant $K>1$ such that

$$\|a_i(t,x)\| \vee \|a_i(t,x)^{-1}\| \vee \|\nabla a_i(t,\cdot)(x)\| \le K, \quad (t,x)\in[0,T]\times\mathbb{R}^d.$$

Moreover, at leat one of the following conditions hold:

(1) $\int_0^T \mathbb{E}\big[\frac{\|a_2(t,X_t^{1,\nu})\|}{1+|X_t^{1,\nu}|^2} + \frac{|b_2(t,X_t^{1,\nu})|+|\Phi^\nu(t,X_t^{1,\nu})|}{1+|X_t^{1,\nu}|}\big]\mathrm{d}t<\infty$;

(2) there exist $1\le V\in C^2(\mathbb{R}^d)$ with $V(x)\to\infty$ as $|x|\to\infty$, and a constant $K>0$ such that

$$L_t^{a_2,b_2}V(x)\le KV(x), \quad \int_0^T \mathbb{E}\Big[\frac{|\langle \Phi^\nu(t,X_t^{1,\nu}), \nabla V(X_t^{1,\nu})\rangle|}{V(X_t^{1,\nu})}\Big]\mathrm{d}t<\infty.$$

It is well known that (H) implies the existence and uniqueness of the diffusion processes $(X_t^{i,\nu})_{i=1,2}$ for any $\nu\in\mathscr{P}$, and the existence of the density functions $(p_t^{i,\nu})_{i=1,2}$, see for instance [4].

As observed in [5, Remark 1.4] that one may have

$$\int_0^t \mathbb{E}\big[|\nabla \log p_s^{1,\nu}|^2(X_s^{1,\nu})\big]\mathrm{d}s<\infty,$$

provided ν has finite information entropy, i.e. $\rho(x):=\frac{\mathrm{d}\nu}{\mathrm{d}x}$ satisfies $\int_{\mathbb{R}^d}(\rho|\log\rho|)(x)\mathrm{d}x<\infty$. In this case, (1.1) provides a non-trivial upper bound for $\mathrm{Ent}(P_t^{1,\nu}|P_t^{2,\nu})$.

However, for a fixed initial value x, i.e. $\nu=\delta_x$, $\mathbb{E}[|\nabla\log p_s^{1,x}|^2(X_s^{1,x})]$ behaves as $\frac{c}{s}$ for some constant $c>0$ and small $s>0$, so that

$$\int_0^t \mathbb{E}[|\nabla\log p_s^{1,x}|^2(X_s^{1,x})]\mathrm{d}s=\infty, \quad t>0.$$

Consequently, the estimate (1.1) becomes invalid when

$$\inf_{(s,x)\in[0,T]\times\mathbb{R}^d}\|a_1(s,x)-a_2(s,x)\|>0. \tag{1.2}$$

To kill the singularity in (1.1) for small $t>0$, we introduce a new technique by constructing an interpolation diffusion process which is coupled with each of the given two diffusion processes respectively, so we call it the bi-coupling argument.

1.1 Entropy estimates for diffusion processes

We make the following assumption (A_1) and (A_2) where b_i may have a Dini continuous term with respect to a Dini function in the class

$$\mathscr{D}:=\Big\{\varphi:[0,\infty)\to[0,\infty) \text{ is increasing and concave}, \varphi(0)=0, \int_0^1\frac{\varphi(s)}{s}\mathrm{d}s<\infty\Big\}.$$

For $\varphi\in\mathscr{D}, t>0$ and a function f on $[0,t]\times\mathbb{R}^d$, let

$$\|f\|_{t,\infty}:=\sup_{x\in\mathbb{R}^d}|f(t,x)|, \quad \|f\|_{r\to t,\infty}:=\sup_{s\in[r,t]}\|f\|_{s,\infty}, \quad r\in[0,t],$$

$$\|f\|_{0\to T,\varphi}:=\sup_{t\in[0,T],x\ne y\in\mathbb{R}^d}\bigg(|f(t,x)|+\frac{|f(t,x)-f(t,y)|}{\varphi(|x-y|)}\bigg).$$

(A_1) For each $i=1,2$, $b_i=b_i^{(0)}+b_i^{(1)}$ is locally bounded, and there exists a constant $K>0$ such that

$$\|b_i^{(0)}\|_{0\to T,\infty}\vee\|\nabla b_i^{(1)}\|_{0\to T,\infty}\vee\|a_i\|_{0\to T,\infty}\vee\|a_i^{-1}\|_{0\to T,\infty}\vee\|\nabla a_i\|_{0\to T,\infty}\le K.$$

(A_2) There exist $i\in\{1,2\}$ and $\varphi\in\mathscr{D}$ such that $\|b_i^{(0)}\|_{T,\varphi}\le K$.

For any $\nu_1,\nu_2\in\mathscr{P}$, let $\mathscr{C}(\nu_1,\nu_2)$ be the set of all couplings of ν_1 and ν_2. Consider the quadratic Wasserstein distance

$$\mathbb{W}_2(\nu_1,\nu_2):=\inf_{\pi\in\mathscr{C}(\nu_1,\nu_2)}\bigg(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^2\pi(\mathrm{d}x,\mathrm{d}y)\bigg)^{\frac{1}{2}}.$$

In the following, $c=c(K,T,d,\varphi)$ stands for a constant depending only on K,T,d and φ.

Theorem 1.1. *Assume (A_1) and (A_2). Then the following assertions hold for some constants $c=c(K,T,d,\varphi)>0$ and $\varepsilon=\varepsilon(K,T,d,\varphi)\in(0,\frac{1}{2}]$.*

(1) *For any $\nu_1,\nu_2\in\mathscr{P}$ and $t\in(0,T]$,*

$$\begin{aligned}\mathrm{Ent}(P_t^{1,\nu_1}|P_t^{2,\nu_2})\le&\frac{c\mathbb{W}_2(\nu_1,\nu_2)^2}{t}+\frac{c}{t}\int_0^t\{\|b_1-b_2\|_{s,\infty}^2+\|a_1-a_2\|_{s,\infty}^2\}\mathrm{d}s\\&+c\Big[\log(1+t^{-1})\|a_1-a_2\|_{\varepsilon t\to t,\infty}^2+\int_{\varepsilon t}^t\|\mathrm{div}(a_1-a_2)\|_{s,\infty}^2\mathrm{d}s\Big].\end{aligned}\tag{1.3}$$

(2) *If there exists a constant $C(K)>0$ such that $\|b_1\|_{0\to T,\infty}\le C(K)$, then*

$$\begin{aligned}\mathrm{Ent}(P_t^{1,\nu_1}|P_t^{2,\nu_2})\le&\frac{c}{t}\bigg(\mathbb{W}_2(\nu_1,\nu_2)^2+\int_0^t\{\|b_1-b_2\|_{s,\infty}^2+\|a_1-a_2\|_{s,\infty}^2\}\mathrm{d}s\bigg)\\&+c\bigg(\|a_1-a_2\|_{\varepsilon t\to t,\infty}^2+\int_{\varepsilon t}^t\|\mathrm{div}(a_1-a_2)\|_{s,\infty}^2\mathrm{d}s\bigg),\quad\nu_1,\nu_2\in\mathscr{P},t\in(0,T].\end{aligned}\tag{1.4}$$

(3) *If there exists a constant $C(K)>0$ such that*

$$\|\nabla^i b_1\|_{0\to T,\infty}+\|\nabla^i a_1\|_{0\to T,\infty}\le C(K),\quad i=1,2,\tag{1.5}$$

then for any $\nu_1,\nu_2\in\mathbb{R}^d$ and $t\in(0,T]$,

$$\begin{aligned}\mathrm{Ent}(P_t^{1,\nu_1}|P_t^{2,\nu_2})\le&\frac{c}{t}\bigg[\mathbb{W}_2(\nu_1,\nu_2)^2+\int_0^t\big(\|b_1-b_2\|_{s,\infty}^2+\|a_1-a_2\|_{t,\infty}^2\big)\mathrm{d}s\bigg]\\&+\int_{\varepsilon t}^t\|\mathrm{div}(a_1-a_2)\|_{s,\infty}^2\mathrm{d}s.\end{aligned}\tag{1.6}$$

1.2 Log-Harnack inequality for DDSDEs

Let $\mathscr{P}_2 := \{\nu \in \mathscr{P} : \nu(|\cdot|^2) < \infty\}$, which is a Polish space under $\mathbb{W}_2$. Consider the following distribution dependent SDE on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b(t, X_t, \mathscr{L}_{X_t})\mathrm{d}t + \sigma(t, X_t, \mathscr{L}_{X_t})\mathrm{d}W_t, \quad t \in [0, T], \tag{1.7}$$

where $\mathscr{L}_{X_t}$ is the distribution of X_t,

$$b : [0, T] \times \mathbb{R}^d \times \mathscr{P}_2 \to \mathbb{R}^d, \quad \sigma : [0, T] \times \mathbb{R}^d \times \mathscr{P}_2 \to \mathbb{R}^d \otimes \mathbb{R}^d$$

are measurable, and W_t is a d-dimensional Brownian motion on a complete filtration probability space $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\in[0,T]}, \mathbb{P})$. When this SDE is well-posed for distributions in $\mathscr{P}_2$, i.e. for any initial value X_0 with $\mathscr{L}_{X_0} \in \mathscr{P}_2$ (correspondingly, any initial distribution $\nu \in \mathscr{P}_2$), the SDE has a unique solution (correspondingly, a unique weak solution) with $(\mathscr{L}_{X_t})_{t\in[0,T]} \in C([0, T]; \mathscr{P}_2)$, the space of all continuous maps from $[0, T]$ to $\mathscr{P}_2$ under the weak topology. In this case, let $P_t^*\nu = \mathscr{L}_{X_t}$ for the solution with $\mathscr{L}_{X_0} = \nu$, and define

$$P_t f(\nu) := \int_{\mathbb{R}^d} f \mathrm{d}(P_t^*\nu), \quad \nu \in \mathscr{P}_2, t \in [0, T], f \in \mathscr{B}_b(\mathbb{R}^d).$$

We investigate the log-Harnack inequality

$$P_t \log f(\nu_1) \le \log P_t f(\nu_2) + \frac{c}{t}\mathbb{W}_2(\mu, \nu)^2, \quad f \in \mathscr{B}_b^+(\mathbb{R}^d), t \in (0, T], \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d), \tag{1.8}$$

where $c > 0$ is a constant, and $\mathscr{B}_b^+(\mathbb{R}^d)$ is the set of all positive functions in $\mathscr{B}_b(\mathbb{R}^d)$. By the definition of Ent and Young's inequality [2, Lemma 2.4], (1.8) is equivalent to the entropy-cost inequality

$$\mathrm{Ent}(P_t^*\nu | P_t^*\mu) \le \frac{c}{t}\mathbb{W}_2(\mu, \nu)^2, \quad t \in (0, T], \mu, \nu \in \mathscr{P}_2(\mathbb{R}^d).$$

When the noise is distribution free, i.e. $\sigma(t, x, \mu) = \sigma(t, x)$ does not depend on the distribution argument μ, (1.8) has been established in [8, 10, 15, 22, 24] under different conditions, see also [6, 7, 23] for extensions to the infinite-dimensional and reflecting models.

However, if the noise coefficient is also distribution dependent, the coupling by change of measures applied in the above references does not apply. Recently, for $\sigma(t, x, \mu) = \sigma(t, \mu)$ independent of the spatial variable x, (1.8) has been established in [11] by using a noise decomposition argument, see also [3] for the study on a special model.

As an application of Theorem 1.1, we are able to establish (1.8) for (1.7) with distribution dependent multiplicative noise. For any $\mu \in C([0, T]; \mathscr{P}_2)$, let

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

Correspondingly to (A_1) and (A_2), we make the following assumption.

- (B) There exists a constant $K > 0$ such that a^μ and $b^\mu = b^{\mu,0} + b^{\mu,1}$ satisfy the following conditions.

(1) For any $\mu \in C([0,T];\mathscr{P}_2)$, b^μ is locally bounded, and for any $(t,x,\mu) \in [0,T]\times\mathbb{R}^d\times\mathscr{P}_2$,

$$\|\nabla b^{\mu,1}\|_{0\to T,\infty} + \|a^\mu\|_{0\to T,\infty} + \|(a^\mu)^{-1}\|_{0\to T,\infty} + \|\nabla a^\mu\|_{0\to T,\infty} \le K.$$

(2) There exists $\varphi \in \mathscr{D}$ such that

$$\|b^{\mu,0}\|_{T,\varphi} \le K, \quad \mu \in C([0,T];\mathscr{P}_2).$$

(3) For any $\nu,\mu \in \mathscr{P}_2$,

$$\|b^\nu - b^\mu\|_{0\to T,\infty} \vee \|a^\nu - a^\mu\|_{0\to T,\infty} \vee \big\|\mathrm{div}(a^\nu - a^\mu)\big\|_{0\to T,\infty} \le K\mathbb{W}_2(\nu,\mu).$$

Theorem 1.2. *Assume (B). Then (1.7) is well-posed for distributions in $\mathscr{P}_2$, and there exists a constant $c = c(K,T,d,\varphi) > 0$ such that (1.8) holds.*

In the next section, we introduce the bi-coupling argument by constructing an interpolation SDE for $X_t^{i,x_i}, i = 1,2$. This SDE has finite entropy with respect to X_t^{1,x_1}, and its density with respect to X_t^{2,x_2} has finite p-moment for some $p > 1$, so that by the entropy inequality in Lemma 2.1, we are able to prove Theorem 1.1 and Theorem 1.2 in Sections 3 and 4 respectively.

2 Bi-coupling and moment estimate on density

Let $\sigma_i = \sqrt{2a_i}, i = 1,2$. According to [14, Theorem 2.1], (A_1) implies the well-posedness of the SDEs:

$$\mathrm{d}X_t^i = b_i(t,X_t^i)\mathrm{d}t + \sigma_i(t,X_t^i)\mathrm{d}W_t, \quad t\in[0,T],\ i=1,2. \tag{2.1}$$

For any $s \in [0,T)$ and $x \in \mathbb{R}^d$, let $X_{s,t}^{i,x}$ be the unique solution for $t \in [s,T]$ with $X_{s,s}^{i,x} = x$. Then $(X_{s,t}^{i,x})_{(t,x)\in[0,T]\times\mathbb{R}^d}$ is the diffusion process generated by $(L_t^{a_i,b_i})_{t\in[s,T]},\ i=1,2$.

For fixed $x_1,x_2 \in \mathbb{R}^d$, let X_t^{i,x_i} solve (2.1) for $X_0^{i,x_i} = x_i$ and $\sigma_i := \sqrt{2a_i}, i = 1,2$. We have

$$P_t^{i,x_i} := \mathscr{L}_{X_t^{i,x_i}}, \quad i=1,2,\ t\in(0,T].$$

To estimate $\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{2,x_2})$ for some $t_1 \in (0,T]$, we choose $t_0 \in (0,\frac{1}{2}t_1]$ and construct a bridge diffusion process $X_t^{\langle t_0\rangle x_1}$ starting at x_1 which is generated by $L_t^{a_1,b_1}$ for $t \in [0,t_0]$ and $L_t^{a_2,b_2}$ for $t \in (t_0,t_1]$. More precisely, let

$$\begin{aligned} &b^{\langle t_0\rangle}(t,\cdot) := 1_{[0,t_0]}(t)b_1(t,\cdot) + 1_{(t_0,t_1]}(t)b_2(t,\cdot),\\ &\sigma^{\langle t_0\rangle}(t,\cdot) := 1_{[0,t_0]}(t)\sigma_1(t,\cdot) + 1_{(t_0,t_1]}(t)\sigma_2(t,\cdot), \quad t\in[0,t_1]. \end{aligned}$$

We consider the interpolation SDE

$$\mathrm{d}X_t^{\langle t_0\rangle x_1} = b^{\langle t_0\rangle}(t,X_t^{\langle t_0\rangle x_1})\mathrm{d}t + \sigma^{\langle t_0\rangle}(t,X_t^{\langle t_0\rangle x_1})\mathrm{d}W_t, \quad X_0^{x_1} = x_1,\ t\in[0,t_1]. \tag{2.2}$$

Let $P_t^{\langle t_0\rangle x_1} := \mathscr{L}_{X_t^{\langle t_0\rangle x_1}}$. We will deduce from (1.1) a finite upper bound for $\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{\langle t_0\rangle x_1})$, where the singularity at $t = 0$ disappears since the distance of diffusion coefficients vanishes for $t \in [0,t_0]$. Moreover, we will estimate the moment for the density of $P_{t_1}^{\langle t_0\rangle x_1}$ with respect to $P_{t_1}^{2,x_2}$, so that by the following Lemma 2.1, we derive the desired upper bound on $\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{2,x_2})$.

Lemma 2.1. *Let μ_1, μ_2 and μ be probability measures on a measurable space $(E, \mathscr{B})$. Then for any $p > 1$,*

$$\mathrm{Ent}(\mu_1|\mu_2) \le p\mathrm{Ent}(\mu_1|\mu) + (p-1)\log\int_E \Big(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_2}\Big)^{\frac{p}{p-1}}\mathrm{d}\mu_2,$$

where the right hand side is set to be infinite if $\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}$ or $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_2}$ does not exist.

Proof. It suffices to prove for the case that $\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}$ and $\frac{\mathrm{d}\mu}{\mathrm{d}\mu_2}$ exist such that the upper bound is finite. In this case, we have

$$\begin{aligned}&\mathrm{Ent}(\mu_1|\mu_2) - \mathrm{Ent}(\mu_1|\mu) = \int_E \Big\{\log\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2} - \log\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}\Big\}\mathrm{d}\mu_1\\ &= \int_E \Big\{\log\frac{\mathrm{d}\mu}{\mathrm{d}\mu_2}\Big\}\mathrm{d}\mu_1 = \frac{p-1}{p}\int_E \Big(\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu_2}\Big)\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_2}\Big)^{\frac{p}{p-1}}\mathrm{d}\mu_2.\end{aligned}$$

Combining with the Young inequality [2, Lemma 2.4], we obtain

$$\mathrm{Ent}(\mu_1|\mu_2) - \mathrm{Ent}(\mu_1|\mu) \le \frac{p-1}{p}\mathrm{Ent}(\mu_1|\mu_2) + \frac{p-1}{p}\log\int_E\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_2}\Big)^{\frac{p}{p-1}}\mathrm{d}\mu_2.$$

□

By Lemma 2.1, for any $p > 1$ we have

$$\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{2,x_2}) \le p\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{\langle t_0\rangle x_1}) + (p-1)\log\int_{\mathbb{R}^d}\bigg(\frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\bigg)^{\frac{p}{p-1}}\mathrm{d}P_{t_1}^{2,x_2}. \tag{2.3}$$

Noting that $a(t,\cdot) - a_1(t,\cdot) = 0$ for $t \in [0, t_0]$, we may apply (1.1) to derive a non-trivial upper bound on the first term in the right hand side of (2.3), see Proposition 3.1 for details. So, in the following, we only estimate the second term.

Proposition 2.2. *Assume (A_1) and (A_2). Then there exist constants $p = p(K,T,d) > 1, \varepsilon = \varepsilon(K,T,d) \in (0, \frac{1}{2}]$ and $c = c(K,T,d) > 0$, such that for any $x_1, x_2 \in \mathbb{R}^d, t_1 \in (0,T]$ and $t_0 = \varepsilon t_1$,*

$$\log\int_{\mathbb{R}^d}\Big(\frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\Big)^{\frac{p}{p-1}}\mathrm{d}P_{t_1}^{2,x_2} \le \frac{c}{t_1}\bigg(|x_1-x_2|^2 + \int_0^{t_1}\{\|a_1-a_2\|_{t,\infty}^2 + \|b_1-b_2\|_{t,\infty}^2\}\mathrm{d}t\bigg).$$

Proof. (a) Let

$$P_t^{\langle t_0\rangle}f(x) := \mathbb{E}[f(X_t^{\langle t_0\rangle x})],\quad P_t^{(2)}f(x) := \mathbb{E}[f(X_t^{2,x})],\quad f \in B_b(\mathbb{R}^d),\ (t,x) \in [0,T]\times\mathbb{R}^d.$$

By first taking $f := n \wedge \big(\frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\big)^{\frac{1}{p-1}}$ then letting $n \to \infty$, we see that the desired estimate follows from

$$\begin{aligned}&\big|P_{t_1}^{\langle t_0\rangle}f(x_1)\big|^p \le \big(P_{t_1}^{(2)}|f|^p(x_2)\big)\\ &\times\exp\bigg[\frac{c(p-1)}{t_1}\bigg(|x_1-x_2|^2 + \int_0^{t_1}\{\|a_1-a_2\|_{t,\infty}^2 + \|b_1-b_2\|_{t,\infty}^2\}\mathrm{d}t\bigg)\bigg],\quad f \in \mathscr{B}_b(\mathbb{R}^d).\end{aligned} \tag{2.4}$$

Let $(P^{(2)}_{s,t})_{0\le s\le t\le T}$ be the semigroup generated by $L_t^{a_2,b_2}$, i.e.

$$P^{(2)}_{s,t}f(x):=\mathbb{E}[f(X^{2,x}_{s,t})],\quad f\in\mathscr{B}_b(\mathbb{R}^d),$$

where $(X^{2,x}_{s,t})_{t\in[s,T]}$ solves

$$\mathrm{d}X^{2,x}_{s,t}=b_2(t,X^{2,x}_{s,t})\mathrm{d}t+\sigma_2(t,X^{2,x}_{s,t})\mathrm{d}W_t,\quad X^{2,x}_{s,s}=x,\ t\in[s,T].$$

By the Markov property and the SDE (2.2), we obtain

$$P^{\langle t_0\rangle}_{t_1}f(x_1)=\mathbb{E}\big[(P^{(2)}_{t_0,t_1}f)(X^{1,x_1}_{t_0})\big],\ \ P^{(2)}_{t_1}f(x_2)=\mathbb{E}\big[(P^{(2)}_{t_0,t_1}f)(X^{2,x_2}_{t_0})\big]. \tag{2.5}$$

By [14, Theorem 2.2] which applies to a more general setting where $b_2^{(0)}$ only satisfies a local integrability condition, there exists constants $p_1=p_1(K,T,d)>0$ and $c_1=c_1(K,T,d)>0$ such that

$$\big|P^{(2)}_{t_0,t_1}f(x)\big|^{p_1}\le\big(P^{(2)}_{t_0,t_1}|f|^{p_1}(y)\big)\mathrm{e}^{\frac{c_1|x-y|^2}{t_1}},\quad f\in\mathscr{B}_b(\mathbb{R}^d),x,y\in\mathbb{R}^d. \tag{2.6}$$

Combining this with (2.5) and Jensen's inequality, for $p:=2p_1$ we obtain

$$\begin{aligned}
\big|P^{\langle t_0\rangle}_{t_1}f(x_1)|^p&=\big|\mathbb{E}[P^{(2)}_{t_0,t_1}f(X^{1,x_1}_{t_0})]\big|^{2p_1}\le\Big(\mathbb{E}\big[|P^{(2)}_{t_0,t_1}f|^{p_1}(X^{1,x_1}_{t_0})\big]\Big)^2\\
&\le\bigg\{\mathbb{E}\Big[\big(P^{(2)}_{t_0,t_1}|f|^{p_1}(X^{2,x_2}_{t_0})\big)\exp\Big(\frac{c_1|X^{1,x_1}_{t_0}-X^{2,x_2}_{t_0}|^2}{t_1}\Big)\Big]\bigg\}^2\\
&\le\big(\mathbb{E}\big[P^{(2)}_{t_0,t_1}|f|^{2p_1}(X^{2,x_2}_{t_0})\big]\big)\mathbb{E}\bigg[\exp\Big(\frac{2c_1|X^{1,x_1}_{t_0}-X^{2,x_2}_{t_0}|^2}{t_1}\Big)\bigg]\\
&=\big(P^{(2)}_{t_1}|f|^p(x_2)\big)\mathbb{E}\bigg[\exp\Big(\frac{2c_1|X^{1,x_1}_{t_0}-X^{2,x_2}_{t_0}|^2}{t_1}\Big)\bigg].
\end{aligned} \tag{2.7}$$

Thus, to prove (2.4), it remains to estimate the expectation term in the upper bound.

(b) Since the exponential term is symmetric in $(X^{1,x_1}_{t_0},X^{2,x_2}_{t_0})$, without loss of generality, in (A_2) we may and do assume that $\|b_1^{(0)}\|_{0\to T,\varphi}\le K$. We shall use Zvonkin's transform to kill this non-Lipschitz term. By [27, Theorem 2.1], for fixed $p,q\in(2,\infty)$ with $\frac{d}{p}+\frac{2}{q}<1$, there exist constants $c_1=c_1(K,T,d,p,q)>0$ and $\beta=\beta(p,q)\in(0,1)$ such that for any $\lambda>0$, the PDE

$$(\partial_t+L_t^{a_1,b_1}-\lambda)u_t=-b_1^{(0)}(t,\cdot),\quad t\in[0,T],u_T=0 \tag{2.8}$$

has a unique solution satisfying

$$\lambda^\beta(\|u\|_{0\to T,\infty}+\|\nabla u\|_{0\to T,\infty})+\|\partial_t u\|_{\tilde L^p_q}+\|\nabla^2u\|_{\tilde L^p_q}\le c_1, \tag{2.9}$$

where

$$\|f\|_{\tilde L^p_q}:=\sup_{z\in\mathbb{R}^d}\bigg(\int_0^T\|1_{B(z,1)}f(t,\cdot)\|^q_{L^p(\mathbb{R}^d)}\mathrm{d}t\bigg)^{\frac{1}{q}}. \tag{2.10}$$

Let $P_{s,t}^{a_1,b_1^{(1)}}$ be the Markov semigroup generated by $L_t^{a_1,b_1^{(1)}}$, and let $p_{s,t}^{a_1,b_1^{(1)}}$ be the heat kernel with respect to the Lebesgue measure. By Duhamel's formula, we have

$$u_s = \int_s^T \mathrm{e}^{-\lambda(t-s)} P_{s,t}^{a_1,b_1^{(1)}} \big\{\nabla_{b_1^{(0)}} u_t + b_1^{(0)}(t,\cdot)\big\}\mathrm{d}t, \quad s \in [0,T]. \tag{2.11}$$

On the other hand, let ∇_x^2 be the Hessian operator in x. By [12, Theorem 1.2], under (A_1) we find a constant $\delta = \delta(K,T,d) > 1$ such that

$$|\nabla_x^2 p_{s,t}^{a_1,b_1^{(1)}}(x,y)| \le \frac{\lambda}{t-s} g_\delta(t-s,x,y), \quad 0 \le s < t \le T, x,y \in \mathbb{R}^d$$

holds for

$$g_\delta(r,x,y) := (\pi\delta r)^{-\frac{d}{2}} \mathrm{e}^{-\frac{|\theta_{s,t}(x)-y|^2}{\delta}}, \quad r > 0, x,y \in \mathbb{R}^d,$$

where $\theta : [0,T] \times [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ is a measurable map. So, letting

$$h_t(y) := \nabla_{b_1^{(0)}(t,y)} u_t(y) + b_1^{(0)}(t,y), \tag{2.12}$$

we obtain

$$\begin{aligned} |\nabla_x^2 u_s(x)| &\le \int_s^T \frac{\mathrm{e}^{-\lambda(t-s)}}{t-s} \big|\nabla_x^2 P_{s,t}^{a_1,b_1^{(1)}}(h_t - h_t(z))(x)\big|_{z=\theta_{s,t}(x)} \mathrm{d}t \\ &\le \int_s^T \frac{\mathrm{e}^{-\lambda(t-s)}}{t-s} \mathrm{d}t \int_{\mathbb{R}^d} \big|\nabla_x^2 p_{s,t}^{a_1,b_1^{(1)}}(x,y)\big| \cdot |h_t(y) - h_t(\theta_{s,t}(x))| \mathrm{d}y. \end{aligned} \tag{2.13}$$

By (A_2), (2.9) for $\lambda \ge 1$, and (2.12), we have

$$|h_t(y) - h_t(\theta_{s,t}(x))| \le (1+c_1)|b_1^{(0)}(t,y) - b_1^{(0)}(t,\theta_{s,t}(x))| + K|\nabla u_t(y) - \nabla u_t(\theta_{s,t}(x))|. \tag{2.14}$$

In the following, we estimate these two terms respectively.

Since φ is concave, we find a constant $c_2 = c_2(K,T,d) > 0$ such that

$$\begin{aligned} &\int_{\mathbb{R}^d} |b_1^{(0)}(t,y) - b_1^{(0)}(t,\theta_{s,t}(x))| g_\delta(t-s,x,y)\mathrm{d}y \\ &\le K \int_{\mathbb{R}^d} \varphi(|y-\theta_{s,t}(x)|) g_\delta(t-s,x,y)\mathrm{d}y \\ &\le K\varphi\bigg(\int_{\mathbb{R}^d} |y-\theta_{s,t}(x)| g_\delta(t-s,x,y)\mathrm{d}y\bigg) \le c_2\varphi\Big(\sqrt{t-s}\Big), \quad 0 \le s < t \le T, x \in \mathbb{R}^d. \end{aligned}$$

Hence,

$$\begin{aligned} &\sup_{s\in[0,T]} \int_s^T \frac{\mathrm{e}^{-\lambda(t-s)}}{t-s} \mathrm{d}t \int_{\mathbb{R}^d} |b_1^{(0)}(t,y) - b_1^{(0)}(t,\theta_{s,t}(x))| g_\delta(t-s,x,y)\mathrm{d}y \\ &\le c_2 \int_0^T \frac{\mathrm{e}^{-\lambda t}\varphi(t^{\frac{1}{2}})}{t-s} \mathrm{d}t =: \varepsilon_1, \end{aligned} \tag{2.15}$$

where $\varepsilon_1 = \varepsilon_1(\lambda, K, T, d, \varphi)$ goes to 0 as $\lambda \to \infty$.

On the other hand, let $\alpha = 1 - \frac{d}{p} \in (0,1)$ and denote $z = \theta_{s,t}(x)$. By the Sobolev embedding theorem, there exists a constant $c_0 > 0$ depending on p and d such that

$$|\nabla u_t(y) - \nabla u_t(z)| \le c_0 |y - z|^\alpha \|1_{B(z,1)} \nabla^2 u_t\|_{L^p(\mathbb{R}^d)}, \quad \text{if } |y - z| < 1.$$

Since (2.9) implies $\|\nabla u_t\| \le c_1$ when $\lambda \ge 1$, we find a constant $c_3 = c_3(K, T, d) > 0$ such that

$$|\nabla u_t(y) - \nabla u_t(\theta_{s,t}(x))| \le c_3 |y - \theta_{s,t}(x)|^\alpha \|1_{B(z,1)} \nabla^2 u_t\|_{L^p(\mathbb{R}^d)}.$$

Noting that $\frac{d}{p} + \frac{2}{q} < 1$ and $\alpha = 1 - \frac{d}{p}$ imply $(1-\alpha)\frac{q}{q-1} < 1$, we find a constant $\varepsilon_2 = \varepsilon_2(\lambda, K, T, d, p, q) > 0$ which goes to 0 as $\lambda \to \infty$, such that

$$\begin{aligned}
&\int_s^T \frac{\mathrm{e}^{-\lambda(t-s)}}{t-s} \mathrm{d}t \int_{\mathbb{R}^d} |\nabla u_t(y) - \nabla u_t(\theta_{s,t}(x))| g_\delta(t-s, x, y) \mathrm{d}y \\
&\le c_3 \bigg(\int_s^T \mathrm{e}^{-\lambda(t-s)} (t-s)^{-(1-\alpha)\frac{q}{q-1}} \mathrm{d}t \bigg)^{\frac{q-1}{q}} \|\nabla^2 u\|_{\tilde{L}^p_q} \le \varepsilon_2, \quad s \in [0, T].
\end{aligned}$$

By (2.9), and combining this with (2.13), (2.14), and (2.15), we find large enough $\lambda = \lambda(K, T, P, \varphi) > 0$ such that $\|\nabla^2 u\|_{0 \to T, \infty} \le \frac{1}{2}$. Combining this with (2.9), we may choose large enough $\lambda > 0$ such that

$$\|u\|_{0 \to T, \infty} \vee \|\nabla u\|_{0 \to T, \infty} \vee \|\nabla^2 u\|_{0 \to T, \infty} \le \frac{1}{2}. \tag{2.16}$$

In particular, letting

$$\tilde{X}_t^{i, x_i} := X_t^{i, x_i} + u_t(X_t^{i, x_i}), \quad i = 1, 2, \tag{2.17}$$

we have

$$\frac{1}{2} |X_t^{1, x_1} - X_t^{2, x_2}| \le |\tilde{X}_t^{1, x_1} - \tilde{X}_t^{2, x_2}| \le 2 |X_t^{1, x_1} - X_t^{2, x_2}|. \tag{2.18}$$

Hence, to bound the exponential moment in (2.7), it suffices to estimate the corresponding term for $|\tilde{X}_{t_0}^{1, x_1} - \tilde{X}_{t_0}^{2, x_2}|^2$ replacing $|X_{t_0}^{1, x_1} - X_{t_0}^{2, x_2}|^2$.

(c) Let I_d be the $d \times d$ identity matrix. By (2.8), (2.17) and Itô's formula, we obtain

$$\begin{aligned}
\mathrm{d}\tilde{X}_t^{1, x_1} &= \big\{\lambda u_t + b_1^{(1)}(t, \cdot)\big\}(X_t^{1, x_1}) \mathrm{d}t + \big\{I_d + \nabla u_t(X_t^{1, x_1})\big\} \sigma_1(t, X_t^{1, x_1}) \mathrm{d}W_t, \\
\mathrm{d}\tilde{X}_t^{2, x_2} &= \big\{\lambda u_t + (L_t^{a_2, b_2} - L_t^{a_1, b_1}) u_t + (b_2 - b_1^{(0)})(t, \cdot)\big\}(X_t^{2, x_2}) \mathrm{d}t \\
&\qquad + \big\{I_d + \nabla u_t(X_t^{2, x_2})\big\} \sigma_2(t, X_t^{2, x_2}) \mathrm{d}W_t.
\end{aligned} \tag{2.19}$$

By (A_1), (2.16), (2.18), and Itô's formula, we find $k_1 = k_1(K, T, d, \varphi) > 0$ such that

$$\mathrm{d}|\tilde{X}_t^{1, x_1} - \tilde{X}_t^{2, x_2}|^2 \le k_1 \big(|\tilde{X}_t^{1, x_1} - \tilde{X}_t^{2, x_2}|^2 + \|a_1 - a_2\|_{t, \infty}^2 + \|b_1 - b_2\|_{t, \infty}^2 \big) \mathrm{d}t + \mathrm{d}M_t, \quad t \in [0, t_0], \tag{2.20}$$

where M_t is a martingale satisfying

$$\mathrm{d}\langle M \rangle_t \le k_1 |\tilde{X}_t^{1, x_1} - \tilde{X}_t^{2, x_2}|^2 \mathrm{d}t. \tag{2.21}$$

For any $n \geq 1$, let

$$\tau_n := t_0 \wedge \inf\big\{t \geq 0: \ |\tilde X_t^{1,x_1} - \tilde X_t^{2,x_2}| \geq n\big\}, \quad \gamma_n := \sup_{t\in[0,\tau_n]} |\tilde X_t^{1,x_1} - \tilde X_t^{2,x_2}|^2.$$

By (2.18) we have

$$|\tilde X_0^{1,x_1} - \tilde X_0^{2,x_2}|^2 \leq 4|x_1 - x_2|^2,$$

which together with (2.20), (2.21) and BDG's inequality implies that for some constant $k_2 = k_2(K, T, d, \varphi) > 1$,

$$\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1\gamma_n}{t_1}}\Big] \leq \mathrm{e}^{\frac{k_2}{t_1}\left[|x_1-x_2|^2+\int_0^{t_1}(\|a_1-a_2\|_{t,\infty}^2+\|b_1-b_2\|_{t,\infty}^2)\mathrm{d}t\right]}\Big(\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1k_2t_0\gamma_n}{t_1}}\Big]\Big)^{\frac{1}{2}}\Big(\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1k_2t_0\gamma_n}{t_1^2}}\Big]\Big)^{\frac{1}{2}}.$$

Taking $\varepsilon := \frac{1}{2k_2(1\vee T)}$, for any $t_0 := \varepsilon t_1$ and $t_1 \in (0, T]$ we have

$$(k_2 t_0) \vee \frac{k_2 t_0}{t_1} \leq \frac{1}{2},$$

so that

$$\begin{aligned}\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1\gamma_n}{t_1}}\Big] &\leq \mathrm{e}^{\frac{k_2}{t_1}\left[|x_1-x_2|^2+\int_0^{t_1}(\|a_1-a_2\|_{t,\infty}^2+\|b_1-b_2\|_{t,\infty}^2)\mathrm{d}t\right]}\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1\gamma_n}{2t_1}}\Big]\\ &\leq \mathrm{e}^{\frac{k_2}{t_1}\left[|x_1-x_2|^2+\int_0^{t_1}(\|a_1-a_2\|_{t,\infty}^2+\|b_1-b_2\|_{t,\infty}^2)\mathrm{d}t\right]}\Big(\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1\gamma_n}{t_1}}\Big]\Big)^{\frac{1}{2}}.\end{aligned}$$

Since γ_n is bounded, this implies

$$\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1\gamma_n}{t_1}}\Big] \leq \mathrm{e}^{\frac{2k_2}{t_1}\left[|x_1-x_2|^2+\int_0^{t_1}(\|a_1-a_2\|_{t,\infty}^2+\|b_1-b_2\|_{t,\infty}^2)\mathrm{d}t\right]}, \quad n \geq 1.$$

Therefore, by Fatou's lemma and (2.18),

$$\begin{aligned}\mathbb{E}\Big[\mathrm{e}^{\frac{2c_1|X_{t_0}^{1,x_1}-X_{t_0}^{2,x_2}|^2}{t_1}}\Big] &\leq \liminf_{n\to\infty}\mathbb{E}\Big[\mathrm{e}^{\frac{2c_1|X_{\tau_n}^{1,x_1}-X_{\tau_n}^{2,x_2}|^2}{t_1}}\Big]\\ &\leq \lim_{n\to\infty}\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1\gamma_n}{t_1}}\Big] \leq \mathrm{e}^{\frac{2k_2}{t_1}\left[|x_1-x_2|^2+\int_0^{t_1}(\|a_1-a_2\|_{t,\infty}^2+\|b_1-b_2\|_{t,\infty}^2)\mathrm{d}t\right]}.\end{aligned}$$

This together with (2.7) implies (2.4) for some constant $c = c(K, T, d, \varphi)$, and hence finishes the proof.

□

3 Proof of Theorem 1.1

By (2.3) and Proposition 2.2, to estimate $\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{2,x_2})$, we apply (1.1) to $\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{\langle t_0\rangle x_1})$. To this end, we present the following result.

Proposition 3.1. *Assume (A_1). Then the following assertions hold.*

(1) *There exists a constant* $c = c(K, T, d) > 0$ *such that*

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla p_s^{1,x}|^2}{p_s^{1,x}}(y)\mathrm{d}y \le c\log(1 + r^{-1}), \quad 0 < r \le t \le T, x \in \mathbb{R}^d. \tag{3.1}$$

(2) *If* $|b_1| \le C(K)$ *for some constant* $C(K) > 0$*, then for some constant* $c = c(K, T, d) > 0$*,*

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla p_s^{1,x}|^2}{p_s^{1,x}}(y)\mathrm{d}y \le c\log\Big(1 + \frac{t}{r}\Big), \quad 0 < r \le t \le T, x \in \mathbb{R}^d. \tag{3.2}$$

(3) *If* (1.5) *holds, then exists a constant* $c = c(K, T, d) > 0$ *such that*

$$\int_{\mathbb{R}^d} \frac{|\nabla p_t^{1,x}|^2}{p_t^{1,x}}(y)\mathrm{d}y \le \frac{c}{t}, \quad t \in (0, T], x \in \mathbb{R}^d. \tag{3.3}$$

In the following two subsections, we prove this result and Theorem 1.1 respectively.

3.1 Proof of Proposition 3.1

We first present a lemma.

Lemma 3.2. *Assume* (A_1) *with the condition on* $\|\nabla a_1\|_{0\to T,\infty}$ *replacing by the weaker one: there exists* $\beta \in (0, 1)$ *such that*

$$\|a_1(t, x) - a_1(t, y)\| \le K|x - y|^\beta, \quad t \in [0, T], x, y \in \mathbb{R}^d.$$

Then the following assertions hold.

(1) *There exists a constant* $c = c(K, T, d, \beta) > 0$ *such that*

$$\Big|\int_{\mathbb{R}^d} (p_t^{1,x} \log p_t^{1,x})(y)\mathrm{d}y\Big| \le c\log(1 + t^{-1}), \quad t \in (0, T], x \in \mathbb{R}^d. \tag{3.4}$$

(2) *If* $|b_1| \le C(K)$ *for some constant* $C(K) > 0$*, then*

$$\begin{aligned} &\Big|\int_{\mathbb{R}^d} (p_r^{1,x} \log p_r^{1,x})(y)\mathrm{d}y - \int_{\mathbb{R}^d} (p_t^{1,x} \log p_t^{1,x})(y)\mathrm{d}y\Big| \\ &\qquad \le c\log\Big(1 + \frac{t}{r}\Big), \quad 0 < r \le t \le T, x \in \mathbb{R}^d. \end{aligned} \tag{3.5}$$

Proof. (1) For any $x \in \mathbb{R}^d$, let $\theta_t(x)$ solve

$$\partial_t\theta_t(x) = b_1(t, \theta_t(x)), \quad \theta_0(x) = x, \quad t \in [0, T]. \tag{3.6}$$

By [12, Theorem 1.2], there exists a constant $c_0 = c_0(K, T, d) > 1$ such that

$$\frac{1}{c_0 t^{\frac{d}{2}}}\mathrm{e}^{-\frac{c_0|\theta_t(x)-y|^2}{t}} \le p_t^{1,x}(y) \le \frac{c_0}{t^{\frac{d}{2}}}\mathrm{e}^{-\frac{|\theta_t(x)-y|^2}{c_0 t}}, \quad x, y \in \mathbb{R}^d, t \in (0, T]. \tag{3.7}$$

Consequently,

$$\int_{\mathbb{R}^d} (p_t^{1,x} \log p_t^{1,x})(y)\mathrm{d}y \le \log[c_0 t^{-\frac{d}{2}}] \int_{\mathbb{R}^d} p_t^{1,x}(y)\mathrm{d}y = \log[c_0 t^{-\frac{d}{2}}], \quad t \in (0,T], x \in \mathbb{R}^d. \tag{3.8}$$

On the other hand, by Jensen's inequality and (3.7), we find a constant $c_1 = c_1(K,T,d) > 0$ such that

$$\begin{aligned} &- \int_{\mathbb{R}^d} (p_t^{1,x} \log p_t^{1,x})(y)\mathrm{d}y = 2\int_{\mathbb{R}^d} p_t^{1,x}(y) \log\{p_t^{1,x}(y)\}^{-\frac{1}{2}}\mathrm{d}y \\ &\le 2\log \int_{\mathbb{R}^d} \{p_t^{1,x}(y)\}^{\frac{1}{2}}\mathrm{d}y \le 2\log\Big[c_0^{\frac{1}{2}} t^{-\frac{d}{4}} (\pi c_0 t)^{\frac{d}{2}}\Big] \le \log[c_1 t^{\frac{d}{2}}]. \end{aligned}$$

This together with (3.8) implies (3.4).

(2) For any $0 < r \le t \le T$, we have

$$\begin{aligned} &I(r,t) := \int_{\mathbb{R}^d} (\rho_r \log \rho_r)(y)\mathrm{d}y - \int_{\mathbb{R}^d} (\rho_t \log \rho_t)(y)\mathrm{d}y = I_1(r,t) + I_2(r,t), \\ &I_1(r,t) := \int_{\mathbb{R}^d} \Big(\rho_r \log\frac{\rho_r}{\rho_t}\Big)(y)\mathrm{d}y, \quad I_2(r,t) := \int_{\mathbb{R}^d} (\rho_r - \rho_t)(y) \log \rho_t(y)\mathrm{d}y. \end{aligned} \tag{3.9}$$

If b_1 is bounded, then (3.6) implies

$$|\theta_t(x) - \theta_r(x)| \le c_1(t-r)$$

for some constant $c_1 > 0$, so that by (3.7), we find a constant $c_2 > 0$ such that

$$\begin{aligned} I_1(r,t) &\le \log\Big[c_0^2 \Big(\frac{t}{r}\Big)^{\frac{d}{2}}\Big] + \frac{c_0^2}{t} \int_{\mathbb{R}^d} |\theta_t(x) - y|^2 r^{-\frac{d}{2}} \mathrm{e}^{-\frac{|\theta_r(x)-y|^2}{c_0 r}}\mathrm{d}y \\ &\le c_2 \log\Big(1 + \frac{t}{r}\Big), \quad 0 < r \le t \le T. \end{aligned} \tag{3.10}$$

On the other hand, by (3.7), we find a constant $c_3 > 0$ such that

$$\begin{aligned} I_2(r,t) &= \int_{\mathbb{R}^d} \big\{(\rho_r - \rho_t)^+ \log \rho_t\big\}(y)\mathrm{d}y - \int_{\mathbb{R}^d} \big\{(\rho_r - \rho_t)^- \log \rho_t\big\}(y)\mathrm{d}y \\ &\le \int_{\mathbb{R}^d} \Big\{(\rho_r - \rho_t)^+(y) \log\big[c_0 t^{-\frac{d}{2}}\big] - (\rho_r - \rho_t)^-(y) \log\big[c_0^{-1} t^{-\frac{d}{2}}\big]\Big\}\mathrm{d}y \\ &\quad + \frac{c_0}{t} \int_{\mathbb{R}^d} (\rho_r - \rho_t)^-(y) |\theta_t(x) - y|^2 \mathrm{d}y \\ &\le \log[t^{-\frac{d}{2}}] \int_{\mathbb{R}^d} (\rho_r - \rho_t)(y)\mathrm{d}y + (\log c_0) \int_{\mathbb{R}^d} |\rho_r - \rho_t|(y)\mathrm{d}y \\ &\quad + \frac{c_0}{t} \int_{\mathbb{R}^d} |\theta_t(x) - y|^2 \rho_t(y)\mathrm{d}y \le c_3. \end{aligned}$$

Combining this with (3.9) and (3.10), we derive (3.5). □

Proof of Proposition 3.1. Let $x \in \mathbb{R}^d$ be fixed, and simply denote $\rho_t := p_t^{1,x}$.

(a) We first consider the smooth case where

$$\|\nabla^i b_1\|_{0\to T,\infty} + \|\nabla^i a_1\|_{0\to T,\infty} < \infty, \quad i \geq 1. \tag{3.11}$$

By [12, Theorem 1.2], there exist a constant $\lambda > 1$ and a measurable map $\theta : [0,T] \to \mathbb{R}^d$ such that

$$\lambda^{-1} t^{-\frac{d+i}{2}} \mathrm{e}^{-\frac{\lambda|\theta_t - y|^2}{t}} \leq \big|\nabla^i \rho_t\big|(y) \leq \lambda t^{-\frac{d+i}{2}} \mathrm{e}^{-\frac{|\theta_t - y|^2}{\lambda t}}, \quad t \in (0,T], y \in \mathbb{R}^d, i = 0,1,2. \tag{3.12}$$

Moreover, by the Kolmogorov forward equation and integration by parts formula, we have

$$\partial_t \rho_t = \mathrm{div}\Big[a_1(t,\cdot)\nabla\rho_t + \rho_t\{\mathrm{div} a_1(t,\cdot) - b_1(t,\cdot)\}\Big], \quad t \in (0,T]. \tag{3.13}$$

By (3.12), (3.13) and integration by parts formula, we obtain

$$\begin{aligned} &\int_{\mathbb{R}^d} \big\{\rho_t \log \rho_t - \rho_r \log \rho_r\big\}(y)\mathrm{d}y = \int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \big\{(1+\log\rho_s)\partial_s\rho_s\big\}(y)\mathrm{d}y \\ &= -\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle a_1(s,\cdot)\nabla\log\rho_s + \mathrm{div} a_1(s,\cdot) - b_1(s,\cdot), \nabla\rho_s\Big\rangle(y)\mathrm{d}y. \end{aligned} \tag{3.14}$$

Since $a_1 \geq K^{-1} I_d$, this implies

$$\begin{aligned} &\int_{\mathbb{R}^d} \big\{\rho_t \log \rho_t - \rho_r \log \rho_r\big\}(y)\mathrm{d}y + \frac{1}{K}\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla\rho_s|^2}{\rho_s}(y)\mathrm{d}y \\ &\leq -\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle \mathrm{div} a_1(s,\cdot) - b_1(s,\cdot), \nabla\rho_s\Big\rangle(y)\mathrm{d}y \\ &= \int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle \big[b_1^{(0)} - \mathrm{div} a_1\big](s,\cdot), \nabla\rho_s\Big\rangle(y)\mathrm{d}y + \int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle b_1^{(1)}(s,\cdot), \nabla\rho_s\Big\rangle(y)\mathrm{d}y. \end{aligned} \tag{3.15}$$

By (3.11), (3.12) and Lemma 3.2, we derive

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla\rho_s|^2}{\rho_s}(y)\mathrm{d}y < \infty. \tag{3.16}$$

Noting that (A_1) implies $|b_1^{(0)} - \mathrm{div} a_1| \leq 2K$, so that

$$\begin{aligned} &\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle \big[b_1^{(0)} - \mathrm{div} a_1\big](s,\cdot), \nabla\rho_s\Big\rangle(y)\mathrm{d}y \\ &\leq \frac{1}{2K}\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla\rho_s|^2}{\rho_s}(y)\mathrm{d}y + 2K^3 \int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \rho_s(y)\mathrm{d}y \\ &= \frac{1}{2K}\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla\rho_s|^2}{\rho_s}(y)\mathrm{d}y + 2K^3(t-r). \end{aligned}$$

Moreover, by the integration by parts formula, (3.12) and $\|\nabla b_1^{(1)}\|_{0\to T,\infty} \leq K$, we obtain

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle b_1^{(1)}(s,\cdot), \nabla\rho_s\Big\rangle(y)\mathrm{d}y = -\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \mathrm{div}\{b_1^{(1)}(s,y)\}\rho_s(y)\mathrm{d}y \leq K(t-r).$$

Combining these with (3.15) and (3.16), we derive

$$\begin{aligned}&\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla \rho_s|^2}{\rho_s}(y)\mathrm{d}y \\ &\le 2K \int_{\mathbb{R}^d} \big\{\rho_r \log \rho_r - \rho_t \log \rho_t\big\}(y)\mathrm{d}y + 2K^2(2K^2+1)(t-r).\end{aligned} \tag{3.17}$$

(b) In general, let $0 \le \psi \in C_0^\infty(\mathbb{R}^d)$ such that $\int_{\mathbb{R}^d} \psi(x)\mathrm{d}x = 1$, and define the smooth mollifier $\mathscr{S}_n$:

$$\mathscr{S}_n f(x) := n^d \int_{\mathbb{R}^d} f(x-y)\psi(ny)\mathrm{d}y, \quad n \ge 1, f \in L^1_{loc}(\mathbb{R}^d).$$

Let

$$b_1^{(n)}(t,\cdot) := \mathscr{S}_n b_1(t,\cdot), \quad a_1^{(n)}(t,\cdot) := \mathscr{S}_n a_1(t,\cdot), \quad n \ge 1.$$

Then $(a_1^{(n)}, b_1^{(n)})$ satisfies (3.11) and (A_1) for the same constant K. So, by step (a) and Lemma 3.2, the density function $\rho_t^{(n)}$ for the diffusion process generated by $L_t^{a_1^{(n)}, b_1^{(n)}}$ satisfies

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla \rho_s^{(n)}|^2}{\rho_s^{(n)}}(y)\mathrm{d}y \le c \log(1+r^{-1}), \quad 0 < r \le t \le T, n \ge 1 \tag{3.18}$$

for some constant $c = c(K,T,d) > 0$. Equivalently, for any

$$f \in C_0^{0,2}([r,t]\times\mathbb{R}^d) := \big\{f \in C_b([r,t]\times\mathbb{R}^d) : \ \nabla f, \nabla^2 f \in C_0([r,t]\times\mathbb{R}^d)\big\},$$

we have

$$\begin{aligned}&\bigg|\int_{[r,t]\times\mathbb{R}^d} \rho_s^{(n)}(y)\Delta f_s(y)\mathrm{d}s\mathrm{d}y\bigg|^2 = \bigg|\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \big\{\langle \nabla \log \rho_s^{(n)}, \nabla f_s\rangle \rho_s^{(n)}\big\}(y)\mathrm{d}y\bigg|^2 \\ &\le c\log(1+r^{-1}) \int_{[r,t]\times\mathbb{R}^d} |\nabla f_s|^2(y)\rho_s^{(n)}(y)\mathrm{d}s\mathrm{d}y, \quad n \ge 1.\end{aligned}$$

By [16, Theorem 11.1.4],

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} \rho_s^{(n)}(y)g(y)\mathrm{d}y = \int_{\mathbb{R}^d} \rho_s(y)g(y)\mathrm{d}y, \quad g \in C_b(\mathbb{R}^d), \quad s \in [r,t].$$

So, the above estimate implies

$$\bigg|\int_{[r,t]\times\mathbb{R}^d} \rho_s(y)\Delta f_s(y)\mathrm{d}s\mathrm{d}y\bigg|^2 \le c\log(1+r^{-1}) \int_{[r,t]\times\mathbb{R}^d} |\nabla f_s|^2(y)\rho_s(y)\mathrm{d}s\mathrm{d}y$$

for any $f \in C_0^{0,2}([r,t]\times\mathbb{R}^d)$. Therefore, (3.1) holds.

(c) If $|b_1| \le C(K)$ for some constat $C(K) > 0$, then (3.5) holds, so that instead of (3.18) we have

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla \rho_s^{(n)}|^2}{\rho_s^{(n)}}(y)\mathrm{d}y \le c \log\Big(1+\frac{t}{r}\Big), \quad 0 < r \le t \le T, n \ge 1.$$

Then the above argument implies (3.2).

(d) If (1.5) holds, then by Malliavin's calculus, see for instance [13] or [25, Remark 2.1], for any $v \in \mathbb{R}^d$ with $|v| = 1$, there exists a martingale $M_t^{1,x,v}$ such that

$$\mathbb{E}[\nabla_v f(X_t^{1,x})] = \mathbb{E}[f(X_t^{1,x}) M_t^{1,x,v}], \quad f \in C_b^1(\mathbb{R}^d), t \in (0,T]$$

and $\mathbb{E}[|M_t^{1,x,v}|^2] \le \frac{c}{t}$ holds for some constant $c = c(T,K,d) > 0$ and all $t \in (0,T]$. This implies

$$\bigg| \int_{\mathbb{R}^d} \big\{ \langle v, \nabla_x \log p_t^{1,x} \rangle f \big\}(y) p_t^{1,x}(y) \mathrm{d}y \bigg|^2 \le \frac{c}{t} \int_{\mathbb{R}^d} f(y)^2 p_t^{1,x}(y) \mathrm{d}y, \quad f \in C_b^1(\mathbb{R}^d),\ |v| = 1.$$

Equivalently,

$$\int_{\mathbb{R}^d} \frac{|\nabla p_t^{1,x}|^2}{p_t^{1,x}}(y) \mathrm{d}y \le \frac{cd}{t}, \qquad t \in (0,T],$$

so that (3.3) holds. □

3.2 Proof of Theorem 1.1

(1) Let $p > 1$ and $\varepsilon \in (0, \frac{1}{2}]$ be in Proposition 2.2. By Proposition (3.1) and (A_1), (H) holds for $\nu = \delta_{x_1}$ and $(a^{\langle t_0 \rangle}, b^{\langle t_0 \rangle})$ replacing (a_2, b_2). By (1.1) with $\nu = \delta_{x_1}$ and (3.1), we find a constant $c_1 = c_1(K,T,d,\varphi) > 0$ such that

$$\begin{aligned}
&\mathrm{Ent}(P_{t_1}^{1,x_1} | P_{t_1}^{\langle t_0 \rangle x_1}) \\
&\le c_1 \bigg[\log(1 + t_1^{-1}) \|a_1 - a_2\|_{\varepsilon t_1 \to t_1, \infty}^2 + \int_{\varepsilon t_1}^{t_1} \big(\|\mathrm{div}(a_1 - a_2)\|_{t,\infty}^2 + \|b_1 - b_2\|_t^2 \big) \mathrm{d}t \bigg], \\
&t_1 \in (0,T], x_1 \in \mathbb{R}^d.
\end{aligned} \tag{3.19}$$

Combining this with (2.3) and Proposition 2.2, we find a constant $c = c(K,T,d,\varphi) > 0$ such that for any $t_1 \in (0,T]$ and $x_1, x_2 \in \mathbb{R}^d$,

$$\begin{aligned}
\mathrm{Ent}(P_{t_1}^{1,x_1} | P_{t_1}^{2,x_2}) \le I_{t_1}(x_1,x_2) &:= \frac{c}{t_1} \bigg(|x_1 - x_2|^2 + \int_0^{t_1} \big\{ \|b_1 - b_2\|_{s,\infty}^2 + \|a_1 - a_2\|_{s,\infty}^2 \big\} \mathrm{d}s \bigg) \\
&\quad + c \bigg(\log(1 + t_1^{-1}) \|a_1 - a_2\|_{\varepsilon t_1 \to t_1, \infty}^2 + \int_{\varepsilon t_1}^{t_1} \|\mathrm{div}(a_1 - a_2)\|_{s,\infty}^2 \mathrm{d}s \bigg).
\end{aligned}$$

Equivalently, for any $t \in (0,T]$ and $f \in \mathscr{B}_b^+(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \big\{ \log f(y) \big\} P_t^{1,x_1}(\mathrm{d}y) \le \log \int_{\mathbb{R}^d} f(y) P_t^{2,x_2}(\mathrm{d}y) + I_t(x_1,x_2), \quad x_1, x_2 \in \mathbb{R}^d. \tag{3.20}$$

Let $\pi \in \mathscr{C}(\nu_1, \nu_2)$ such that

$$\mathbb{W}_2(\nu_1,\nu_2)^2 = \int_{\mathbb{R}^d \times \mathbb{R}^d} |x_1 - x_2|^2 \pi(\mathrm{d}x_1, \mathrm{d}x_2).$$

we obtain

$$\begin{aligned}
&\mathrm{Ent}(P_t^{1,\nu_1}|P_t^{2,\nu_2}) = \sup_{0<f\in\mathscr{B}_b(\mathbb{R}^d)}\bigg\{\int_{\mathbb{R}^d}\big\{\log f(y)\big\}P_t^{1,\nu_1}(\mathrm{d}y) - \log\int_{\mathbb{R}^d} f(y)P_t^{2,\nu_2}(\mathrm{d}y)\bigg\}\\
&\le \int_{\mathbb{R}^d\times\mathbb{R}^d} I_t(x_1,x_2)\pi(\mathrm{d}x_1,\mathrm{d}x_2)\\
&= \frac{c}{t}\bigg(\mathbb{W}_2(\nu_1,\nu_2)^2 + \int_0^t\big\{\|b_1-b_2\|_{s,\infty}^2 + \|a_1-a_2\|_{s,\infty}^2\big\}\mathrm{d}s\bigg)\\
&\quad + c\bigg(\log(1+t^{-1})\|a_1-a_2\|_{\varepsilon t\to t,\infty}^2 + \int_{\varepsilon t}^t \|\mathrm{div}(a_1-a_2)\|_{s,\infty}^2\mathrm{d}s\bigg).
\end{aligned}$$

Hence, (1.3) holds.

(2) Let $|b_1|\le C(K)$ for some constant $C(K)>0$. By (1.1), (3.2) and noting that $\frac{t_1}{t_0}=\varepsilon^{-1}$ for $t_0=\varepsilon t_1$, we find a constant $c_1=c_1(K,T,d,\varphi)>0$ such that instead of (3.19),

$$\begin{aligned}
&\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{\langle t_0\rangle x_1}) \le c_1\|a_1-a_2\|_{\varepsilon t_1\to t_1,\infty}^2 + c_1\int_{\varepsilon t_1}^{t_1}\big[\|\mathrm{div}(a_1-a_2)\|_{t,\infty}^2 + \|b_1-b_2\|_{t,\infty}^2\big]\mathrm{d}t,\\
&\quad t_1\in(0,T],\ x_1\in\mathbb{R}^d.
\end{aligned}$$

By repeating the above argument with this estimate replacing (3.19), we derive (1.4) for some constant $c=c(K,T,d,\varphi)>0$.

(3) Let (1.5) hold. By By (1.1), (3.3) and $t_0=\varepsilon t_1$, we find a constant $c_1=c_1(K,T,d,\varphi)>0$ such that for any $t_1\in(0,T]$ and $x_1\in\mathbb{R}^d$,

$$\begin{aligned}
&\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{\langle t_0\rangle x_1}) \le c_1\int_{\varepsilon t_1}^{t_1}\frac{1}{t}\|a_1-a_2\|_{t,\infty}^2\mathrm{d}t + c_1\int_{\varepsilon t_1}^{t_1}\big[\|\mathrm{div}(a_1-a_2)\|_{t,\infty}^2 + \|b_1-b_2\|_{t,\infty}^2\big]\mathrm{d}t,\\
&\le \frac{c_1}{\varepsilon t_1}\int_{\varepsilon t_1}^{t_1}\|a_1-a_2\|_{t,\infty}^2\mathrm{d}t + c_1\int_{\varepsilon t_1}^{t_1}\big[\|\mathrm{div}(a_1-a_2)\|_{t,\infty}^2 + \|b_1-b_2\|_{t,\infty}^2\big]\mathrm{d}t.
\end{aligned}$$

Then as explained above that using this estimate to replace (3.19), we derive (1.6) for some constant $c=c(K,T,d,\varphi)>0$.

4 Proof of Theorem 1.2

By (B), for any $\mu\in\mathscr{P}_2$, $b^\mu(t,x):=b(t,x,\mu)$ has decomposition $b^{0,\mu}+b^{1,\mu}$ such that $b^{1,\mu}$ is locally bounded and

$$|b^{0,\mu}|\vee\|\nabla b^{1,\mu}\|\le K.$$

Let $b^{(1)}:=b^{1,\delta_0}$, where δ_0 is the Dirac measure at 0, and let $b^{(0,\mu)}:=b^\mu-b^{(1)}$. Then (B) implies

$$|\nabla b^{(1)}|\le K,\quad |b^{(0,\mu)}|\le K+K\mu(|\cdot|^2)^{\frac12}.$$

This together with the the condition on σ included in (B) implies assumptions (A_0) and (A_1) in [9] for $k=2$. Therefore, by [9, Theorem 1.1], (1.7) is well-posed for distributions in $\mathscr{P}_2$, and there exists a constant $c>0$ such that

$$\sup_{t\in[0,T]}\mathbb{E}[|X_t|^2]\le c(1+\mathbb{E}[|X_0|^2])<\infty \tag{4.1}$$

holds for any solution with $\mathscr{L}_{X_0} \in \mathscr{P}_2$. So, it remains to verify (1.8).

For $\nu_i \in \mathscr{P}_2, i = 1, 2$, and $(t, x) \in [0, T] \times \mathbb{R}^d$, let

$$
\begin{aligned}
&a_i(t,x) := a(t,x,P_t^*\nu_i) = \frac{1}{2}(\sigma\sigma^*)(t,x,P_t^*\nu_i),\\
&b_i(t,x) := b(t,x,P_t^*\nu_i), \quad b_i^{(k)}(t,x) := b_i^{k,P_t^*\nu_i}(t,x), \quad k = 0,1.
\end{aligned} \tag{4.2}
$$

By Theorem 1.1, under (B), there exists a constant $c_1 = c_1(K, T, d, \varphi) > 0$ such that for any $t \in (0, T]$,

$$
\begin{aligned}
&\mathrm{Ent}(P_t^*\nu_1|P_t^*\nu_2) \le \frac{c_1}{t}\mathbb{W}_2(\nu_1,\nu_2)^2\\
&+ c_1\|b_1 - b_2\|_{t,\infty}^2 + c_1\log(1+t^{-1})\|a_1 - a_2\|_{t,\infty}^2 + c_1 t\|\mathrm{div}(a_1 - a_2)\|_{t,\infty}^2\\
&\le \frac{c_1}{t}\mathbb{W}_2(\nu_1,\nu_2)^2 + c_1K^2\big\{1 + \log(1+t^{-1}) + t\big\}\sup_{s\in[0,t]}\mathbb{W}_2(P_s^*\nu_1, P_s^*\nu_2)^2.
\end{aligned}
$$

Then there exists a constant $c_2 = c_2(K, T, d, \varphi) > 0$ such that

$$
\mathrm{Ent}(P_t^*\nu_1|P_t^*\nu_2) \le \frac{c_1}{t}\mathbb{W}_2(\nu_1,\nu_2)^2 + \frac{c_2}{t}\sup_{s\in[0,t]}\mathbb{W}_2(P_s^*\nu_1,P_s^*\nu_2)^2, \quad t \in (0,T].
$$

Combining this with the following result, we derive (1.8) for some constant $c > 0$, and hence finish the proof of Theorem 1.2.

Proposition 4.1. *Assume (B). Then there exists a constant $c > 0$ such that*

$$
\mathbb{W}_2(P_t^*\nu_1, P_t^*\nu_2) \le c\mathbb{W}_2(\nu_1,\nu_2), \quad t \in [0,T], \nu_1,\nu_2 \in \mathscr{P}_2.
$$

Proof. Let a_i and b_i be in (4.2), and let u_t be in (2.8) for large enough $\lambda > 0$ such that (2.16) holds. Let X_0^1, X_0^2 be $\mathscr{F}_0$-measurable such that

$$
\mathscr{L}_{X_0^i} = \nu_i, \quad i = 1,2, \quad \mathbb{E}[|X_0 - X_0^2|^2] = \mathbb{W}_2(\nu_1,\nu_2)^2. \tag{4.3}
$$

Let X_t^i solve (2.1) with initial value X_0^i. We have $\mathscr{L}_{X_t^i} = P_t^*\nu_i$, so that

$$
\mathbb{W}_2(P_t^*\nu_1, P_t^*\nu_2)^2 \le \mathbb{E}[|X_t^1 - X_t^2|^2], \quad t \in [0,T]. \tag{4.4}
$$

Let $\tilde{X}_t^i = X_t^i + u_t(X_t^i), i = 1, 2$. Then

$$
\frac{1}{2}|X_t^1 - X_t^2| \le |\tilde{X}_t^1 - \tilde{X}_t^2| \le 2|X_t^1 - X_t^2|, \quad t \in [0,T], \tag{4.5}
$$

and similarly to (2.19), by (2.8), (1.7) for X_t^i and Itô's formula, we have

$$
\begin{aligned}
\mathrm{d}\tilde{X}_t^1 &= \big\{\lambda u_t + b_1^{(1)}(t,\cdot)\big\}(X_t^1)\mathrm{d}t + \big\{I_d + \nabla u_t(X_t^1)\big\}\sigma_1(t,X_t^1)\mathrm{d}W_t,\\
\mathrm{d}\tilde{X}_t^2 &= \big\{\lambda u_t + (L_t^{a_2,b_2} - L_t^{a_1,b_1})u_t + (b_2 - b_1^{(0)})(t,\cdot)\big\}(X_t^2)\mathrm{d}t\\
&\quad + \big\{I_d + \nabla u_t(X_t^2)\big\}\sigma_2(t,X_t^2)\mathrm{d}W_t.
\end{aligned}
$$

Combining this with $(B)(1)$, (2.16), (4.3) and Itô's formula, we find $k_1 = k_1(K, T, d, \varphi) > 0$ such that

$$\mathrm{d}|\tilde{X}_t^1 - \tilde{X}_t^2|^2 \le k_1\big(|\tilde{X}_t^1 - \tilde{X}_t^2|^2 + \|a_1 - a_2\|_{t,\infty}^2 + \|b_1 - b_2\|_{t,\infty}^2\big)\mathrm{d}t + \mathrm{d}M_t, \quad t \in [0, T].$$

Noting that $(B)(3)$ and (4.2) imply

$$\|a_1 - a_2\|_{t,\infty}^2 + \|b_1 - b_2\|_{t,\infty}^2 \le 2K^2\xi_t, \quad \xi_t := \sup_{s\in[0,t]} \mathbb{W}_2(P_s^*\nu_1, P_s^*\nu_2)^2,$$

and due to (2.16), (4.3) and (4.4)

$$\mathbb{E}[|\tilde{X}_0^1 - \tilde{X}_0^2|^2] \le 4\mathbb{W}_2(\nu_1, \nu_2)^2, \quad \mathbb{E}[|\tilde{X}_t^1 - \tilde{X}_t^2|^2] \ge \frac{1}{4}\mathbb{E}[|X_t^1 - X_t^2|^2] \ge \frac{1}{4}\mathbb{W}_2(P_t^*\nu_1, P_t^*\nu_2)^2,$$

we find a constant $k_2 = k_2(K, T, d, \varphi) > 0$ such that

$$\xi_t \le k_2\mathbb{W}_2(\nu_1, \nu_2)^2 + k_2\int_0^t \xi_s \mathrm{d}s, \quad t \in [0, T].$$

Since (4.1) implies $\xi_t < \infty$, by Gronwall's inequality, this implies

$$\sup_{t\in[0,T]} \mathbb{W}_2(P_t^*\nu_1, P_t^*\nu_2)^2 = \xi_T \le k_2\mathrm{e}^{k_2 T}\mathbb{W}_2(\nu_1, \nu_2)^2.$$

So, the proof is finished. □

Acknowledgement. The authors would like to thank Professor Xing Huang for useful conversations and corrections. The work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project-ID317210226-SFB 1283.

References

[1] L. Ambrosio, N. Gigli, G. Savaré, *Bakry-Emery curvature-dimension condition and Riemannian Ricci curvature bounds,* Ann. Probab. 43(2015), 339–401.

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

[3] Y. Bai, X. Huang, *Log-Harnack inequality and exponential ergodicity for distribution dependent CKLS and Vasicek Model,* arXiv:2108.02623v5.

[4] V. I. Bogachev, N. V. Krylov, M. Röckner, S. V. Shaposhnikov, *Fokker-Planck-Kolmogorov equations,* American Math. Soc. 2015.

[5] 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.

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

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

[8] X. Huang, F.-Y. Wang, *Distribution dependent SDEs with singular coefficients,* Stochastic Process. Appl. 129(2019), 4747-4770.

[9] X. Huang, F.-Y. Wang, *Singular McKean-Vlasov (reflecting) SDEs with distribution dependent noise,* J. Math. Anal. Appl. 514(2022), 126301.

[10] X. Huang, F.-Y. Wang, *Log-Harnack inequality and Bismut formula for singular McKean-Vlasov SDEs,* arXiv:2207.11536.

[11] X. Huang, F.-Y. Wang, *Regularities and exponential ergodicity in entropy for SDEs driven by distribution dependent noise,* 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] D. Nualart, *The Malliavin Calculus and Related Topics,* Springer, 2006.

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

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

[16] D. W. Stroock, S.R.S. Varadhan, *Multidimensional Diffusion Processes,* Springer, Berlin, 1979.

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

[18] F.-Y. Wang, *Logarithmic Sobolev inequalities on noncompact Riemannian manifolds,* Probab. Theo. Relat. Fields 109(1997),417-424.

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

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

[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.

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

[23] F.-Y. Wang, *Distribution dependent reflecting stochastic differential equations,* to appear in Science in China Math. arXiv:2106.12737.

[24] F.-Y. Wang, *Derivative formula for singular McKean-Vlasov SDEs,* arXiv:2109.02030.

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

[26] P. Xia, L. Xie, X. Zhang, G. Zhao, *$L^q(L^p)$-theory of stochastic differential equations,* Stoch. Proc. Appl. 130(2020), 5188–5211.

[27] S.-Q. Zhang, C. Yuan, *A study on Zvonkin's transformation for stochastic differential equations with singular drift and related applications,* J. Diff. Equat. 297(2021), 277–319.