Bi-Coupling Method and Applications *

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

June 24, 2025

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 entropy-cost inequality is established for McKean-Vlasov SDEs with spatial-distribution dependent noise, which is open for a long time and has potential applications in optimal transport, information theory and mean field particle systems.

AMS subject Classification: 35A01, 35D35.
Keywords: Entropy-cost inequality, bi-coupling, diffusion process, McKean-Vlasov SDEs.

1 Introduction

The main purpose of this paper is to establish the entropy-cost inequality for McKean-Vlasov SDEs with spatial-distribution dependent noise, which has been open for a long time due to the essential difficulty caused by the distribution dependence of noise. To overcome this difficulty, we develop a new coupling argument, called bi-coupling, to cancel the short time singularity in the entropy upper bound for two diffusions presented in [8].

In this part, we first introduce the background of the study from applied areas including the information theory, optimal transport and mean field particle systems, then explain the main difficulty of the study, and finally figure out the main idea of the present study and the structure of the paper.

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

1.1 Background of the study

Let $\mathscr{P}_2$ be the space of probability measures on $\mathbb{R}^d$ having finite second moments, which is a Polish space under the quadratic Wasserstein distance

$$(1.1) \qquad \mathbb{W}_2(\mu,\nu) := \inf_{\pi\in\mathscr{C}(\mu,\nu)} \left(\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 \pi(\mathrm{d}x,\mathrm{d}y) \right)^{\frac{1}{2}}, \quad \mu,\nu\in\mathscr{P}_2,$$

where $\mathscr{C}(\mu,\nu)$ is the set of all couplings of μ and ν. In the theory of optimal transport, $\mathbb{W}_2$ refers to the optimal transportation cost induced by the quadratic cost function, and $(\mathscr{P}_2,\mathbb{W}_2)$ is called the Wasserstein space where a nice analysis and geometry structure has been developed, see for instance Otto's celebrated paper [20] and Villani's monograph [28].

In the information theory, the relative entropy functional describes the chaos of a distribution with respect to a reference measure, which refers to the difference of Shannon entropies for two distributions, and is known as the Kullback-Leibler divergence or the information divergence [17]. For two probability measures μ and ν, the relative entropy of ν with respect μ is defined as

$$\mathrm{Ent}(\nu|\mu) := \begin{cases} \int_{\mathbb{R}^d} \{\log \frac{\mathrm{d}\nu}{\mathrm{d}\mu}\} \mathrm{d}\nu, & \text{if } \frac{\mathrm{d}\nu}{\mathrm{d}\mu} \text{ exists,} \\ \infty, & \text{otherwise.} \end{cases}$$

Both Wasserstein distance and relative entropy have wide applications in applied areas including deep learning and Bayesian statistics [9]. When μ is the standard Gaussian measure on $\mathbb{R}^d$, Talagrand [27] found the beautiful inequality

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

where the constant 2 is sharp. This inequality was then extended in [21, 6] as

$$(1.2) \qquad \mathbb{W}_2(\nu,\mu)^2 \le C\,\mathrm{Ent}(\nu|\mu), \quad \nu\in\mathscr{P}_2$$

for a constant $C>0$ and a probability measure μ satisfying the log-Sobolev inequality

$$\mu(f^2\log f^2) := \int_{\mathbb{R}^d} f^2\log f^2 \mathrm{d}\mu \le C\mu(|\nabla f|^2), \quad f\in C_b^1(\mathbb{R}^d), \mu(f^2)=1.$$

The inequality (1.2) enables one to estimate the cost from above by using the entropy.

However, comparing with the Wasserstein distance $\mathbb{W}_2$, the entropy is usually harder to estimate from above. For instance, $\mathbb{W}_2$ between the distributions of two SDEs can be bounded by the expectation of the distance square of the solutions, which is easily derived using Itô's formula. But the entropy between solutions of SDEs is harder to estimate from above, since the heat kernels (distribution densities) are unknown. So, it is crucial to establish the inverse Talagrand inequality by bounding the entropy using $\mathbb{W}_2$.

In general, $\mathbb{W}_2(\mu,\nu)^2$ can not dominate $\mathrm{Ent}(\nu|\mu)$ since the former is finite for any $\mu,\nu\in\mathscr{P}_2$ but the latter becomes infinite when ν is not absolutely continuous with respect to μ. So, to derive an inverse Talagrand inequality, we consider the entropy between two stochastic systems for which the entropy decays in time according to the H-theorem in information theory. In this spirit, a sharp entropy-cost inequality was found by the second named author [30] for diffusion processes on a manifold M. According to [30, Theorem 1.1], for any constant $K \in \mathbb{R}$, the Bakry-Emery curvature of the diffusion process is bounded below by $K \in \mathbb{R}$ if and only if the following entropy-cost inequality holds:

$$\text{Ent}(P_t^*\mu|P_t^*\nu) \le \frac{K}{2(\mathrm{e}^{2Kt}-1)}\mathbb{W}_2^\rho(\mu,\nu)^2, \quad \mu,\nu \in \mathscr{P}_2(M), \quad t>0, \tag{1.3}$$

where $P_t^*\mu$ is the distribution of the diffusion process at time t with initial distribution μ, $\mathbb{W}_2^\rho$ is the quadratic Wasserstein distance induced by the Riemannian distance ρ on M (when $M=\mathbb{R}^d$ it reduces to $\mathbb{W}_2$ defined in (1.1)), and $\mathscr{P}_2(M)$ is the set of probability measures on M having finite second moment. This inequality has applications for both short and long times:

- For small time, (1.3) describes an instant finite property of the entropy, i.e. even though the initial entropy is infinite, the entropy at any time $t>0$ becomes finite, and the short time behavior of the entropy behaves like t^{-1};
- For long time, (1.3) provides exponential decay of P_t^* in entropy by using that in $\mathbb{W}_2$ which is easier to verify in applications.

The inequality (1.3) is equivalent to the log-Harnack inequality (see [31])

$$P_t \log f(x) \le \log P_t f(y) + \frac{K\rho(x,y)^2}{2(\mathrm{e}^{2Kt}-1)}, \quad x,y\in\mathbb{R}^d, t>0, f\in\mathscr{B}^+(M),$$

where $\mathscr{B}^+(M)$ is the space of all uniformly positive measurable functions on M, and $P_tf(x) := \int_{\mathbb{R}^d} f(y)\mathrm{d}(P_t^*\delta_x)$ is the associated diffusion semigroup. As a member in the family of dimension-free Hananck inequalities (see [29, 30, 32]), the log-Harnack inequality has crucial applications in optimal transport, curvature on Riemennian manifolds or metric measure spaces, see for instance [2, 24, 30, 31].

In this paper, we aim to establish the entropy-cost inequality of type (1.3) for the nonlinear Fokker-Planck equation on $\mathscr{P}_2$:

$$\partial_t\mu_t = (L_{t,\mu_t}^{a,b})^*\mu_t, \quad t\in[0,T], \tag{1.4}$$

where $T>0$ is a fixed time, and for any $(t,\mu)\in[0,T]\times\mathscr{P}_2$, $(L_{t,\mu}^{a,b})^*$ is the $L^2(\mathbb{R}^d)$-adjoint operator of

$$L_{t,\mu}^{a,b} := \sum_{i,j=1}^d a^{ij}(t,\cdot,\mu)\partial_i\partial_j + \sum_{i=1}^d b^i(t,\cdot,\mu)\partial_i. \tag{1.5}$$

Recall that a continuous map $\mu_\cdot : [0,T]\to\mathscr{P}_2$ is called a solution to (1.4), if for any $f\in C_0^\infty(\mathbb{R}^d)$ we have $\int_0^t |\mu_s(L_{s,\mu_s}f)|\mathrm{d}s<\infty$ and

$$\mu_t(f) = \mu_0(f) + \int_0^t \mu_s(L_{s,\mu_s}f)\mathrm{d}s, \quad t\in[0,T].$$

By the propagation of chaos, see [25], under reasonable conditions we have

$$\mu_t = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \delta_{X_t^{i,N}} \text{ in } L^2(\Omega\to\mathscr{P}_2;\mathbb{P}),$$

where for every $N\in\mathbb{N}$, $(X_t^{i,N})_{1\le i\le N}$ is the associated mean field particle system with N many particles, and μ_t is the distribution of the solution X_t to the following McKean-Vlasov SDE:

$$\text{d}X_t = b(t,X_t,\mathscr{L}_{X_t})\text{d}t + \sigma(t,X_t,\mathscr{L}_{X_t})\text{d}W_t,\quad t\in[0,T], \tag{1.6}$$

where $\mathscr{L}_{X_t}$ is the distribution of X_t, W_t is the d-dimensional Brownian motion under a standard probability base $(\Omega,\mathscr{F},\{\mathscr{F}_t\}_{t\in[0,T]},\mathbb{P})$, $\sigma := \sqrt{2a}$, and (a,b) comes from $L_{t,\mu}^{a,b}$ in (1.5). According to [5], under a mild integrability condition, (1.4) is well-posed in $\mathscr{P}_2$ if and only if (1.6) has weak well-posedness for distributions in $\mathscr{P}_2$, and in this case $\mu_t = P_t^*\mu := \mathscr{L}_{X_t^\mu}$ is the unique solution to the nonlinear Fokker-Planck equation (1.4) with $\mu_0=\mu$, where X_t^μ solves (1.6) with $\mathscr{L}_{X_0}=\mu$.

We intend to find a constant $c>0$ such that

$$\text{Ent}(P_t^*\mu|P_t^*\nu) \le \frac{c}{t}\mathbb{W}_2(\mu,\nu)^2,\quad t\in(0,T],\mu,\nu\in\mathscr{P}_2. \tag{1.7}$$

When the noise is distribution-free, i.e. $\sigma(t,x,\mu)=\sigma(t,x)$ does not depend on μ, (1.7) has already been derived and applied in the literature, (1.7) has been established in [13, 16, 24, 33, 35] under different conditions, see also [11, 12, 34] for extensions to the infinite-dimensional and reflecting models. When 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, (1.7) has been established in [15] by using a noise decomposition argument, see also [4] for the study on a special model.

However, when the noise is spatial-distribution dependent, this type inequality has been open for a long time until the new coupling technique (bi-coupling) is developed in the present paper, for which we construct a new diffusion process which is coupled with the other two processes respectively, see Section 2 below for details.

We would like to indicated that after an earlier version of this paper is available online (arXiv:2302.13500), the bi-coupling method has been applied in [22, 10] to different models to derive new estimates on entropy and probability distances, so that the efficient and originality of this new method has been illustrated.

1.2 Existing entropy inequality and difficulty of the present study

Noting that $P_t^*\mu$ is the distribution of the diffusion process X_t^μ generated by $(L_{t,P_t^*\mu}^{a,b})_{t\in[0,T]}$, the left hand side in (1.7) is the entropy between the distributions of two diffusion processes generated by $L_{t,P_t^*\mu}^{a,b}$ and $L_{t,P_t^*\nu}^{a,b}$ respectively. So, the study reduces to estimate the entropy between two different diffusion processes.

In general, 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

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

where $\mathscr{B}_b(\mathbb{R}^d)$ is the space of all bounded measurable functions on $\mathbb{R}^d$. 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 intend to estimate the relative entropy

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

for $t\in(0,T]$ and $\nu_1,\nu_2\in\mathscr{P}_2$. Before moving on, let us recall a nice entropy inequality derived by Bogachev, Röckner and Shaposhnikov [8]. 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 [8, Theorem 1.1], the entropy inequality

$$\mathrm{Ent}(P_t^{1,\nu}|P_t^{2,\nu})\le\frac{1}{2}\int_0^t\mathbb{E}\big[\big|a_2(s,X_s^{1,\nu})^{-\frac{1}{2}}\Phi^\nu(s,X_s^{1,\nu})\big|^2\big]\mathrm{d}s,\quad t\in(0,T]\tag{1.8}$$

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 least 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 [7].

As observed in [8, 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, (1.8) provides a non-trivial upper bound for $\mathrm{Ent}(P_t^{1,\nu}|P_t^{2,\nu})$.

However, when $X_s^{1,x}$ is the standard Brownian motion starting from a fixed initial value x, i.e. $\nu=\delta_x$, we have

$$\mathbb{E}[|\nabla\log p_s^{1,x}|^2(X_s^{1,x})]=\frac{1}{s^2}\mathbb{E}|X_s^{1,x}-x|^2=\frac{1}{s}.$$

So, for elliptic diffusions the best possible short time estimate on $\mathbb{E}[|\nabla\log p_s^{1,x}|^2(X_s^{1,x})]$ behaves like s^{-1}, 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 (1.8) becomes trivial when

$$\inf_{(s,x)\in[0,T]\times\mathbb{R}^d}\|a_1(s,x)-a_2(s,x)\|>0. \tag{1.9}$$

So, the key point of the present study is to cancel the small time singularity in (1.8), which stimulates us to develop a new coupling method, i.e. the bi-coupling method in Section 2 below.

1.3 Main idea and structure of the paper

To kill the singularity in (1.8) for small $t>0$, in Section 2 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. In Section 3 we apply the bi-coupling to estimate the entropy between two diffusion processes, and as an application, in Section 4 we establish the entropy-cost inequality (1.7) for the McKean-Vlasov SDE (1.6).

To measure the singularity/regularity of coefficients in (1.6), we introduce the following class of Dini functions

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

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

$$\|f\|_{t,\infty} := \sup_{x \in \mathbb{R}^d} |f(t,x)|, \quad \|f\|_{r \to t,\infty} := \sup_{s \in [r,t]} \|f\|_{s,\infty}, \quad r \in [0,t],$$
$$\|f\|_{0 \to T,\varphi} := \sup_{t \in [0,T], x \neq y \in \mathbb{R}^d} \left(|f(t,x)| + \frac{|f(t,x) - f(t,y)|}{\varphi(|x-y|)} \right).$$

In the following, $c = c(K, T, d, \varphi)$ stands for a constant depending only on K, T, d and φ given in (A_1) and (A_2).

2 Bi-coupling method and density estimates

Let $\sigma_i = \sqrt{2a_i}, i = 1, 2$. Consider SDEs:

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

We make the following assumptions (A_1) and (A_2) where b_i may have a Dini continuous term with respect to some $\varphi \in \mathscr{D}$.

(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 $\|b_i^{(0)}\|_{0 \to T,\varphi} \le K$.

According to [23, Theorem 2.1], (A_1) implies the well-posedness of (2.1). For any $s \in [0,T)$ and $x \in \mathbb{R}^d$, let $X_{s,t}^{i,x}$ be the unique solution of (2.1) 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 $X_t^{i,x_i} := X_{0,t}^{i,x_i}$ solve (2.1) for $X_0^{i,x_i} = x_i$. 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

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

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 (1.8) 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 entropy inequality (2.3), we derive the desired upper bound on $\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{2,x_2})$. We remark that (2.3) has been presented in [12] for $p = 2$, but in the present study we shall need the inequality for $p > 2$ as required in the dimension-free Harnack inequality due to [23], see the proof of Proposition 2.3 for details.

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

$$\mathrm{Ent}(\mu_1|\mu_2) \le p\mathrm{Ent}(\mu_1|\mu) + (p-1)\log \int_E \Big(\frac{\mathrm{d}\mu}{\mathrm{d}\mu_2}\Big)^{\frac{p}{p-1}} \mathrm{d}\mu_2, \tag{2.3}$$

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 [3, 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

$$\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{2,x_2}) \le p\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{\langle t_0\rangle x_1}) + (p-1)\log\int_{\mathbb{R}^d}\bigg(\frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\bigg)^{\frac{p}{p-1}}\mathrm{d}P_{t_1}^{2,x_2}. \tag{2.4}$$

Noting that $a(t,\cdot) - a_1(t,\cdot) = 0$ for $t \in [0, t_0]$, we may apply (1.8) to derive a non-trivial upper bound on the first term in the right hand side of (2.4), see Proposition 3.2 for details. So, in the following, we only estimate the second term. To this end, we need the following simple lemma.

Lemma 2.2. *Let $\xi_t \geq 0$ be a continuous semi-martingale such that*

$$\mathrm{d}\xi_t \leq k_1 \xi_t \mathrm{d}t + \mathrm{d}A_t + \mathrm{d}M_t, \quad t \in [0,T],$$

where $k_1 > 0$ is a constant, A_t is an increasing function with $A_0 = 0$, and M_t is a local martingale with

$$\mathrm{d}\langle M \rangle_t \leq k_1 \xi_t \mathrm{d}t.$$

Then for any $t_0 \in (0, T \wedge k_1^{-1})$ and constants $\lambda, k > 0$ such that

$$k(1 - k_1 t_0) \geq k_1 \Big(1 + \frac{\lambda}{2}\Big), \tag{2.5}$$

we have

$$\mathbb{E} \exp\Big[\frac{\lambda \xi_{t_0}}{1 + k t_0}\Big] \leq \exp\big[\lambda \xi_0 + \lambda A_{t_0}\big].$$

Proof. Let $\eta_t := \exp\big[\frac{\lambda \xi_t}{1+kt}\big]$. By Itô's formula, we find a local martingale $\tilde{M}_t$ such that

$$\begin{aligned}\mathrm{d}\eta_t &= \eta_t \Big\{\frac{\lambda}{1+kt}\mathrm{d}\xi_t + \frac{\lambda^2}{2(1+kt)^2}\mathrm{d}\langle M \rangle_t - \frac{k\lambda \xi_t}{(1+kt)^2}\mathrm{d}t\Big\} + \mathrm{d}\tilde{M}_t \\ &\leq \eta_t \xi_t \Big\{\frac{\lambda k_1}{1+kt} + \frac{\lambda^2 k_1}{2(1+kt)^2} - \frac{k\lambda}{(1+kt)^2}\Big\}\mathrm{d}t + \lambda \eta_t \mathrm{d}A_t + \mathrm{d}\tilde{M}_t, \quad t \in [0,T].\end{aligned}$$

By (2.5) we have

$$\frac{\lambda k_1}{1+kt} + \frac{\lambda^2 k_1}{2(1+kt)^2} - \frac{k\lambda}{(1+kt)^2} \leq 0, \quad t \in [0, t_0],$$

so that

$$\mathrm{d}\eta_t \leq \lambda \eta_t \mathrm{d}A_t + \mathrm{d}\tilde{M}_t, \quad t \in [0, t_0].$$

By Gronwall's lemma, this implies

$$\mathbb{E}[\eta_{t_0}] \leq \eta_0 \mathrm{e}^{\lambda A_{t_0}},$$

which coincides with the desired estimate.

□

Proposition 2.3. *Assume (A_1) and (A_2). Then there exist constants $p = p(K,T,d) > 2, \varepsilon = \varepsilon(K,T,d) \in (0, \frac{1}{2}]$ 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 $t_0 = \varepsilon t_1$,*

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

Proof. (a) Recall that $\mathscr{B}_b(\mathbb{R}^d)$ is the space of all bounded measurable functions on $\mathbb{R}^d$, and let

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

Then the desired estimate follows from the inequality

$$\begin{aligned}&\big|P_{t_1}^{\langle t_0 \rangle} f(x_1)\big|^p \leq \big(P_{t_1}^{(2)} |f|^p(x_2)\big) \\ &\times \exp\Big[\frac{c(p-1)}{t_1}\Big(|x_1 - x_2|^2 + \int_0^{t_1} \big\{\|a_1 - a_2\|_{t,\infty}^2 + \|b_1 - b_2\|_{t,\infty}^2\big\}\mathrm{d}t\Big)\Big], \quad f \in \mathscr{B}_b(\mathbb{R}^d).\end{aligned} \tag{2.6}$$

Indeed, taking $f := \big(n \wedge \frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\big)^{\frac{1}{p-1}}$ for $n \geq 1$, this inequality implies

$$\begin{aligned}&\bigg(\int_{\mathbb{R}^d}\Big(n\wedge \frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\Big)^{\frac{p}{p-1}}\mathrm{d}P_{t_1}^{2,x_2}\bigg)^p \leq \bigg(\int_{\mathbb{R}^d}\Big(n\wedge \frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\Big)^{\frac{1}{p-1}}\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}\bigg)^p\\
&\leq \bigg(\int_{\mathbb{R}^d}\Big(n\wedge \frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\Big)^{\frac{p}{p-1}}\mathrm{d}P_{t_1}^{2,x_2}\bigg)\\
&\quad\times \exp\bigg[\frac{c(p-1)}{t_1}\bigg(|x_1-x_2|^2+\int_0^{t_1}\big\{\|a_1-a_2\|_{t,\infty}^2+\|b_1-b_2\|_{t,\infty}^2\big\}\mathrm{d}t\bigg)\bigg].\end{aligned}$$

Taking log in both sides we derive

$$\begin{aligned}&\log\int_{\mathbb{R}^d}\Big(n\wedge \frac{\mathrm{d}P_{t_1}^{\langle t_0\rangle x_1}}{\mathrm{d}P_{t_1}^{2,x_2}}\Big)^{\frac{p}{p-1}}\mathrm{d}P_{t_1}^{2,x_2}\\
&\leq \frac{c}{t_1}\bigg(|x_1-x_2|^2+\int_0^{t_1}\big\{\|a_1-a_2\|_{t,\infty}^2+\|b_1-b_2\|_{t,\infty}^2\big\}\mathrm{d}t\bigg),\quad n\geq 1,\end{aligned}$$

which implies the desired estimate as $n \to \infty$. So, it remains to find constants $p > 2$ and $c > 0$ such that (2.6) holds.

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

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

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

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

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

$$P_{t_1}^{\langle t_0\rangle}f(x_1) = \mathbb{E}\big[(P_{t_0,t_1}^{(2)}f)(X_{t_0}^{1,x_1})\big],\quad P_{t_1}^{(2)}f(x_2)=\mathbb{E}\big[(P_{t_0,t_1}^{(2)}f)(X_{t_0}^{2,x_2})\big]. \tag{2.7}$$

By [23, Theorem 2.2] which applies to a more general setting where $b_2^{(0)}$ only satisfies a local integrability condition, and noting that $t_1 - t_0 = (1-\varepsilon)t_1$, we find constants $p_1 = p_1(K,T,d) > 1 \vee \frac{d}{2}$ and $c_1 = c_1(K,T,d,\varepsilon) > 0$ such that

$$\big|P_{t_0,t_1}^{(2)}f(x)\big|^{p_1} \leq \big(P_{t_0,t_1}^{(2)}|f|^{p_1}(y)\big)\mathrm{e}^{\frac{c_1|x-y|^2}{t_1}},\quad f\in\mathscr{B}_b(\mathbb{R}^d), x,y\in\mathbb{R}^d. \tag{2.8}$$

Combining this with (2.7) and Jensen's inequality, for $p := 2p_1 > 2 \vee d$ we obtain

$$\begin{aligned}&\big|P_{t_1}^{\langle t_0\rangle}f(x_1)|^p = \big|\mathbb{E}[P_{t_0,t_1}^{(2)}f(X_{t_0}^{1,x_1})]\big|^{2p_1} \leq \Big(\mathbb{E}\big[|P_{t_0,t_1}^{(2)}f|^{p_1}(X_{t_0}^{1,x_1})\big]\Big)^2\\
&\leq \bigg\{\mathbb{E}\Big[\big(P_{t_0,t_1}^{(2)}|f|^{p_1}(X_{t_0}^{2,x_2})\big)\exp\Big(\frac{c_1|X_{t_0}^{1,x_1}-X_{t_0}^{2,x_2}|^2}{t_1}\Big)\Big]\bigg\}^2\\
&\leq \big(\mathbb{E}\big[P_{t_0,t_1}^{(2)}|f|^{2p_1}(X_{t_0}^{2,x_2})\big]\big)\mathbb{E}\Big[\exp\Big(\frac{2c_1|X_{t_0}^{1,x_1}-X_{t_0}^{2,x_2}|^2}{t_1}\Big)\Big]\\
&= \big(P_{t_1}^{(2)}|f|^p(x_2)\big)\mathbb{E}\Big[\exp\Big(\frac{2c_1|X_{t_0}^{1,x_1}-X_{t_0}^{2,x_2}|^2}{t_1}\Big)\Big].\end{aligned} \tag{2.9}$$

Thus, to prove (2.6), it remains to estimate the expectation term in the upper bound.

(b) Since the exponential term is symmetric in $(X_{t_0}^{1,x_1}, X_{t_0}^{2,x_2})$, 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 (A_1), $b_1^{(0)}$ is bounded, and noting that $p := 2p_1 > 2 \vee d$, for a fixed constant $q > 2$ such that $\frac{d}{p} + \frac{2}{q} < 1$, we have $\|b_1^{(0)}\|_{\tilde L_q^p} < \infty$. So, according to [38, Theorem 2.1], 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

$$(\partial_t + L_t^{a_1,b_1} - \lambda)u_t = -b_1^{(0)}(t,\cdot), \quad t \in [0,T], u_T = 0 \tag{2.10}$$

has a unique solution satisfying

$$\lambda^\beta(\|u\|_{0\to T,\infty} + \|\nabla u\|_{0\to T,\infty}) + \|\partial_t u\|_{\tilde L_q^p} + \|\nabla^2 u\|_{\tilde L_q^p} \le c_1, \tag{2.11}$$

where for any measurable function g on $[0,T] \times \mathbb{R}^d$,

$$\|g\|_{\tilde L_q^p} := \sup_{z\in\mathbb{R}^d} \bigg(\int_0^T \|1_{B(z,1)} g(t,\cdot)\|_{L^p(\mathbb{R}^d)}^q \mathrm{d}t \bigg)^{\frac{1}{q}}. \tag{2.12}$$

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

$$u_s = \int_s^T \mathrm{e}^{-\lambda(t-s)} P_{s,t}^{a_1,b_1^{(1)}} \big\{ \nabla_{b_1^{(0)}} u_t + b_1^{(0)}(t,\cdot) \big\} \mathrm{d}t, \quad s \in [0,T]. \tag{2.13}$$

Let ∇_x^2 be the Hessian operator in x. By [18, 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

$$g_\delta(r,x,y) := (\pi\delta r)^{-\frac{d}{2}} \mathrm{e}^{-\frac{|\theta_{s,t}(x)-y|^2}{\delta r}}, \quad r > 0, x, y \in \mathbb{R}^d, \tag{2.14}$$

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

$$h_t(y) := (b_1^{(0)}(t,y) \cdot \nabla) u_t(y) + b_1^{(0)}(t,y), \tag{2.15}$$

and denoting by (∇_x, ∇_x^2) the gradient and Hessian operators in $x \in \mathbb{R}^d$, we obtain

$$\begin{aligned} |\nabla_x^2 u_s(x)| &\le \int_s^T \frac{\mathrm{e}^{-\lambda(t-s)}}{t-s} \big| \nabla_x^2 P_{s,t}^{a_1,b_1^{(1)}} (h_t - h_t(z))(x) \big|_{z=\theta_{s,t}(x)} \mathrm{d}t \\ &\le \int_s^T \frac{\mathrm{e}^{-\lambda(t-s)}}{t-s} \mathrm{d}t \int_{\mathbb{R}^d} \big| \nabla_x^2 p_{s,t}^{a_1,b_1^{(1)}}(x,y) \big| \cdot |h_t(y) - h_t(\theta_{s,t}(x))| \mathrm{d}y. \end{aligned} \tag{2.16}$$

By (A_2), (2.11) for $\lambda \geq 1$, and (2.15), we have

$$|h_t(y) - h_t(\theta_{s,t}(x))| \leq (1 + c_1)|b_1^{(0)}(t, y) - b_1^{(0)}(t, \theta_{s,t}(x))| + K|\nabla u_t(y) - \nabla u_t(\theta_{s,t}(x))|. \tag{2.17}$$

In the following, we estimate these two terms in the upper bound respectively.

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

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

Hence,

$$\begin{aligned}&\sup_{s\in[0,T]} \int_s^T \frac{\mathrm{e}^{-\lambda(t-s)}}{t - s} \mathrm{d}t \int_{\mathbb{R}^d} |b_1^{(0)}(t, y) - b_1^{(0)}(t, \theta_{s,t}(x))| g_\delta(t - s, x, y) \mathrm{d}y \\ &\leq c_2 \int_0^T \frac{\mathrm{e}^{-\lambda t} \varphi(t^{\frac{1}{2}})}{t} \mathrm{d}t =: \varepsilon_1,\end{aligned} \tag{2.18}$$

where $\varepsilon_1 = \varepsilon_1(\lambda, K, T, d, \varphi)$. Since $\varphi \in \mathscr{D}$ implies

$$\int_0^T \frac{\varphi(t^{\frac{1}{2}})}{t} \mathrm{d}t = 2 \int_0^{T^{\frac{1}{2}}} \frac{\varphi(s)}{s} \mathrm{d}s < \infty,$$

by the dominated convergence theorem we derive $\lim_{\lambda\to\infty} \varepsilon_1 = 0$.

On the other hand, let $\alpha = 1 - \frac{d}{p} \in (0, 1)$. By the Sobolev embedding theorem, see e.g. [1], there exists a constant $c_0 > 0$ depending on p and d such that

$$\sup_{z\neq y\in B(z,1)} \frac{|f(y) - f(z)|}{|y - z|^\alpha} \leq c_0 \|1_{B(z,1)}(|f| + |\nabla f|)\|_{L^p}, \quad z \in \mathbb{R}^d, \quad f \in W_{loc}^{1,p}(\mathbb{R}^d).$$

So,

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

Noting that $\frac{d}{p} + \frac{2}{q} < 1$ and $\alpha = 1 - \frac{d}{p}$ imply $(1 - \alpha)\frac{q}{q-1} < 1$, by combining this with (2.11) and (2.14), we find constants $c_3 = c_3(p, d) > 0$ and $\varepsilon_2 = \varepsilon_2(\lambda, K, T, d, p, q) > 0$, where $\varepsilon_2 \to 0$ as $\lambda \to \infty$, such that

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

By (2.11), and combining this with (2.16), (2.17), and (2.18), 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 (2.11), we may choose large enough $\lambda > 0$ such that

$$\|u\|_{0\to T,\infty} \vee \|\nabla u\|_{0\to T,\infty} \vee \|\nabla^2 u\|_{0\to T,\infty} \le \frac{1}{2}. \tag{2.19}$$

In particular, letting

$$\tilde{X}_t^{i,x_i} := X_t^{i,x_i} + u_t(X_t^{i,x_i}), \quad i = 1,2, \tag{2.20}$$

we have

$$\frac{1}{2}|X_t^{1,x_1} - X_t^{2,x_2}| \le |\tilde{X}_t^{1,x_1} - \tilde{X}_t^{2,x_2}| \le 2|X_t^{1,x_1} - X_t^{2,x_2}|. \tag{2.21}$$

Hence, to bound the exponential moment in (2.9), 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 (2.10), (2.20) and Itô's formula, we obtain

$$\begin{aligned} \mathrm{d}\tilde{X}_t^{1,x_1} &= \big\{\lambda u_t + b_1^{(1)}(t,\cdot)\big\}(X_t^{1,x_1})\mathrm{d}t + \big\{I_d + \nabla u_t(X_t^{1,x_1})\big\}\sigma_1(t,X_t^{1,x_1})\mathrm{d}W_t,\\ \mathrm{d}\tilde{X}_t^{2,x_2} &= \big\{\lambda u_t + (L_t^{a_2,b_2} - L_t^{a_1,b_1})u_t + (b_2 - b_1^{(0)})(t,\cdot)\big\}(X_t^{2,x_2})\mathrm{d}t\\ &\qquad + \big\{I_d + \nabla u_t(X_t^{2,x_2})\big\}\sigma_2(t,X_t^{2,x_2})\mathrm{d}W_t. \end{aligned} \tag{2.22}$$

By (A_1), (2.19), (2.21), and Itô's formula, we find $k_1 = k_1(K,T,d,\varphi) > 0$ such that

$$\mathrm{d}|\tilde{X}_t^{1,x_1} - \tilde{X}_t^{2,x_2}|^2 \le k_1|\tilde{X}_t^{1,x_1} - \tilde{X}_t^{2,x_2}|^2\mathrm{d}t + \mathrm{d}A_t + \mathrm{d}M_t, \quad t\in[0,t_0], \tag{2.23}$$

where

$$A_t := k_1\int_0^t \big(\|a_1 - a_2\|_{s,\infty}^2 + \|b_1 - b_2\|_{s,\infty}^2\big)\mathrm{d}s, \tag{2.24}$$

and M_t is a martingale satisfying

$$\mathrm{d}\langle M\rangle_t \le k_1|\tilde{X}_t^{1,x_1} - \tilde{X}_t^{2,x_2}|^2\mathrm{d}t. \tag{2.25}$$

For any $n\ge 1$, let

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

By (2.21) we have

$$|\tilde{X}_0^{1,x_1} - \tilde{X}_0^{2,x_2}|^2 \le 4|x_1 - x_2|^2. \tag{2.26}$$

Moreover, to apply Lemma 2.2, let

$$t_0 := \frac{t_1}{2[(Tk_1 + 4k_1c_1)\vee 1]}, \quad \lambda := \frac{8c_1(1+kt_0)}{t_1}, \quad k = \frac{k_1}{1-k_1t_0}\Big(1 + \frac{\lambda}{2}\Big),$$

so that (2.5) holds and

$$\frac{\lambda}{1+kt_0} = \frac{8c_1}{t_1}.$$

Combining this with (2.23)-(2.26), we may apply Lemma 2.2 for $\xi_t = |\tilde X_t^{1,x_1} - \tilde X_t^{2,x_2}|^2$ to find a constant $k_2 = k_2(K,T,d,\varphi) > 0$ such that

$$\mathbb{E}\Big[\mathrm{e}^{\frac{8c_1}{t_1}|\tilde X_{t_0}^{1,x_1} - \tilde X_{t_0}^{2,x_2}|^2}\Big] \le \mathrm{e}^{\frac{k_2}{t_1}\left(|x_1-x_2|^2 + \int_0^{t_0}(\|a_1-a_2\|_{t,\infty}^2 + \|b_1-b_2\|_{t,\infty}^2)\mathrm{d}t\right)}.$$

This together with (2.9) implies (2.6) for some constant $c = c(K,T,d,\varphi)$, and hence finishes the proof.

□

3 Entropy estimates between two diffusion processes

With the bi-coupling method and density estimates addressed in Section 2, we are able to prove the following result on entropy upper bound estimates for diffusion processes with arbitrary initial distributions in $\mathscr{P}_2$, for which the existing estimates may be invalid as explained in Section 1.2.

Theorem 3.1. *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 $t \in (0,T]$,*

$$\begin{aligned}\mathrm{Ent}(P_t^{1,\nu_1}|P_t^{2,\nu_2}) \le &\frac{c\mathbb{W}_2(\nu_1,\nu_2)^2}{t} + \frac{c}{t}\int_0^t \big\{\|b_1-b_2\|_{s,\infty}^2 + \|a_1-a_2\|_{s,\infty}^2\big\}\mathrm{d}s \\ &+ c\Big[\|a_1-a_2\|_{\varepsilon t\to t,\infty}^2 + \int_{\varepsilon t}^t \|\mathrm{div}(a_1-a_2)\|_{s,\infty}^2\mathrm{d}s\Big].\end{aligned} \tag{3.1}$$

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

$$\|\nabla^i b_1\|_{0\to T,\infty} + \|\nabla^i a_1\|_{0\to T,\infty} \le C(K), \quad i = 1,2, \tag{3.2}$$

then for any $\nu_1, \nu_2 \in \mathscr{P}$ and $t \in (0,T]$,

$$\begin{aligned}\mathrm{Ent}(P_t^{1,\nu_1}|P_t^{2,\nu_2}) \le &\frac{c}{t}\Big[\mathbb{W}_2(\nu_1,\nu_2)^2 + \int_0^t \big(\|b_1-b_2\|_{s,\infty}^2 + \|a_1-a_2\|_{s,\infty}^2\big)\mathrm{d}s\Big] \\ &+ \int_{\varepsilon t}^t \|\mathrm{div}(a_1-a_2)\|_{s,\infty}^2\mathrm{d}s.\end{aligned} \tag{3.3}$$

To prove Theorem 3.1, we shall apply (2.4), where the second term in the upper bound has been estimated in Proposition 2.3, and the first term will be estimated by using (1.8) and the following result.

Proposition 3.2. *Assume (A_1). Then the following assertions hold.*

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

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla p_s^{1,x}|^2}{p_s^{1,x}}(y)\mathrm{d}y \le c \log\Big(1 + \frac{t}{r}\Big), \quad 0 < r \le t \le T, x \in \mathbb{R}^d. \tag{3.4}$$

(2) *If* (3.2) *holds, then exists a constant* $c = c(K,T,d) > 0$ *such that*

$$\int_{\mathbb{R}^d} \frac{|\nabla p_t^{1,x}|^2}{p_t^{1,x}}(y)\mathrm{d}y \le \frac{c}{t}, \quad t \in (0,T], x \in \mathbb{R}^d. \tag{3.5}$$

To prove Proposition 3.2, we first present the following lemma.

Lemma 3.3. *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 there exists a constant $c = c(K,T,d,\beta) > 0$ *such that*

$$\begin{aligned} &\Big| \int_{\mathbb{R}^d} (p_r^{1,x} \log p_r^{1,x})(y)\mathrm{d}y - \int_{\mathbb{R}^d} (p_t^{1,x} \log p_t^{1,x})(y)\mathrm{d}y \Big| \\ &\qquad \le c \log\Big(1 + \frac{t}{r}\Big), \quad 0 < r \le t \le T, x \in \mathbb{R}^d. \end{aligned} \tag{3.6}$$

Proof. Let $x \in \mathbb{R}^d$ be fixed. Simply denote $\rho_t(y) := p_t^{1,x}(y), t \in (0,T], y \in \mathbb{R}^d$. Let $\theta_t(x)$ solve

$$\partial_t \theta_t(x) = b_1(t, \theta_t(x)), \quad \theta_0(x) = x, \quad t \in [0,T]. \tag{3.7}$$

By [18, Theorem 1.2], there exists a constant $c_0 = c_0(K,T,d) > 1$ such that

$$\frac{1}{c_0 t^{\frac{d}{2}}} \mathrm{e}^{-\frac{c_0|\theta_t(x)-y|^2}{t}} \le \rho_t(y) \le \frac{c_0}{t^{\frac{d}{2}}} \mathrm{e}^{-\frac{|\theta_t(x)-y|^2}{c_0 t}}, \quad x,y \in \mathbb{R}^d, t \in (0,T]. \tag{3.8}$$

Consequently,

$$\int_{\mathbb{R}^d} \rho_t(y) \log \rho_t(y)\mathrm{d}y \le \int_{\mathbb{R}^d} \rho_t(y) \log[c_0 t^{-\frac{d}{2}}]\mathrm{d}y = \log[c_0 t^{-\frac{d}{2}}], \quad t \in (0,T]. \tag{3.9}$$

On the other hand, by (3.8) and Jensen's inequality, there exists a constant $c_1 > 0$ such that

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

Hence,

$$\int_{\mathbb{R}^d} \rho_t(y) \log \rho_t(y)\mathrm{d}y \ge \log[c_1 t^{-\frac{d}{2}}], \quad t \in (0,T].$$

Combining this with (3.9), we derive find a constant $c > 0$ such that

$$\int_{\mathbb{R}^d} \rho_r(y) \log \rho_r(y) \mathrm{d}y - \int_{\mathbb{R}^d} \rho_t(y) \log \rho_t(y) \mathrm{d}y \geq \log[c_1 r^{-\frac{d}{2}}] - \log[c_0 t^{-\frac{d}{2}}] \geq -c \log\Big(1 + \frac{t}{r}\Big),$$

and similarly,

$$\int_{\mathbb{R}^d} \rho_r(y) \log \rho_r(y) \mathrm{d}y - \int_{\mathbb{R}^d} \rho_t(y) \log \rho_t(y) \mathrm{d}y \leq \log[c_0 r^{-\frac{d}{2}}] - \log[c_1 t^{-\frac{d}{2}}] \leq c \log\Big(1 + \frac{t}{r}\Big).$$

So, (3.6) holds. □

Proof of Proposition 3.2. 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

$$\|\nabla^i b_1\|_{0\to T,\infty} + \|\nabla^i a_1\|_{0\to T,\infty} < \infty, \quad i \geq 1. \tag{3.10}$$

By [18, Theorem 1.2], there exist a constant $\lambda > 1$ and a measurable map $\theta : [0, T] \to \mathbb{R}^d$ such that

$$\lambda^{-1} t^{-\frac{d+i}{2}} \mathrm{e}^{-\frac{\lambda|\theta_t - y|^2}{t}} \leq \big|\nabla^i \rho_t\big|(y) \leq \lambda t^{-\frac{d+i}{2}} \mathrm{e}^{-\frac{|\theta_t - y|^2}{\lambda t}}, \quad t \in (0, T], y \in \mathbb{R}^d, i = 0, 1, 2. \tag{3.11}$$

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

$$\partial_t \rho_t = \mathrm{div}\Big[a_1(t, \cdot)\nabla \rho_t + \rho_t\{\mathrm{div} a_1(t, \cdot) - b_1(t, \cdot)\}\Big], \quad t \in (0, T]. \tag{3.12}$$

By (3.11), (3.12) and integration by parts formula, we obtain

$$\begin{aligned} &\int_{\mathbb{R}^d} \big\{\rho_t \log \rho_t - \rho_r \log \rho_r\big\}(y)\mathrm{d}y = \int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \big\{(1 + \log \rho_s)\partial_s \rho_s\big\}(y)\mathrm{d}y \\ &= -\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle a_1(s, \cdot)\nabla \log \rho_s + \mathrm{div} a_1(s, \cdot) - b_1(s, \cdot), \nabla \rho_s\Big\rangle(y)\mathrm{d}y. \end{aligned} \tag{3.13}$$

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

$$\begin{aligned} &\int_{\mathbb{R}^d} \big\{\rho_t \log \rho_t - \rho_r \log \rho_r\big\}(y)\mathrm{d}y + \frac{1}{K}\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla \rho_s|^2}{\rho_s}(y)\mathrm{d}y \\ &\leq -\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle \mathrm{div} a_1(s, \cdot) - b_1(s, \cdot), \nabla \rho_s\Big\rangle(y)\mathrm{d}y \\ &= \int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle \big[b_1^{(0)} - \mathrm{div} a_1\big](s, \cdot), \nabla \rho_s\Big\rangle(y)\mathrm{d}y + \int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \Big\langle b_1^{(1)}(s, \cdot), \nabla \rho_s\Big\rangle(y)\mathrm{d}y. \end{aligned} \tag{3.14}$$

By (3.10), (3.11) and Lemma 3.3, we derive

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla \rho_s|^2}{\rho_s}(y)\mathrm{d}y < \infty. \tag{3.15}$$

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

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

Moreover, by the integration by parts formula, (3.11) and $\|\nabla b_1^{(1)}\|_{0\to T,\infty} \le K$, we obtain

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

Combining these with (3.14) and (3.15), we derive

$$\begin{aligned}&\int_r^t \mathrm{d}s\int_{\mathbb{R}^d}\frac{|\nabla\rho_s|^2}{\rho_s}(y)\mathrm{d}y\\ &\le 2K\int_{\mathbb{R}^d}\big\{\rho_r\log\rho_r - \rho_t\log\rho_t\big\}(y)\mathrm{d}y + 2K^2(2K^2+1)(t-r).\end{aligned} \tag{3.16}$$

(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 (3.10) and (A_1) for the same constant K. So, by step (a) and Lemma 3.3, the density function $\rho_t^{n)}$ for the diffusion process generated by $L_t^{a_1^{n)},b_1^{n)}}$ satisfies

$$\int_r^t \mathrm{d}s \int_{\mathbb{R}^d} \frac{|\nabla\rho_s^{n)}|^2}{\rho_s^{n)}}(y)\mathrm{d}y \le c\log\Big(1+\frac{t}{r}\Big), \quad 0 < r \le t \le T, n \ge 1 \tag{3.17}$$

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

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

By [26, 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

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

for any $f\in C_0^{0,2}([r,t]\times\mathbb{R}^d)$. Therefore, (3.4) holds.

(c) If (3.2) holds, then by Malliavin's calculus, see for instance [19] or [36, 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

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

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 (3.5) holds. □

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. (1) Let $p>1$ and $\varepsilon\in(0,\frac{1}{2}]$ be in Proposition 2.3. By Proposition 3.2 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 (1.8) with $\nu=\delta_{x_1}$ and (3.4), we find a constant $c_1=c_1(K,T,d,\varphi)>0$ such that

$$\begin{aligned}&\mathrm{Ent}(P_{t_1}^{1,x_1}|P_{t_1}^{\langle t_0\rangle x_1})\\ &\le c_1\bigg[\|a_1-a_2\|_{\varepsilon t_1\to t_1,\infty}^2+\int_{\varepsilon t_1}^{t_1}\big(\|\mathrm{div}(a_1-a_2)\|_{t,\infty}^2+\|b_1-b_2\|_t^2\big)\mathrm{d}t\bigg],\\ &\quad t_1\in(0,T],x_1\in\mathbb{R}^d.\end{aligned}\tag{3.18}$$

Combining this with (2.4) and Proposition 2.3, we find a constant $c=c(K,T,d,\varphi)>0$ such that for any $t_1\in(0,T]$ and $x_1,x_2\in\mathbb{R}^d$,

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

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

$$\int_{\mathbb{R}^d} \big\{\log f(y)\big\} P_t^{1,x_1}(\mathrm{d}y) \le \log \int_{\mathbb{R}^d} f(y) P_t^{2,x_2}(\mathrm{d}y) + I_t(x_1,x_2), \quad x_1,x_2 \in \mathbb{R}^d. \tag{3.19}$$

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

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

Hence, (3.1) holds.

(2) Let (3.2) hold. By (1.8) and (3.5), we find a constant $c_1 = c_1(K,T,d,\varphi) > 0$ such that for any $t \in (0,T]$ and $x_1 \in \mathbb{R}^d$,

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

Then as explained above that using this estimate to replace (3.18), we derive (3.3) for some constant $c = c(K,T,d,\varphi) > 0$. □

4 Application to McKean-Vlasov SDEs

As an application of Theorem 3.1, we are able to establish (1.7) for (1.6) 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 4.1. *Assume* (B). *Then* (1.6) *is well-posed for distributions in* $\mathscr{P}_2$, *and there exists a constant* $c=c(K,T,d,\varphi)>0$ *such that* (1.7) *holds.*

Proof. 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 [14] for $k=2$. Therefore, by [14, Theorem 1.1], (1.6) 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 $\mathscr{L}_{X_0}\in\mathscr{P}_2$.

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

$$\begin{aligned}&a_i(t,x):=a(t,x,P_t^*\nu_i)=\frac12(\sigma\sigma^*)(t,x,P_t^*\nu_i),\\ &b_i(t,x):=b(t,x,P_t^*\nu_i),\quad b_i^{(k)}(t,x):=b_i^{k,P_t^*\nu_i}(t,x),\quad k=0,1.\end{aligned}\tag{4.2}$$

By Theorem 3.1, under (B), there exists a constant $c_1=c_1(K,T,d,\varphi)>0$ such that for any $t\in(0,T]$,

$$\begin{aligned}&\mathrm{Ent}(P_t^*\nu_1|P_t^*\nu_2)\le\frac{c_1}{t}\mathbb{W}_2(\nu_1,\nu_2)^2\\ &+c_1\|b_1-b_2\|_{t,\infty}^2+c_1\log(1+t^{-1})\|a_1-a_2\|_{t,\infty}^2+c_1t\|\mathrm{div}(a_1-a_2)\|_{t,\infty}^2\\ &\le\frac{c_1}{t}\mathbb{W}_2(\nu_1,\nu_2)^2+c_1K^2\big\{1+\log(1+t^{-1})+t\big\}\sup_{s\in[0,t]}\mathbb{W}_2(P_s^*\nu_1,P_s^*\nu_2)^2.\end{aligned}$$

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 Proposition 4.2, we derive (1.7) for some constant $c>0$, and hence finish the proof. □

Proposition 4.2. *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 (2.10) for large enough $\lambda > 0$ such that (2.19) 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 (2.22), by (2.10), (1.6) for X_t^i and Itô's formula, we have

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

Combining this with $(B)(1)$, (2.19), (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 (2.19), (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

$$\sup_{t\in[0,T]}\mathbb{W}_2(P_t^*\nu_1, P_t^*\nu_2)^2 = \xi_T \le k_2\mathrm{e}^{k_2T}\mathbb{W}_2(\nu_1,\nu_2)^2.$$

So, the proof is finished. □

Remark 4.1. After an earlier version of this paper is available online, the bi-coupling argument developed here has been applied in [10, 22] for singular and degenerate models, where condition $(B)(3)$ is weakened in [10] by using $\mathbb{W}_\psi + \mathbb{W}_k$ replacing $\mathbb{W}_2$, see [10, Theorem 1.3, Remark 1.2] for details. We believe that with additional efforts this new coupling argument will enables one to derive the entropy-cost inequality for McKean-Vlasov SDEs with singular potentials, where $b_t(x,\mu)$ is given by

$$b_t(x,\mu) := \int_{\mathbb{R}^d} V(x-y)\mu(\mathrm{d}y)$$

for V being a singular potential such as the Coulomb potential $V(x) = |x|^{2-d}$ for $d > 2$ and $V(x) = \log|x|$ for $d = 2$. This will be addressed in a forthcoming paper.

Acknowledgement. The authors would like to thank Professor Xing Huang and the referees for useful conversations and corrections.

Author Contributions. The manuscript has been developed in collaboration by the two stated authors.

Data Availibility. No datasets were generated or analysed during the current study.

Declarations Conflict of interest. Partial financial support was received from the National Key R&D Program of China (NO. 2022YFA1006000, 2020YFA0712900) and NNSFC (11921001).

References

[1] R. A. Adams, J. F. Fournier, *Sobolev Spaces,* Academic Press, 2003.

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

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

[4] Y. Bai, X. Huang, *Log-Harnack inequality and exponential ergodicity for distribution dependent CKLS and Vasicek Model,* J. Theor. Probab. 36(2023), 1902-1921.

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

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

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

[8] 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.

[9] I. Goodfellow, Y. Bengio, Y., A. Courville, *Deep Learning (Vol. 1)*, MIT Press, Cambridge 2016.

[10] X. Huang, P. Ren, F.-Y. Wang, *Probability distance estimates between diffusion processes and applications to singular McKean-Vlasov SDEs,* J. Diff. Equat. 420(2025), 376-399.

[11] 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.

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

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

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

[15] X. Huang, F.-Y. Wang, *Regularities and exponential ergodicity in entropy for SDEs driven by distribution dependent noise,* Bernoulli 30(2024), 3303-3323.

[16] X. Huang, F.-Y. Wang, *Log-Harnack inequality and Bismut formula for McKean-Vlasov SDEs with singularities in all variables,* arXiv:2207.11536.

[17] S. Kullback, R. A. Leibler, *On information and sufficiency,* Ann. Math. Statist. 22(1951), 79-86.

[18] 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.

[19] D. Nualart, *The Malliavin Calculus and Related Topics,* Springer, 2006.

[20] F. Otto, *The geometry of dissipative evolution equations: the porous media equation,* Comm. Part. Diff. Equations 26(2001), 101-174.

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

[22] Z. Qian, P. Ren, F.-Y. Wang, *Entropy estimate for degenerate SDEs with applications to nonlinear kinetic Fokker-Planck equations,* SIAM J. Math. Anal. 56 (2024), no. 4, 5330-5349.

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

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

[25] A.-S. Sznitman, *Topics in Propagations of Chaos,* Lecture notes in Math. Vol. 1464, pp. 165-251, Springer, Berlin.

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

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

[28] C. Villani, *Optimal transport: Old and New Part 1,2,* Springer, 2009.

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

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

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

[32] 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.

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

[34] F.-Y. Wang, *Distribution dependent reflecting stochastic differential equations,* Sci. China Math. 66(2023), 2411-2456.

[35] F.-Y. Wang, *Derivative formula for singular McKean-Vlasov SDEs,* Comm. Pure Appl. Anal. 22(2023), 1866–1898.

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

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

[38] 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.