UNADJUSTED LANGEVIN ALGORITHMS FOR SDES WITH HÖLDER DRIFT

XIANG LI, FENG-YU WANG, AND LIHU XU

ABSTRACT. Consider the following stochastic differential equation for $(X_t)_{t\geq 0}$ on $\mathbb{R}^d$ and its Euler-Maruyama (EM) approximation $(Y_{t_n})_{n\in\mathbb{Z}^+}$:

$$\begin{aligned} \mathrm{d}X_t &= b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \\ Y_{t_{n+1}} &= Y_{t_n} + \eta_{n+1} b(Y_{t_n}) + \sigma(Y_{t_n})\left(B_{t_{n+1}} - B_{t_n}\right), \end{aligned}$$

where $b : \mathbb{R}^d \mapsto \mathbb{R}^d,\ \ \sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are measurable, B_t is the d-dimensional Brownian motion, $t_0 := 0, t_n := \sum_{k=1}^n \eta_k$ for constants $\eta_k > 0$ satisfying $\lim_{k\to\infty}\eta_k = 0$ and $\sum_{k=1}^\infty \eta_k = \infty$. Under (partial) dissipation conditions ensuring the ergodicity, we obtain explicit convergence rates of $\mathbb{W}_p(\mathscr{L}(Y_{t_n}), \mathscr{L}(X_{t_n})) + \mathbb{W}_p(\mathscr{L}(Y_{t_n}), \mu) \to 0$ as $n \to \infty$, where $\mathbb{W}_p$ is the L^p-Wasserstein distance for certain $p \in [0,\infty)$, $\mathscr{L}(\xi)$ is the distribution of random variable ξ, and μ is the unique invariant probability measure of $(X_t)_{t\geq 0}$. Comparing with the existing results where b is at least C^2-smooth, our estimates apply to Hölder continuous drift and can be sharp in several specific situations.

CONTENTS

1. Introduction 1
2. $\mathbb{W}_p$-estimate for $p \in [0,1]$: partial dissipation case 4
2.1. Main result and an example 4
2.2. Proof of Theorem 2.1 6
3. $\mathbb{W}_p$-estimate for $p > 1$: the uniform dissipation case 22
3.1. Main result and an example of bridge regression 22
3.2. Proof of Theorem 3.1 24
References 27

1. INTRODUCTION

We consider the following time homogenous stochastic differential equation(SDE) on $\mathbb{R}^d$:

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \ \ t\geq 0, \tag{1.1}$$

where $b : \mathbb{R}^d \mapsto \mathbb{R}^d,\ \ \sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are measurable, and B_t is the d-dimensional Brownian motion under a probability base $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$. Under conditions allowing singular drift, the

2020 *Mathematics Subject Classification.* Primary 65C05, 65C40, 60H10; Secondary 60H35, 62E17.
Key words and phrases. Unadjusted langevin algorithm, singular drift SDEs, Euler-Maruyama scheme with decreasing step, Wasserstein distance, total variation distance, convergence rate.
F. Y. Wang is supported in part by the National Key R&D Program of China (2022YFA1006000, 2020YFA0712900) and NNSFC (11921001). L. Xu is supported by NNSFC grant (No. 12071499) and University of Macau grant MYRG2018- 00133-FST.

well-posedness, regularity estimates and exponential ergodicity have been derived for (1.1), see for instance [10, 21, 26, 28, 25, 27] and references therein.

In recent years, stochastic algorithms arising from statistics and machine learning have been intensively developed to simulate the invariant probability measure for a stochastic system, where a practical algorithm is the unadjusted Langevin algorithm (ULA) using the Euler-Maruyama(EM) scheme of the associated SDE. See for instance [8, 4, 5, 16, 6, 19] for the background of the study.

For a sequence of step sizes $\{\eta_k\}_{k\geq 1}$, the EM Scheme of (1.1) is defined by the following induction

$$Y_{t_{k+1}} = Y_{t_k} + \eta_{k+1} b(Y_{t_k}) + \sigma(Y_{t_k})\left(B_{t_{k+1}} - B_{t_k}\right), \qquad Y_{t_0} = Y_0 = X_0,\ k \in \mathbb{Z}^+, \tag{1.2}$$

with $Y_0 = X_0 = x$, where $t_k := \sum_{i=1}^k \eta_i$ and $B_{t_{k+1}} - B_{t_k}$ can be identified as $\sqrt{\eta_{k+1}}\zeta_k$ for i.i.d. d-dimensional standard normal distributed random variables $\{\zeta_k\}_{k\geq 1}$. The associated continuous time Euler Scheme is defined by

$$Y_t = Y_{t_k} + (t - t_k) b(Y_{t_k}) + \sigma(Y_{t_k})\left(B_t - B_{t_k}\right), \qquad t \in [t_k, t_{k+1}), k \geq 0, Y_0 = X_0. \tag{1.3}$$

In addition to the aforementioned ULA algorithm, the equation (1.2) can also be interpreted as a noisy gradient descent(GD) algorithm as $b(x) = -\nabla V(x)$ with $V(x)$ being a loss function in a certain optimization problem. Motivated by these two algorithms and their variants, the Euler-Maruyama(EM) scheme of SDEs have been extensively investigated under different assumptions and settings, see for instance [1, 4, 5, 6, 17, 22].

In this paper, we investigate the convergence rate of

$$\mathbb{W}_p(\mathscr{L}(X_{t_n}), \mathscr{L}(Y_{t_n})) + \mathbb{W}_p(\mathscr{L}(Y_{t_n}), \mu) \to 0 \text{ as } n \to \infty,$$

where $p \in [0, \infty)$, μ is the unique invariant probability measure of $(X_t)_{t\geq 0}$, $\mathscr{L}(\xi)$ is the distribution of a random variable ξ, and for $\mathscr{C}(\mu_1, \mu_2)$ being the class of couplings of probability measures μ_1, μ_2 on $\mathbb{R}^d$,

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

is the L^p-Wasserstein distance. By Kantorovich's dual formula, for any $p \in [0, 1]$ we have

$$\mathbb{W}_p(\mu_1, \mu_2) = \sup_{f \in \mathscr{B}_b, [f]_p \leq 1} |\mu_1(f) - \mu_2(f)|,$$

where $\mathscr{B}_b$ is the set of all bounded measurable functions on $\mathbb{R}^d$, and

$$[f]_p := \sup_{x\neq y} \frac{|f(x) - f(y)|}{|x - y|^p}$$

is the p-order modulus of continuity for f. In particular,

$$2\mathbb{W}_0(\mu_1, \mu_2) = \|\mu_1 - \mu_2\|_{TV} := \sup_{|f|\leq 1} |\mu_1(f) - \mu_2(f)|$$

is the total variation distance.

It is worthy to point out that for $b(x) = -\nabla V(x)$ and $\sigma = I_d$ (the $d\times d$ identity matrix), where V is a strictly convex function with bounded $\nabla^2 V$, the convergence rate under the Kullback-Leibler divergence (i.e. the relative entropy) has been studied in [4, 1, 17].

By the ergodicity, we have $\mathscr{L}(X_{t_n}) \to \mu$ as $t_n \to \infty$, where $\mathscr{L}(X_{t_n})$ is the distribution of X_{t_n} and μ is the unique invariant probability measure. So, it is essential to assume $t_n \to \infty$ as $n \to \infty$, i.e. $\sum_{k=1}^{\infty} \eta_k = \infty$. On the other hand, to approximate $\mathscr{L}(X_{t_n})$ by $\mathscr{L}(Y_{t_n})$, the step size η_n should decay to 0 as $n \to \infty$. Thus, throughout the paper we make the following assumption.

$(A0)$ The step size sequence $\{\eta_k > 0\}_{k\geq 1}$ is non-increasing and satisfies

$$\lim_{k\to\infty} \eta_k = 0, \qquad \sum_{k=1}^{\infty} \eta_k = \infty.$$

A typical examples of step size $\{\eta_k\}_{k\geq 1}$ satisfying $(A0)$ is $\eta_k = \theta k^{-a}$ for some $\theta \in (0,\infty)$ and $a \in (0,1]$.

Before moving on, let us recall some existing estimates on the convergence rate presented in the literature for $\eta_k = \frac{1}{k}$.

When $b = -\nabla V$ for V being strictly convex and $\nabla^2 V$ being bounded, Durmus and Moulines [5, 6] showed that $\mathbb{W}_2(\mathscr{L}(Y_{t_n}),\mu)$ and $\mathbb{W}_0(\mathscr{L}(Y_{t_n}),\mu)$ are both bounded by $O(d^{\frac{1}{2}} n^{-\frac{1}{2}})$, and that their bounds can be improved to $O(dn^{-1})$ and $O(d^{\frac{1}{2}} n^{-\frac{3}{4}})$ respectively if $V \in C^3$ has Lipschitz continuous first and second order derivatives.

Recently, Pagès and Panloup [19] studied the non-asymptotic bounds related to the EM scheme (1.2), for which they replaced the strongly convex assumption on V in [6] with a weaker assumption on the drift b and allows a multiplicative noise (i.e. σ is non-constant). They obtained an $O(n^{-1+\epsilon})$ upper bound with $\epsilon \in (0,1)$, for $\mathbb{W}_0(\mathscr{L}(Y_{t_n}),\mu)$ under the conditions $b,\sigma \in C^6$, and an $O(n^{-1}\log(n))$ upper bound for $\mathbb{W}_1(\mathscr{L}(Y_{t_n}),\mu)$ under the conditions $b,\sigma \in C^4$. When σ is a constant matrix, their upper bounds for both distances can be improved to $O(n^{-1})$ under the weaker condition that $b \in C^3$. Note that their results fail to provide an explicit dependence on the dimension d.

However, in many important applications such as lasso and bridge regressions, the drift b is much more singular than those required in the above references. This motivates us to study the convergence of the unadjusted Langevin algorithm for SDE (1.1) in the case $b \in C^\alpha$ with $\alpha \in (0,2]$ for partially filling this gap. If $b \in C^\alpha$ with $\alpha \in (0,1]$, it means that b is α-Hölder continuous, while if $b \in C^\alpha$ with $\alpha \in (1,2]$, it means that ∇b is $(\alpha-1)$-Hölder continuous.

To compare our main results with the above introduced estimates, we simply let $\eta_k = \frac{1}{k}$ and briefly summarize our results as follows.

1. When b is partially dissipative with $b \in C^\alpha$ for some $\alpha \in (0,2]$, and $\sigma(x) = \sigma$ is constant, Theorem 2.1 implies that

$$\mathbb{W}_1(\mathscr{L}(Y_{t_n}),\mu) + \mathbb{W}_0(\mathscr{L}(Y_{t_n}),\mu) \leq O(d^{\frac{1}{2}+1_{\{1<\alpha\leq 2\}}} n^{-\frac{\alpha}{2}}).$$

   This estimate is also extended to $\mathbb{W}_p$ for $p \in (0,1)$.When $\alpha = 2$, and $p = 0,1$, the order of n in our result matches the optimal order shown in [19, Theorem 2.3]. Note that our result is obtained under weaker condition on b rather than $b \in C^3$ as in [19].
2. When b is uniformly dissipative and in C^α for some $\alpha \in (0,1]$, Theorem 3.1 implies

$$\mathbb{W}_p(\mathscr{L}(Y_{t_n}),\mathscr{L}(X_{t_n})) \leq O(d^{\frac{1}{2}} n^{-\frac{\alpha}{2}})$$

   for any $p \in (1,\infty)$. In particular, when $\alpha = 1$ and $p = 2$ this estimate has the same order as that in [6].

Let us briefly discuss our approach to proving the main results. In the uniform dissipation case, we use a synchronous coupling of the solutions of (1.1) and (1.3) and compare their difference directly. However, this method is not applicable in the case of partial dissipation. Alternatively, we apply a domino decomposition in [19], whose probability version was recently established in [3] by a generalized Lindeberg principle and seems more powerful in applications. Comparing to [19], it seems to us that our one-step error estimation is a little more delicate, and thus allows us to achieve the optimal rate under a weaker assumption on b. More precisely, we establish an integration by parts formula for the unadjusted Langevin algorithm rather than directly estimate a Malliavin matrix, which helps us to get estimates explicitly depending on the dimension d. Girsanov's theorem and Pinsker's inequality also play an important role in our proof.

The paper is organized as follows. When the coefficients only satisfy a partial dissipation condition, we consider the additive noise where $\sigma(x) = \sigma$ does not depend on x and t, and estimate $\mathbb{W}_p(\mathscr{L}(Y_{t_n}), \mu)$ in Section 2 for $p \in [0, 1]$. In Section 3, we estimate $\mathbb{W}_p(\mathscr{L}(Y_{t_n}), \mathscr{L}(X_{t_k}))$ for $p > 1$ under the uniform dissipation condition, where the time in-homogenous case is also considered.

Throughout this paper, let $|\cdot|$ be the Euclidean norm, let $\|\cdot\|_{\rm op}$ and $\|\cdot\|_{\rm HS}$ denote the operator norm and Hilbert-Schmidt norm for matrices respectively. We will use $c = c(\cdots)$ to denote a constant c which only depends on the quantities "$\cdots$". The notations $C^k(\mathcal{X}, \mathcal{Y})$ is used to denote the classes of continuous functions mapping from $\mathcal{X}$ to $\mathcal{Y}$ that have continuous first, second,... and k-th order partial derivatives and if the context of the function's space is clear, we will abbreviate the notation of the space as C^k.

2. $\mathbb{W}_p$-ESTIMATE FOR $p \in [0, 1]$: PARTIAL DISSIPATION CASE

2.1. Main result and an example.

In this section, we study the $\mathbb{W}_p$-estimate for $p \in [0, 1]$ under the following assumptions

$(A1)$ Let $\alpha \in (0, 2]$. $\sigma(x) = \sigma$ does not depend on x and is invertible, and there exist constants $K_1, K_2 \in (0, \infty)$ such that

$$\|\sigma\|_{\rm op} \vee \|\sigma^{-1}\|_{\rm op} \le K_1$$

and b is partially dissipative, i.e.

$$\langle b(x) - b(y), x - y\rangle \le K_1 - K_2|x - y|^2, \quad x, y \in \mathbb{R}^d.$$

Moreover,

- when $\alpha \le 1$,

$$|b(x) - b(y)| \le K_1(|x - y| + |x - y|^\alpha), \quad x, y \in \mathbb{R}^d;$$

- when $\alpha \in (1, 2]$, $b \in C^\alpha(\mathbb{R}^d)$ and

$$\|\nabla b\|_{\rm op} \le K_1, \quad \|\nabla b(x) - \nabla b(y)\|_{\rm op} \le K_1|x - y|^{\alpha-1}, \quad x, y \in \mathbb{R}^d.$$

This assumption implies the well-posedness of (1.1), see [21, 29] for the well-posedness of more singular SDEs. The following result improves [19, Theorem 2.2]. In particular, when $\alpha = 2$ and $p > 0$, the order n^{-1} in (2.3) is sharp for $\eta_k \sim \frac{1}{k}$ is sharp since it is reached by the Ornstein-Uhlenbeck process, for which $\mathbb{W}_0(\mathscr{L}(Y^x_{t_n}), \mu)$ can be computed explicitly (see [19, Section 4.6]).

Theorem 2.1. *Assume* $(A0)$ *and* $(A1)$ *for some* $\alpha \in (0,2]$. *Then there exist constants* $c_1 = c_1(K_1, K_2, \eta, \alpha), c_2 = c_2(K_1, K_2) \in (0, \infty)$ *such that the following statements hold for any* $x \in \mathbb{R}^d$ *and* $n \geq 2$.

(1) *Let* $\alpha \in (0,1]$. *We have*

$$\begin{aligned}&\mathbb{W}_1\left(\mathscr{L}(X^x_{t_n}), \mathscr{L}(Y^x_{t_n})\right) + \mathbb{W}_1\left(\mu, \mathscr{L}(Y^x_{t_n})\right) \\ &\leq c_1 d^{\frac{1}{2}}(1+|x|)\sum_{k=1}^{n-1} \mathrm{e}^{-c_2(t_n-t_k)}\eta_k^{1+\frac{\alpha}{2}},\end{aligned} \tag{2.1}$$

$$\begin{aligned}&\mathbb{W}_0\left(\mathscr{L}(X^x_{t_n}), \mathscr{L}(Y^x_{t_n})\right) + \mathbb{W}_0\left(\mu, \mathscr{L}(Y^x_{t_n})\right) \\ &\leq c_1 d^{\frac{1}{2}}(1+|x|)\left[\sum_{k=1}^{n-1} \mathrm{e}^{-c_2(t_n-t_k)}(t_n-t_k)^{-\frac{1}{2}}\eta_k^{1+\frac{\alpha}{2}} + \eta_n^{\frac{1+(\alpha\wedge 1)}{2}}\right].\end{aligned} \tag{2.2}$$

In particular, if we choose $\eta_k = \frac{\theta}{k}$ *for some constant* $\theta > \frac{\alpha}{2c_2}$, *then there exists a constant* $c_1' = c_1'(K_1, K_2, \theta, \alpha) \in (0, \infty)$ *such that*

$$\mathbb{W}_i\left(\mathscr{L}(X^x_{t_n}), \mathscr{L}(Y^x_{t_n})\right) + \mathbb{W}_i\left(\mu, \mathscr{L}(Y^x_{t_n})\right) \leq c_1' d^{\frac{1}{2}} n^{-\frac{\alpha}{2}}, \;\; i = 0, 1. \tag{2.3}$$

(2) *Let* $\alpha \in (1,2]$. *We have*

$$\begin{aligned}&\mathbb{W}_1\left(\mathscr{L}(X^x_{t_n}), \mathscr{L}(Y^x_{t_n})\right) + \mathbb{W}_1\left(\mu, \mathscr{L}(Y^x_{t_n})\right) \\ &\leq c_1 d^{\frac{3}{2}}(1+|x|^2)\sum_{k=1}^{n-1} \mathrm{e}^{-c_2(t_n-t_k)}\eta_k^{1+\frac{\alpha}{2}},\end{aligned} \tag{2.4}$$

$$\begin{aligned}&\mathbb{W}_0\left(\mathscr{L}(X^x_{t_n}), \mathscr{L}(Y^x_{t_n})\right) + \mathbb{W}_0\left(\mu, \mathscr{L}(Y^x_{t_n})\right) \\ &\leq c_1 d^{\frac{3}{2}}(1+|x|^2)\left[\sum_{k=1}^{n-1} \mathrm{e}^{-c_2(t_n-t_k)}(t_n-t_k)^{-\frac{1}{2}}\eta_k^{1+\frac{\alpha}{2}} + \eta_n^{\frac{1+(\alpha\wedge 1)}{2}}\right].\end{aligned} \tag{2.5}$$

In particular, if we choose $\eta_k = \frac{\theta}{k}$ *for some constant* $\theta > \frac{\alpha}{2c_2}$, *then there exists a constant* $c_1' = c_1'(K_1, K_2, \theta, \alpha) \in (0, \infty)$ *such that*

$$\mathbb{W}_i\left(\mathscr{L}(X^x_{t_n}), \mathscr{L}(Y^x_{t_n})\right) + \mathbb{W}_i\left(\mu, \mathscr{L}(Y^x_{t_n})\right) \leq c_1' d^{\frac{3}{2}} n^{-\frac{\alpha}{2}}, \;\; i = 0, 1. \tag{2.6}$$

(3) *For any* $p \in (0,1)$,

$$\begin{aligned}&\mathbb{W}_p\left(\mathscr{L}(X^x_{t_n}), \mathscr{L}(Y^x_{t_n})\right) + \mathbb{W}_p\left(\mu, \mathscr{L}(Y^x_{t_n})\right) \\ &\leq c_1 d^{1+\frac{p}{2}+1_{\{1<\alpha\leq 2\}}}\left(1+|x|^{1+1_{\{1<\alpha\leq 2\}}}\right)\left[\sum_{k=1}^{n-1} \mathrm{e}^{-c_2(t_n-t_k)}(t_n-t_k)^{-\frac{1}{2}}\eta_k^{1+\frac{\alpha}{2}} + \eta_n^{\frac{1+(\alpha\wedge 1)}{2}}\right].\end{aligned}$$

Example 2.1. Let us consider using the ULA algorithm for sampling from a d-dimensional distribution with density function $\frac{1}{Z}e^{-\frac{1}{2}|x|^2+\frac{1}{4}|x|^{\alpha+1}}, \alpha \in (0,1]$ where Z is the normalization constant. In this case, the iterative formula for the ULA algorithm with decreasing step size $\eta_k, k \geq 1$ is as follow

$$Y_{k+1} = Y_k + \eta_{k+1}\left(-Y_k + \frac{\alpha+1}{4}\left|Y_k\right|^{\alpha-1}Y_k\right) + \sqrt{2\eta_{k+1}}\zeta_{k+1}, \quad k \geq 0,$$

where $\zeta_1, \dots, \zeta_k, \dots$ are i.i.d. d-dimensional standard normal distributed random variables, and the corresponding SDE should be

$$\mathrm{d}X_t = \left(-X_t + \frac{\alpha+1}{4}|X_t|^{\alpha-1}X_t\right)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad t \geq 0.$$

It can be easily verified that in this case, the drift coefficient $b(x) = -x + \frac{\alpha+1}{4}|x|^{\alpha-1}x$ is not uniformly dissipative (in other words, V is non-convex for V satisfying $b = -\nabla V$), and non-differentiable near the origin if $\alpha < 1$. However, this example satisfies our assumptions in $(A1)$. Hence, according to Theorem 2.1, if η_k is chosen to be $\frac{\theta}{k}$ for some suitable $\theta > 0$, then the approximation error of this algorithm under $\mathbb{W}_1 + \mathbb{W}_0$ has convergence rate $O(d^{\frac{1}{2}}n^{-\frac{\alpha}{2}})$.

2.2. **Proof of Theorem 2.1.**

We first present the following moment estimates on Y_t and X_t which are crucial for the proofs.

Lemma 2.2. *Assume that $(A0)$ holds, (1.1) is well-posed and there exist positive constants κ_1, κ_2 such that*

$$\langle x, b(x)\rangle \leq \kappa_1 - \kappa_2|x|^2, \quad |b(x)| \leq \kappa_2(1+|x|), \quad \|\sigma\|_{\mathrm{op}} \leq \kappa_2, \quad \forall x \in \mathbb{R}^d. \tag{2.7}$$

Then for any $p \in (0,\infty)$, there exists a constant $\kappa = \kappa(\kappa_1, \kappa_2, p) \in (0,\infty)$ such that

$$\sup_{t\geq 0} \mathbb{E}|X_t^x|^p \leq \kappa(d^{\frac{p}{2}} + |x|^p), \quad \forall x \in \mathbb{R}^d. \tag{2.8}$$

If moreover

$$|b(x) - b(y)| \leq K(|x-y| + |x-y|^\alpha), \quad \forall x, y \in \mathbb{R}^d \tag{2.9}$$

holds for some constant $K \in (0,\infty)$ and $\alpha \in (0,1]$, then there exists $\kappa' = \kappa'(\kappa_1, \kappa_2, K, \eta, \alpha, p) \in (0,\infty)$ such that

$$\sup_{t\geq 0} \mathbb{E}|Y_t^x|^p \leq \kappa'(d^{\frac{p}{2}} + |x|^p), \quad \forall x \in \mathbb{R}^d. \tag{2.10}$$

Proof. By Jensen's inequality, we only need to prove for $p \geq 2$.

(1) Proof of (2.8). All constants c_i in this step depend only on κ_1, κ_2 and p. By (2.7), $\|\sigma\|_{\mathrm{HS}}^2 \leq d\|\sigma\|_{\mathrm{op}}^2$ and Young's inequality, there exist constants $c_1, c_2 \in (0,\infty)$ such that, for any $x, y \in \mathbb{R}^d$

$$\begin{aligned}
&p|x|^{p-2}\langle x, b(x)\rangle + \frac{1}{2}p(p-1)|x|^{p-2}\|\sigma\|_{\mathrm{HS}}^2 \\
&\leq -\kappa_2 p|x|^p + \kappa_1 p|x|^{p-2} + \frac{1}{2}\kappa_2 p(p-1)d|x|^{p-2} \\
&\leq c_1 d^{\frac{p}{2}} - c_2|x|^p.
\end{aligned} \tag{2.11}$$

So, by Itô's formula and $\|\sigma\|_{\mathrm{op}} \leq \|\sigma\|_{\mathrm{HS}}$,

$$\begin{aligned}
&\mathrm{d}|X_t^x|^p - \mathrm{d}M_t \\
&= \left[p|X_t^x|^{p-2}\langle X_t^x, b(X_t^x)\rangle + \frac{1}{2}p|X_t^x|^{p-2}\|\sigma\|_{\mathrm{HS}}^2 + \frac{1}{2}p(p-2)\,|X_t^x|^{p-4}\,|\sigma X_t^x|^2\right]\mathrm{d}t \\
&\leq \left[p|X_t^x|^{p-2}\langle X_t^x, b(X_t^x)\rangle + \frac{1}{2}p(p-1)|X_t^x|^{p-2}\|\sigma\|_{\mathrm{HS}}^2\right]\mathrm{d}t
\end{aligned}$$

$$\le \big(c_1 d^{\frac{p}{2}} - c_2|X_t^x|^p\big)\mathrm{d}t$$

holds for some martingale M_t. Then it follows from Gronwall's inequality that

$$\mathbb{E}|X_t^x|^p \le |x|^p \mathrm{e}^{-c_2 t} + c_1 d^{\frac{p}{2}} \int_0^t \mathrm{e}^{-c_2(t-s)}\mathrm{d}s \le \frac{c_1 d^{\frac{p}{2}}}{c_2} + |x|^p,$$

so that (2.8) follows.

(2) Proof of (2.10). All constants c_i in this step depend only on $\kappa_1, \kappa_2, K, \eta, \alpha$ and p. For any $k \ge 1$, by (2.7), (2.9) and (2.11), and noting that

$$|Y_t^x - Y_{t_{k-1}}^x| + |Y_t^x - Y_{t_{k-1}}^x|^\alpha \le c(1 + |Y_t^x - Y_{t_{k-1}}^x|)$$

holds for some constant $c > 0$, we find a martingale M_t such that for $t \in [t_{k-1}, t_k]$

$$\begin{aligned}
\mathrm{d}|Y_t^x|^p - \mathrm{d}M_t &\le \Big(p|Y_t^x|^{p-2}\langle Y_t^x, b(Y_{t_{k-1}}^x)\rangle + \frac{1}{2}p(p-1)|Y_t^x|^{p-2}\|\sigma\|_{\mathrm{HS}}^2\Big)\mathrm{d}t\\
&\le \Big(p|Y_t^x|^{p-2}\langle Y_t^x, b(Y_t^x)\rangle + \frac{1}{2}p(p-1)|Y_t^x|^{p-2}\|\sigma\|_{\mathrm{HS}}^2\Big)\mathrm{d}t\\
&\quad + p|Y_t^x|^{p-1}|b(Y_t^x) - b(Y_{t_{k-1}}^x)|\mathrm{d}t\\
&\le \Big[-\kappa_2 p\,|Y_t^x|^p + \frac{1}{2}\kappa_2 p(p-1)d\,|Y_t^x|^{p-2} + Kp\,|Y_t^x|^{p-1}\Big(\Big|Y_t^x - Y_{t_{k-1}}^x\Big| + 1\Big)\Big]\,\mathrm{d}t\\
&\le \Big(-\frac{c_2}{2}|Y_t^x|^p + c_3\big\{d^{\frac{p}{2}} + |Y_t^x - Y_{t_{k-1}}^x|^p\big\}\Big)\mathrm{d}t,
\end{aligned}$$

where the last step follows from the Young's inequality. This implies

$$\begin{aligned}
\mathbb{E}|Y_{t_k}^x|^p &\le \mathrm{e}^{-\frac{1}{2}c_2\eta_k}\mathbb{E}|Y_{t_{k-1}}^x|^p + c_3 d^{\frac{p}{2}}\eta_k\\
&\quad + 2c_3\int_{t_{k-1}}^{t_k} \mathbb{E}[|Y_t^x - Y_{t_{k-1}}^x|^p]\mathrm{d}t, \quad k \ge 1.
\end{aligned} \tag{2.12}$$

By the boundedness of σ and the linear growth of b, we find a constant $c_4 > 0$ such that

$$\begin{aligned}
\mathbb{E}|Y_t^x - Y_{t_{k-1}}^x|^p &= \mathbb{E}\big|(t - t_{k-1})b(Y_{t_{k-1}}^x) + \sigma(B_t - B_{t_{k-1}})\big|^p\\
&\le c_4\left[\eta_k^p(1 + \mathbb{E}|Y_{t_{k-1}}^x|^p) + \eta_k^{\frac{p}{2}} d^{\frac{p}{2}}\right], \quad t \in [t_{k-1}, t_k], k \ge 1.
\end{aligned}$$

Since $\{\eta_k\}_{k\ge1}$ is non-increasiong, combining this with (2.12), we can find a constant $c_5 > 0$ such that

$$\begin{aligned}
\mathbb{E}|Y_{t_k}^x|^p &\le \Big(\mathrm{e}^{-\frac{1}{2}c_2\eta_k} + 2c_3c_4\eta_k^{1+p}\Big)\mathbb{E}|Y_{t_{k-1}}^x|^p + c_3 d^{\frac{p}{2}}\eta_k + 2c_3c_4\left[\eta_k^{1+p} + \eta_k^{1+\frac{p}{2}} d^{\frac{p}{2}}\right]\\
&\le \Big(\mathrm{e}^{-\frac{1}{2}c_2\eta_k} + c_5\eta_k^{1+p}\Big)\mathbb{E}|Y_{t_{k-1}}^x|^p + c_5 d^{\frac{p}{2}}\eta_k, \quad k \ge 1.
\end{aligned} \tag{2.13}$$

Since $\eta_k \to 0$ as $k \to \infty$, we can find a $k_0 \ge 1$ such that

$$0 \le \mathrm{e}^{-\frac{1}{2}c_2\eta_k} + c_5\eta_k^{1+p} \le 1 - \frac{1}{4}c_2\eta_k, \quad k \ge k_0. \tag{2.14}$$

Consequently,

$$c_6 := \sup_{k\ge1}\prod_{i=1}^k \Big(\mathrm{e}^{-\frac{1}{2}c_2\eta_j} + c_5\eta_j^{1+p}\Big) < \infty.$$

So, iterating the estimate (2.13) yields

$$\mathbb{E}|Y_{t_k}|^p \le c_6|x|^p + c_5 d^{\frac{p}{2}}\left[\eta_k + \sum_{i=1}^{k-1}\eta_i \prod_{j=i+1}^{k}\left(\mathrm{e}^{-\frac{1}{2}c_2\eta_j} + c_5\eta_j^{1+p}\right)\right]. \tag{2.15}$$

By (2.14) we obtain

$$\begin{aligned}
&\sum_{i=1}^{k-1}\eta_i\prod_{j=i+1}^{k}\left(\mathrm{e}^{-\frac{1}{2}c_2\eta_j}+c_5\eta_j^{1+p}\right) \le c_6\sum_{i=1}^{k_0-1}\eta_i + \sum_{i=k_0}^{k-1}\eta_i\prod_{j=i+1}^{k}\left(1-\frac{1}{4}c_2\eta_j\right)\\
&= c_6\sum_{i=1}^{k_0-1}\eta_i + \frac{4}{c_2}\sum_{i=k_0}^{k-1}\left[1-(1-\frac{1}{4}c_2\eta_i)\right]\prod_{j=i+1}^{k}\left(1-\frac{1}{4}c_2\eta_j\right)\\
&= c_6\sum_{i=1}^{k_0-1}\eta_i + \frac{4}{c_2}\sum_{i=k_0}^{k-1}\left[\prod_{j=i+1}^{k}\left(1-\frac{1}{4}c_2\eta_j\right) - \prod_{j=i}^{k}\left(1-\frac{1}{4}c_2\eta_j\right)\right]\\
&\le c_6\sum_{i=1}^{k_0-1}\eta_i + \frac{4}{c_2} < \infty, \quad \forall k \ge k_0.
\end{aligned}$$

Combining this with (2.15), we prove (2.10) for some constant κ' and $k \ge k_0$, while for $k \le k_0$ the estimate follows from (2.13). □

Lemma 2.3. *Assume that the conditions in Lemma 2.2 hold. Then, for any $p > 0$ there exists a constant $\kappa = \kappa(\kappa_1, \kappa_2, p) \in (0,\infty)$ such that*

$$\mathbb{E}|X_t^x - x|^p \le \kappa(d^{\frac{p}{2}} + |x|^p)(1\wedge t)^{\frac{p}{2}\wedge 1}, \ \ x\in\mathbb{R}^d, \ \ t\ge 0.$$

If moreover (2.9) holds, then there exists $\kappa' = \kappa'(\kappa_1,\kappa_2,\eta,K,p,\alpha) \in (0,\infty)$ such that

$$\mathbb{E}|Y_t^x - Y_{t_{k-1}}^x|^p \le \kappa'(d^{\frac{p}{2}} + |x|^p)\eta_k^{\frac{p}{2}}, \ \ x\in\mathbb{R}^d, \ k\ge 1, \ t\in[t_{k-1},t_k].$$

Proof. Again, by Jensen's inequality, we only need to prove for $p \ge 2$.

(1) Applying Itô's formula, (2.11) and Young's inequality gives us

$$\begin{aligned}
&\mathrm{d}\,|X_t^x - x|^p - \mathrm{d}M_t\\
&\le \left[p|X_t^x - x|^{p-2}\langle X_t^x - x, b(X_t^x)\rangle + \frac{1}{2}p(p-1)|X_t^x-x|^{p-2}\|\sigma\|_{\mathrm{HS}}^2\right]\mathrm{d}t\\
&\le \left(c_1 d^{\frac{p}{2}} - c_2\,|X_t^x - x|^p + p|X_t^x - x|^{p-2}\langle X_t^x - x, b(X_t^x) - b(X_t^x - x)\rangle\right)\mathrm{d}t\\
&\le \left(c_1 d^{\frac{p}{2}} - c_2\,|X_t^x - x|^p + 2K_1 p|X_t^x - x|^{p-1}(1+|x|)\right)\mathrm{d}t\\
&\le \left(c_1'\left(d^{\frac{p}{2}} + |x|^p\right) - c_2'\,|X_t^x - x|^p\right)\mathrm{d}t
\end{aligned}$$

where the third inequality is a consequence of (2.9) and M_t is the Martingale term. Since $X_0^x - x = 0$, it follows from Gronwall's inequality that

$$\mathbb{E}\,|X_t^x - x|^p \le \frac{c_1'(d^{\frac{p}{2}} + |x|^p)}{c_2'}\left(1 - e^{-c_2' t}\right) \le \kappa(d^{\frac{p}{2}} + |x|^p)\,(1\wedge t)\,, \qquad \forall x\in\mathbb{R}^d, t\ge 0,$$

for some $\kappa = \kappa(\kappa_1, \kappa_2, p)$.

(2) For the second one, notice that for $t \in [t_k, t_{k+1})$, the conditional distribution of $Y_t^x - Y_{t_k}^x$ given $Y_{t_k}^x$ is

$$Y_t^x - Y_{t_k}^x|_{Y_{t_k}^x} \sim \mathcal{N}\left((t-t_k)b(Y_{t_k}^x), (t-t_k)\sigma\sigma^T\right).$$

Hence, it follows from (2.7) and (2.10) that for some constants $c_3' = c_3'(p) > 0$ and $c_4' = c_4'(p, \kappa_2) > 0$,

$$\begin{aligned}\mathbb{E}|Y_t^x - Y_{t_k}^x|^p &\le c_3'\left[(t-t_k)^p\mathbb{E}|b(Y_{t_k}^x)|^p + (t-t_k)^{\frac{p}{2}}\|\sigma\|_{\mathrm{HS}}^p\right]\\ &\le c_3'\kappa_2^p\left[(t-t_k)^p\left(1+\mathbb{E}|Y_{t_k}^x|\right)^p + (t-t_k)^{\frac{p}{2}}d^{\frac{p}{2}}\right]\\ &\le c_4'\left(d^{\frac{p}{2}} + |x|^p\right)(t-t_k)^{\frac{p}{2}}.\end{aligned}$$

So the proof is complete. □

To prove Theorem 2.1, we consider the SDEs (1.1) and (1.3) on each time interval $[t_k, t_{k+1})$ for $k \in \mathbb{Z}^+$, where $t_0 := 0$. For any $x \in \mathbb{R}^d$ and $k \in \mathbb{Z}^+$, let $(Y_{t_k,t}(x))_{t\in[t_k,t_{k+1}]}$ solve the SDE

$$\mathrm{d}Y_{t_k,t}(x) = b(x)\mathrm{d}t + \sigma\mathrm{d}B_t, \quad X_{t_k,t_k}(x) = Y_{t_k,t_k}(x) = x, t \in [t_k, t_{k+1}].$$

Define

$$Q_{t_k,t_{k+1}}f(x) := \mathbb{E}[f(Y_{t_k,t_{k+1}}(x))], \;\; Q_{t_k,t_n} := Q_{t_k,t_{k+1}}Q_{t_{k+1},t_{k+2}}\cdots Q_{t_{n-1},t_n}, \;\; n \ge k+1.$$

Correspondingly, for any $s \ge 0$ and $x \in \mathbb{R}^d$, we let $\{X_{s,t}(x)\}_{t\ge s}$ solve the SDE

$$\mathrm{d}X_{s,t}(x) = b(X_{s,t}(x))\mathrm{d}t + \sigma\mathrm{d}B_t, \;\; t \ge s, X_{s,s}(x) = x.$$

Then the Markov semigroup P_t associated with (1.1) satisfies

$$P_{t-s}f(x) = P_{s,t}f(x) := \mathbb{E}[f(X_{s,t}(x))], \quad t \ge s \ge 0. \tag{2.16}$$

Let $Q_{0,0} = P_0$ be the identity operator. We have the domino decomposition

$$P_{t_n} - Q_{0,t_n} = \sum_{k=1}^n Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,t_n}, \;\; n \in \mathbb{N}.$$

Combining this with

$$\mathbb{W}_p\left(\mathscr{L}(X_{t_n}^x), \mathscr{L}(Y_{t_n}^x)\right) = \sup_{[f]_p\le 1}|\mathbb{E}[f(X_{t_n}^x) - f(Y_{t_n}^x)]| = \sup_{[f]_p\le 1}|P_{0,t_n}f(x) - Q_{0,t_n}f(x)|, \;\; n \ge 1,$$

we derive

$$\mathbb{W}_p\left(\mathscr{L}(X_{t_n}^x), \mathscr{L}(Y_{t_n}^x)\right) = \sup_{[f]_p\le 1}\left|\sum_{k=1}^n Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,t_n}f(x)\right|. \tag{2.17}$$

To prove Theorem 2.1 using this formula, we need the following derivative estimates on P_t.

Lemma 2.4. *Assume* (A0) *and* (A1). *There exist constants* $\kappa_1 = \kappa_1(K_1, K_2), \kappa_2 = \kappa_2(K_1, K_2) \in (0, \infty)$ *such that*

$$\|\nabla P_t f\|_\infty \le \kappa_1 \mathrm{e}^{-\kappa_2 t}\|\nabla f\|_\infty, \;\; t > 0, \tag{2.18}$$

$$\|\nabla P_t f\|_\infty \le \kappa_1 \mathrm{e}^{-\kappa_2 t}t^{-\frac{1}{2}}\|f\|_\infty, \;\; t > 0. \tag{2.19}$$

Moreover, when $\alpha \geq 1$,

$$\| \|\nabla^2 P_t f\|_{\text{op}}\|_\infty \leq \kappa_1 \mathrm{e}^{-\kappa_2 t}\big(t^{-\frac{1}{2}} + \sqrt{d}\, 1_{\{\alpha<2\}}\big)\|\nabla f\|_\infty, \ \ t>0. \tag{2.20}$$

Proof. (a) By [7, Corollary 2.3], $(A1)$ implies

$$\begin{aligned}\sup_{\|\nabla f\|_\infty \leq 1} |P_t f(x) - P_t f(y)| &= \mathbb{W}_1(\mathscr{L}(X_t^x), \mathscr{L}(X_t^y)) \\ &\leq \kappa_1 \mathrm{e}^{-\kappa_2 t}|x-y| \qquad \forall\ t\geq 0, x, y \in \mathbb{R}^d,\end{aligned}$$

for some constants $\kappa_1, \kappa_2 \in (0,\infty)$ depending only on K_1, K_2. Consequently, (2.18) holds.

Next, by [20, Theorem 3.4], there exists constant $k_1 \in (0,\infty)$ depending only on K_1, K_2 such that

$$\|\nabla_v P_t f\|_\infty \leq k_1 t^{-\frac{1}{2}}\|f\|_\infty |v|, \ \ t\in(0,1], \forall v \in \mathbb{R}^d.$$

Combining this with (2.18) and the semigroup property of P_t, we prove (2.19) for some constants $\kappa_1, \kappa_2 \in (0,\infty)$ depending only on K_1, K_2.

(b) When $\alpha = 2$, we have $\|\nabla b\|_{\text{op}} \vee \|\nabla^2 b\|_{\text{op}} \leq K_1$. In general, we let $\|\nabla^2 b\|_{\text{op}} \leq K_1'$ for some constant K_1' possibly different from K_1. Then for any $v, w \in \mathbb{R}^d$,

$$\phi_t(v) := \nabla_v X_t^x := \lim_{\varepsilon\downarrow 0}\frac{X_t^{x+\varepsilon v} - X_t^x}{\varepsilon}, \ \ \varphi_t(v,w) := \lim_{\varepsilon\downarrow 0}\frac{\nabla_v X_t^{x+\varepsilon w} - \nabla_v X_t^x}{\varepsilon}$$

exist and solve the equations

$$\begin{aligned}&\partial_t \phi_t(v) = \nabla_{\phi_t(v)} b(X_t^x), \quad \phi_0(v) = v,\\ &\partial_t \varphi_t(v,w) = \nabla_{\varphi_t(v,w)} b(X_t^x) + \nabla_{\phi_t(w)}\nabla_{\phi_t(v)} b(X_t^x), \ \ \varphi_0(v,w) = 0.\end{aligned}$$

Consequently,

$$\sup_{|v|\leq 1}|\phi_t(v)| \leq \mathrm{e}^{K_1}, \quad \sup_{|v|,|w|\leq 1}|\varphi_t(v,w)| \leq K_1'\mathrm{e}^{3K_1}, \quad t\in[0,1]. \tag{2.21}$$

Combining this with the Bismut–Elworthy–Li formula

$$\nabla_v P_t f(x) = \frac{1}{t}\mathbb{E}\bigg[f(X_t^x)\int_0^t \langle \sigma^{-1}\phi_s(v), \mathrm{d}B_s\rangle\bigg],$$

we derive

$$\nabla_w \nabla_v P_t f(x) = I_1 + I_2,$$

where

$$\begin{aligned}I_1 &:= \frac{1}{t}\mathbb{E}\bigg[\langle \nabla f(X_t^x), \phi_t(w)\rangle \int_0^t \langle \sigma^{-1}\phi_s(v), \mathrm{d}B_s\rangle\bigg],\\ I_2 &:= \frac{1}{t}\mathbb{E}\bigg[f(X_t^x)\int_0^t \langle \sigma^{-1}\varphi_s(v,w), \mathrm{d}B_s\rangle\bigg].\end{aligned} \tag{2.22}$$

By (2.21) and $\|\sigma^{-1}\|_{\text{op}} \leq K_1$, for any $x \in \mathbb{R}^d$ and $t \in (0,1]$,

$$\begin{aligned}&|I_1| \leq \|\nabla f\|_\infty K_1 \mathrm{e}^{2K_1} t^{-\frac{1}{2}},\\ &|I_2| = \frac{1}{t}\bigg|\mathbb{E}\bigg[\big\{f(X_t^x) - P_t f(x)\big\}\int_0^t \langle \sigma^{-1}\varphi_s(v,w), \mathrm{d}B_s\rangle\bigg]\bigg|\\ &\qquad \leq \frac{1}{t}\big(\mathbb{E}|f(X_t^x) - P_t f(x)|^2\big)^{\frac{1}{2}} K_1 K_1' \mathrm{e}^{3K_1}.\end{aligned} \tag{2.23}$$

Noting that (2.18) implies

$$
\begin{aligned}
\mathbb{E}|f(X_t^x) - P_tf(x)|^2 &= P_tf^2(x) - (P_tf(x))^2 \\
&= \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} P_s(P_{t-s}f)^2(x)\mathrm{d}s \\
&= \int_0^t P_s\left(L\left(P_{t-s}f\right)^2 - 2P_{t-s}f\cdot LP_{t-s}f\right)(x)\mathrm{d}s \\
&= \int_0^t P_s|\sigma^*\nabla P_{t-s}f|^2(x)\mathrm{d}s \le \|\nabla f\|_\infty^2 (K_1\kappa_1)^2 t,\ \ t>0,
\end{aligned}
\tag{2.24}
$$

where $L := \mathrm{tr}\{\sigma\sigma^*\nabla^2\} + b\cdot\nabla$ is the generator associated with (1.1). We derived (2.20) for $t\in(0,1]$ for some larger constant $\kappa_1 = \kappa_1(K_1,K_2)$ since $K_1' = K_1$ under $(A1)$. And for $t\in(1,\infty)$, the desired result follows from

$$
\begin{aligned}
\|\|\nabla^2 P_tf\|_{\mathrm{op}}\|_\infty &= \|\|\nabla^2 P_1P_{t-1}f\|_{\mathrm{op}}\|_\infty \\
&\le \kappa_1 e^{-\kappa_2}\|\nabla P_{t-1}f\|_\infty \le \kappa_1 e^{-\kappa_2 t}\|\nabla f\|_\infty,
\end{aligned}
$$

where the last inequality is a consequence of (2.18).

Now, let $\alpha\in[1,2)$. Let $\tilde b(x) = \mathbb{E}[b(x+B_1)]$. Then

$$
\begin{aligned}
&|\nabla_v\tilde b(x)| \le \mathbb{E}[|\nabla_v b(x+B_1)|] \le K_1,\ \ |v|\le 1,\\
&|b(x)-\tilde b(x)| \le K_1\mathbb{E}|B_1| \le K_1\sqrt{d}.
\end{aligned}
\tag{2.25}
$$

By the Bismut formula,

$$\nabla_v\tilde b(x) = \mathbb{E}\big[b(x+B_1)\langle v,B_1\rangle\big].$$

For $w\in\mathbb{R}^d$ with $|w|\le 1$, $\|\nabla b\|_{\mathrm{op}}\le K_1$ yields

$$\big|\nabla_w\nabla_v\tilde b(x)\big| = \left|\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\mathbb{E}\Big[\big\{b(x+B_1+\varepsilon w) - b(x+B_1)\big\}\langle v,B_1\rangle\Big]\right| \le K_1.$$

Let $\tilde P_t$ generated by $\tilde L := \mathrm{tr}\{\sigma\sigma^*\nabla^2\} + \tilde b\cdot\nabla$. By the result for $\alpha=2$ case, we can find a constant $c_1 = c_1(K_1)\in(0,\infty)$ such that the semigroup satisfies

$$\sup_{x\in\mathbb{R}^d,|v|\vee|w|\le 1,\|\nabla f\|_\infty\le 1}|\nabla_w\nabla_v\tilde P_tf(x)| \le c_1 t^{-\frac12},\ \ t\in(0,1]. \tag{2.26}$$

Combining this with (2.18), (2.25), $\|\nabla b\|_{\mathrm{op}}\le K_1$ and the formula

$$P_tf = \tilde P_tf + \int_0^t \tilde P_s\nabla_{b-\tilde b}P_{t-s}f\mathrm{d}s,$$

we derive that for any $x\in\mathbb{R}^d, |v|\vee|w|\le 1, t\in(0,1]$ and $\|\nabla f\|_\infty\le 1$,

$$
\begin{aligned}
|\nabla_w\nabla_vP_tf(x)| &= \left|\nabla_w\nabla_v\tilde P_tf(x) + \int_0^t \nabla_w\nabla_v\tilde P_s\nabla_{b-\tilde b}P_{t-s}f(x)\mathrm{d}s\right| \\
&\le c_1t^{-\frac12} + c_1\int_0^t s^{-\frac12}\|\nabla\nabla_{b-\tilde b}P_{t-s}f\|_\infty\mathrm{d}s.
\end{aligned}
$$

By denoting $\xi_t := \sup_{x\in\mathbb{R}^d, |v|\vee|w|\le 1, \|\nabla f\|_\infty\le 1} |\nabla_w\nabla_v P_t f(x)|$, we then have

$$\xi_t \le c_1 t^{-\frac{1}{2}} + c_1 K_1\sqrt{d}\int_0^t s^{-\frac{1}{2}}\xi_{t-s}\mathrm{d}s,\ \ t\in(0,1].$$

Let $a_t = \sup_{s\in[0,t]}\{s^{\frac{1}{2}}\xi_s\}$ and notice that $\int_0^1 s^{-\frac{1}{2}}(1-s)^{-\frac{1}{2}}\,\mathrm{d}s = B(1,1) = 1$ with $B(\cdot,\cdot)$ being the beta function. It follows that

$$\begin{aligned} a_t &\le c_1 + c_1\sqrt{d}K_1 t^{\frac{1}{2}}\left(\int_0^t s^{-\frac{1}{2}}(t-s)^{-\frac{1}{2}}\mathrm{d}s\right)a_t \\ &= c_1 + c_1\sqrt{d}K_1 t^{\frac{1}{2}} B(1,1)a_t. \end{aligned}$$

Solving this inequality yields, for any $t\in(0,t_0/2)$ with $t_0 = \frac{1}{(c_1K_1)^2 d}$,

$$a_t \le \frac{c_1}{1-c_1\sqrt{d}K_1 t^{\frac{1}{2}}} \le 2c_1,$$

which implies

$$\|\|\nabla^2 P_t f\|_{\mathrm{op}}\|_\infty = \xi_t\|\nabla f\|_\infty \le 2c_1 t^{-\frac{1}{2}}\|\nabla f\|_\infty.$$

Similar as before, for $t\in(t_0/2,\infty)$, the proof is finished by

$$\begin{aligned} \|\|\nabla^2 P_t f\|_{\mathrm{op}}\|_\infty &= \|\|\nabla^2 P_{\frac{t_0}{4}} P_{t-\frac{t_0}{4}} f\|_{\mathrm{op}}\|_\infty \\ &\le 4c_1 t_0^{-\frac{1}{2}}\|\nabla P_{t-\frac{t_0}{4}} f\|_\infty \le \kappa_1\sqrt{d}t^{-\frac{1}{2}}e^{-\kappa_2 t}\|\nabla f\|_\infty, \end{aligned}$$

for some $\kappa_1,\kappa_2\in(0,\infty)$. □

Lemma 2.5. *Assume* $(A0)$ *and* $(A1)$. *Then there exists a constant* $\kappa = \kappa(K_1, K_2, \eta, \alpha)$ *such that the following statements hold for any* $x\in\mathbb{R}^d$ *and* $k\in\mathbb{N}$.

(1) *When* $\alpha\in(0,1]$,

$$\begin{aligned} &\sup_{\|\nabla f\|_\infty\le 1}\left|Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,t}f(x)\right| \\ &\le \kappa d^{\frac{1}{2}}(1+|x|)\mathrm{e}^{-\kappa_2(t-t_k)}\eta_k^{1+\frac{\alpha}{2}},\quad t>t_k. \end{aligned}$$

(2) *When* $\alpha\in(1,2]$,

$$\begin{aligned} &\sup_{\|\nabla f\|_\infty\le 1}\left|Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,t}f(x)\right| \\ &\le \kappa d^{\frac{3}{2}}(1+|x|^2)\mathrm{e}^{-\kappa_2(t-t_k)}\left(\eta_k^{1+\frac{\alpha}{2}} + (t-t_k)^{-\frac{1}{2}}\eta_k^3\right),\quad t>t_k. \end{aligned}\tag{2.27}$$

Proof. (a) Let $\alpha\in(0,1]$. By Lemma 2.3 and and Lemma 2.2, we can find a constant $k_1 = k_1(K_1,K_2)\in(0,\infty)$ such that

$$\begin{aligned} &\mathbb{E}[|X_{t_{k-1},t}(Y^x_{t_{k-1}}) - Y^x_{t_{k-1}}| + |X_{t_{k-1},t}(Y^x_{t_{k-1}}) - Y^x_{t_{k-1}}|^\alpha] \\ &= \mathbb{E}\Big\{\left(\mathbb{E}[|X_{t_{k-1},t_k}(z) - z| + |X_{t_{k-1},t_k}(z)-z|^\alpha]\right)_{z=Y^x_{t_{k-1}}}\Big\} \\ &\le \kappa\left[\left(d^{\frac{1}{2}} + \mathbb{E}\left|Y^x_{t_{k-1}}\right|\right)\eta_k^{\frac{1}{2}} + \left(d^{\frac{\alpha}{2}} + \mathbb{E}\left|Y^x_{t_{k-1}}\right|^\alpha\right)\eta_k^{\frac{\alpha}{2}}\right] \\ &\le k_1 d^{\frac{1}{2}}(1+|x|)\eta_k^{\frac{\alpha}{2}},\quad t\in[t_{k-1},t_k]. \end{aligned}\tag{2.28}$$

Meanwhile, by (1.1), (1.2), (2.18) and $(A1)$, we obtain

$$
\begin{aligned}
&\sup_{\|\nabla f\|_\infty\le 1}\left|Q_{0,t_{k-1}}(P_{t_{k-1},t_k}-Q_{t_{k-1},t_k})P_{t_k,t}f(x)\right|\\
&\le \|\nabla P_{t_k,t}f\|_\infty \sup_{\|\nabla g\|_\infty\le 1}\left|Q_{0,t_{k-1}}(P_{t_{k-1},t_k}-Q_{t_{k-1},t_k})g(x)\right|\\
&\le \kappa_1 \mathrm{e}^{-\kappa_2(t-t_k)}\sup_{\|\nabla g\|_\infty\le 1}\left|\mathbb{E}[g(X_{t_{k-1},t_k}(Y^x_{t_{k-1}}))-g(Y_{t_{k-1},t_k}(Y^x_{t_{k-1}}))]\right|\\
&\le \kappa_1 \mathrm{e}^{-\kappa_2(t-t_k)}\mathbb{E}\left|X_{t_{k-1},t_k}(Y^x_{t_{k-1}}))-Y_{t_{k-1},t_k}(Y^x_{t_{k-1}})\right|\\
&=\kappa_1 \mathrm{e}^{-\kappa_2(t-t_k)}\int_{t_{k-1}}^{t_k}\mathbb{E}[|b(X_{t_{k-1},t}(Y^x_{t_{k-1}}))-b(Y^x_{t_{k-1}})|]\mathrm{d}t\\
&\le \kappa_1 K_1 \mathrm{e}^{-\kappa_2(t-t_k)}\int_{t_{k-1}}^{t_k}\mathbb{E}\Big\{\big(\mathbb{E}[|X_{t_{k-1},s}(z)-z|+|X_{t_{k-1},s}(z)-z|^\alpha]\big)_{z=Y^x_{t_{k-1}}}\Big\}\mathrm{d}s.
\end{aligned}
$$

Combining this with (2.28) we prove the first assertion.

(b) Let $\alpha\in(1,2]$. For any function g on $\mathbb{R}^d$ with $\|\nabla^i g\|_\infty<\infty, i=1,2$, we have

$$
\begin{aligned}
g(z)-g(y)&=\int_0^1\nabla_{z-y}g(y+r(z-y))\mathrm{d}r\\
&=\int_0^1\nabla_{z-y}g(y+r(z-y))-\nabla_{z-y}g(y)\mathrm{d}r+(\nabla_{z-y}g(y)-\nabla_{z-y}g(u))+\nabla_{z-y}g(u)\\
&=\int_0^1\int_0^1 r_1\nabla_{z-y}\nabla_{z-y}g(y+r_1r_2(z-y))\mathrm{d}r_2\mathrm{d}r_1\\
&\quad+\int_0^1\nabla_{y-u}\nabla_{z-y}g(u+r(y-u))\mathrm{d}r+\nabla_{z-y}g(u),\quad u,y,z\in\mathbb{R}^d.
\end{aligned}
$$

Let $\|\nabla f\|_\infty\le 1$. We shall apply this formula for

$$g=P_{t_k,t}f,\quad z=X_{t_{k-1},t_k}(Y^x_{t_{k-1}}),\quad y=Y_{t_{k-1},t_k}(Y^x_{t_{k-1}}),\quad u=Y^x_{t_{k-1}}.$$

Let

$$\Delta_k:=X_{t_{k-1},t_k}(Y^x_{t_{k-1}})-Y_{t_{k-1},t_k}(Y^x_{t_{k-1}}),\quad \tilde\Delta_k:=Y^x_{t_k}-Y^x_{t_{k-1}}. \tag{2.29}$$

By (2.16) and noting that $Y_{t_{k-1},t_k}(Y^x_{t_k-1})=Y^x_{t_k}$, we deduce from the above formula that

$$
\begin{aligned}
&\left|Q_{0,t_{k-1}}(P_{t_{k-1},t_k}-Q_{t_{k-1},t_k})P_{t_k,t}f(x)\right|\\
&=\left|\mathbb{E}\left[P_{t_k,t}f\left(X_{t_{k-1},t_k}(Y^x_{t_{k-1}})\right)\right]-\mathbb{E}\left[P_{t_k,t_n}f\left(Y_{t_{k-1},t_k}(Y^x_{t_{k-1}})\right)\right]\right|\\
&\le\left|\mathbb{E}\int_0^1\int_0^1 r_1\nabla_{\Delta_k}\nabla_{\Delta_k}P_{t_k,t}f\left(Y_{t_{k-1},t_k}(Y^x_{t_{k-1}})+r_1r_2\Delta_k\right)\mathrm{d}r_2\mathrm{d}r_1\right|\\
&\quad+\left|\mathbb{E}\int_0^1\nabla_{\tilde\Delta_k}\nabla_{\Delta_k}P_{t_k,t}f\left(Y^x_{t_{k-1}}+r\left(Y^x_{t_k}-Y^x_{t_{k-1}}\right)\right)\mathrm{d}r\right|\\
&\quad+\left|\mathbb{E}[\nabla_{\Delta_k}P_{t_k,t}f(Y^x_{t_{k-1}})]\right|\\
&:=J_1+J_2+J_3,
\end{aligned}
\tag{2.30}
$$

To estimate these terms, we need to bound Δ_k and $\tilde{\Delta}_k$. For $t \in [t_{k-1}, t_k]$, define

$$\Delta_{k,t} := X_{t_{k-1},t}(Y^x_{t_{k-1}}) - Y_{t_{k-1},t}(Y^x_{t_{k-1}}) = X_{t_{k-1},t}(Y^x_{t_{k-1}}) - Y^x_t, \tag{2.31}$$

$$\tilde{\Delta}_{k,t} := Y_{t_{k-1},t}(Y^x_{t_{k-1}}) - Y^x_{t_{k-1}} = Y^x_t - Y^x_{t_{k-1}}. \tag{2.32}$$

By Lemma 2.3, we can find a constant $C_0 = C_0(K_1, K_2, \eta, \alpha) \in (0, \infty)$ such that

$$\mathbb{E}|\tilde{\Delta}_{k,t}|^4 \le C_0 d^2 (1 + |x|^4)(t - t_{k-1})^2, \quad \forall t \in [t_{k-1}, t_k]. \tag{2.33}$$

We claim that there exist constants $C_i = C_i(K_1, K_2, \eta, \alpha) \in (0, \infty), i = 1, 2$, such that

$$\sup_{t \in [t_{k-1}, t_k]} \mathbb{E}|\Delta_{k,t}|^4 \le C_1 d^2 (1 + |x|^4)\eta_k^6, \tag{2.34}$$

$$\mathbb{E}\big|\mathbb{E}[\Delta_k | Y^x_{t_{k-1}}]\big| \le C_2 d^{\frac{\alpha}{2}} (1 + |x|^\alpha)\eta_k^{1+\frac{\alpha}{2}}. \tag{2.35}$$

Indeed, by (2.33) and that $(A1)$ for $\alpha \in (1, 2]$,

$$\begin{aligned}
\mathbb{E}|\Delta_{k,t}|^4 &= \mathbb{E}|X_{t_{k-1},t}(Y^x_{t_{k-1}}) - Y_{t_{k-1},t}(Y^x_{t_{k-1}})|^4 \\
&\le \eta_k^3 \int_{t_{k-1}}^t \mathbb{E}|b(X_{t_{k-1},s}(Y^x_{t_{k-1}})) - b(Y^x_{t_{k-1}})|^4 \mathrm{d}s \\
&\le 8K_1^2\eta_k^3 \int_{t_{k-1}}^t \mathbb{E}\big\{|X_{t_{k-1},s}(Y^x_{t_{k-1}}) - Y_{t_{k-1},s}(Y^x_{t_{k-1}})|^4 + |Y_{t_{k-1},s}(Y^x_{t_{k-1}}) - Y^x_{t_{k-1}}|^4\big\}\mathrm{d}s \\
&\le 8K_1^2\eta_k^3 \int_{t_{k-1}}^t \mathbb{E}|\Delta_{k,s}|^4 \mathrm{d}s + 8K_1^2 C_0 d^2(1 + |x|^4)\eta_k^6, \quad t \in [t_{k-1}, t_k].
\end{aligned}$$

By Grownwall's inequality, we find a constant $C_1 \in (0, \infty)$ depending on K_1, K_2, η and α, such that (2.34) holds. On the other hand, by $(A1)$ for $\alpha \in (1, 2]$, we have

$$\begin{aligned}
|\mathbb{E}[\Delta_k | Y_{t^x_{k-1}}]| &= |\mathbb{E}[X_{t_{k-1},t_k}(z) - Y_{t_{k-1},t_k}(z)]|_{z=Y_{t^x_{k-1}}} \\
&= \left| \int_{t_{k-1}}^{t_k} \mathbb{E}[b\left(X_{t_{k-1},t}(z)\right) - b(z)]\mathrm{d}t \right|_{z=Y_{t^x_{k-1}}} \\
&= \left| \int_{t_{k-1}}^{t_k} \int_0^1 \mathbb{E}\left[\nabla_{X_{t_{k-1},t}(z)-z} b\left(z + r(X_{t_{k-1},t}(z) - z)\right)\right] \mathrm{d}r\mathrm{d}t \right|_{z=Y_{t^x_{k-1}}} \\
&\le \left| \int_{t_{k-1}}^{t_k} \int_0^1 \mathbb{E}\left[\nabla_{X_{t_{k-1},t}(z)-z} \left(b\left(z + r(X_{t_{k-1},t}(z) - z)\right) - b(z)\right)\right] \mathrm{d}r\mathrm{d}t \right|_{z=Y_{t^x_{k-1}}} \\
&\quad + \left| \int_{t_{k-1}}^{t_k} \mathbb{E}[\nabla_{X_{t_{k-1},t}(z)-z} b(z)]\mathrm{d}t \right|_{z=Y_{t^x_{k-1}}} \\
&\le K_1 \int_{t_{k-1}}^{t_k} \left(\mathbb{E}|X_{t_{k-1},t}(z) - z|^\alpha\right)_{z=Y_{t^x_{k-1}}} \mathrm{d}t + K_1 \int_{t_{k-1}}^{t_k} \mathrm{d}t \int_{t_{k-1}}^t \mathbb{E}|b(X_{t_{k-1},s}(z))|_{z=Y_{t^x_{k-1}}} \mathrm{d}s.
\end{aligned}$$

So, by the linear growth of b, Lemma 2.2 and Lemma 2.3, we find constants $k_1, k_2 \in (0, \infty)$ depending on K_1, K_2, η and α, such that

$$\mathbb{E}|\mathbb{E}[\Delta_k | Y_{t^x_{k-1}}]| \le k_1 \eta_k^{1+\frac{\alpha}{2}} \mathbb{E}(d^{\frac{\alpha}{2}} + |Y^x_{t_{k-1}}|^\alpha)$$

$$\le k_2 d^{\frac{\alpha}{2}}(1+|x|^\alpha)\eta_k^{1+\frac{\alpha}{2}}.$$

Hence, (2.35) holds. By (2.33), (2.34), (2.35) and Lemma 2.4, we find constants $p_1, p_2 \in (0,\infty)$ depending only on K_1, K_2, η and α, such that $\|\nabla f\|_\infty \le 1$ implies

$$\begin{aligned} J_1 &\le \big\|\|\nabla^2 P_{t_k,t} f\|_{\rm op}\big\|_\infty \mathbb{E}|\Delta_k|^2 \\ &\le p_1 d^{\frac{3}{2}}(1+|x|^2)\mathrm{e}^{-\kappa_2(t-t_k)}(t-t_k)^{-\frac{1}{2}}\eta_k^3, \end{aligned}$$

and

$$\begin{aligned} J_3 &= \big|\mathbb{E}\langle \nabla P_{t_k,t} f(Y^x_{t_{k-1}}), \mathbb{E}[\Delta_k|Y^x_{t_{k-1}}]\rangle\big| \\ &\le \big\|\|\nabla P_{t_k,t} f\|_{\rm op}\big\|_\infty \mathbb{E}\big|\mathbb{E}[\Delta_k|Y^x_{t_{k-1}}]\big| \\ &\le p_2 d^{\frac{\alpha}{2}}(1+|x|^2)\mathrm{e}^{-\kappa_2(t-t_k)}\eta_k^{1+\frac{\alpha}{2}}, \quad t>t_k. \end{aligned}$$

Combining this with (2.30), we derive the desired estimate provided we find a constant $\kappa>0$ such that

$$J_2 \le \kappa d^{\frac{3}{2}}(1+|x|^2)\mathrm{e}^{-\kappa_2(t-t_k)}\eta_k^{2\wedge\frac{3\alpha}{2}}, \quad t>t_k. \tag{2.36}$$

(c) To verify (2.36), we apply the integration by parts formula for Malliavin's derivative. It suffices to prove for large k, say $k\ge 2$. Let

$$m := \inf\big\{i\in\mathbb{Z}^+ : \ t_k - t_i \le 3\eta_1\big\}.$$

We have $0\le m\le k-2$ and

$$\eta_1 \le t_{k-1} - t_m \le 3\eta_1. \tag{2.37}$$

Let $\{e_l\}_{1\le l\le d}$ be the canonical orthonormal basis in $\mathbb{R}^d$. For each $1\le l\le d$, let $h_l(0)=0$,

$$h_l'(t) := \sum_{i=m}^{k-1} 1_{[t_i,t_{i+1})}(t)\sigma^{-1}\Big(\frac{e_l}{t_k-t_m} - \frac{t_i-t_m}{t_k-t_m}\nabla_{e_l} b(Y^x_{t_i})\Big), \ \ t\in[0,t_k].$$

Since $h_l(t)$ is adapted with respect to the filtration generated by the Brownian motion, it is clear that (see [18, Proposition 1.3.11])

$$\delta(h_l) := \int_0^{t_k} \langle h_l'(t), \mathrm{d}W_2\rangle = \int_{t_m}^{t_k} \langle h_l'(t), \mathrm{d}W_2\rangle,$$

where δ is the divergence operator in Malliavin calculus. By (2.37), there exist constants $c_1, c_2>0$ depending only on K_1, η such that

$$|h_l'(t)| \le c_1 1_{[t_m,t_k]}(t), \ \ \mathbb{E}[\delta(h_l)^4] \le c_1\mathbb{E}\Big(\int_{t_m}^{t_k} |h_l'(t)|^2 \mathrm{d}t\Big)^2 \le c_2. \tag{2.38}$$

Let D_{h_l} be the Malliavin derivative along h_l. We claim that

$$D_{h_l} Y^x_{t_i} = \frac{t_i-t_m}{t_k-t_m} e_l, \quad m\le i\le k. \tag{2.39}$$

The formula is trivial for $i = m$. If it holds for some $m \le i \le k-1$, by the definition of Y_t^x, we have

$$\begin{aligned} D_{h_l} Y_{t_{i+1}}^x &= D_{h_l} Y_{t_i}^x + \eta_{i+1} D_{h_l} b(Y_{t_i}^x) + \sigma\left(h_l(t_{i+1}) - h_l(t_i)\right) \\ &= \frac{t_i - t_m}{t_k - t_m} e_l + \nabla_{e_l} b(Y_{t_i}^x) \frac{(t_{i+1} - t_i)(t_i - t_m)}{t_k - t_m} + \sigma(h_l(t_{i+1}) - h_l(t_i)) \\ &= \frac{t_{i+1} - t_m}{t_k - t_m} e_l, \end{aligned}$$

so that by induction we derive (2.39), which together with $(A1)$ and (2.37) yields

$$|D_{h_l}(Y_t^x - Y_{t_{k-1}}^x)| \le \left|\nabla_{D_{h_l} Y_{t_{k-1}}^x} b(Y_{t_{k=1}}^x)(t - t_{k-1})\right| + |\sigma(h_l(t) - h_l(t_{k-1}))| \le c_3 \eta_k \tag{2.40}$$

for some constant $c_3 > 0$ and all $t \in [t_{k-1}, t_k]$. In particular,

$$|D_{h_l} \tilde{\Delta}_k| = \left|D_{h_l} Y_{t_k}^x - D_{h_l} Y_{t_{k-1}}^x\right| \le c_3 \eta_k. \tag{2.41}$$

Moreover, $\Delta_{k,t}$ in (2.31) satisfies

$$D_{h_l} \Delta_{k,t} = D_{h_l} X_{t_{k-1},t}(Y_{t_{k-1}}^x) - D_{h_l} Y_t^x,$$

so that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} D_{h_l} \Delta_{k,t} &= \left(\nabla_{D_{h_l} X_{t_{k-1},t}(Y_{t_{k-1}}^x)} b\right)\left(X_{t_{k-1},t}(Y_{t_{k-1}}^x)\right) - \left(\nabla_{D_{h_l} Y_{t_{k-1}}^x} b\right)(Y_{t_{k-1}}^x) \\ &= \left(\nabla_{D_{h_l} \Delta_{k,t}} b\right)\left(X_{t_{k-1},t}(Y_{t_{k-1}}^x)\right) + \left(\nabla_{D_{h_l}(Y_t^x - Y_{t_{k-1}}^x)} b\right)\left(X_{t_{k-1},t}(Y_{t_{k-1}}^x)\right) \\ &\quad + \left(\nabla_{D_{h_l} Y_{t_{k-1}}^x} b\right)\left(X_{t_{k-1},t}(Y_{t_{k-1}}^x)\right) - \left(\nabla_{D_{h_l} Y_{t_{k-1}}^x} b\right)(Y_{t_{k-1}}^x) \\ &:= \left(\nabla_{D_{h_l} \Delta_{k,t}} b\right)\left(X_{t_{k-1},t}(Y_{t_{k-1}}^x)\right) + \mathcal{R}_k(t). \end{aligned}$$

Note that $D_{h_l} \Delta_{k,t_{k-1}} = 0$. So solving this ODE yields,

$$D_{h_l} \Delta_{k,t} = \int_{t_{k-1}}^t \exp\left(\int_s^t \nabla b\left(X_{t_{k-1},r}(Y_{t_{k-1}}^x)\right) \mathrm{d}r\right) \mathcal{R}_k(s) \mathrm{d}s, \quad \forall t \in [t_{k-1}, t_k].$$

Combining this with (2.34), (2.39), (2.40) and $(A1)$, we can find constants $c_4, c_5 > 0$ such that

$$\begin{aligned} \mathbb{E}|D_{h_l} \Delta_k|^2 &= \mathbb{E}|D_{h_l} \Delta_{k,t_k}|^2 \\ &\le \eta_k \int_{t_{k-1}}^{t_k} \exp\left(2K_1 \eta_k\right) \mathbb{E}\left|\mathcal{R}_k(s)\right|^2 \mathrm{d}s \\ &\le \eta_k \int_{t_{k-1}}^{t_k} K_1^2 \exp\left(2K_1 \eta_k\right) \mathbb{E}\left(\left|D_{h_l}(Y_s^x - Y_{t_{k-1}}^x)\right|^2 + |\Delta_{k,s}|^{2(\alpha-1)}\right) \mathrm{d}s \\ &\le c_4 \eta_k^2 \left(\eta_k^2 + \sup_{t \in [t_{k-1}, t_k]} \mathbb{E}|\Delta_{k,t}|^{2(\alpha-1)}\right) \\ &\le c_5 d^{\alpha-1}(1 + |x|^2)^{\alpha-1} \eta_k^{2+\min\{2, 3(\alpha-1)\}}. \end{aligned} \tag{2.42}$$

Let

$$\Xi_r := (1-r) Y_{t_{k-1}}^x + r Y_{t_k}^x, \quad r \in [0,1].$$

By (2.37) and (2.39), it is clear that

$$D_{h_l}\Xi_r = \Big(\frac{(1-r)(t_{k-1}-t_m)}{t_k-t_m} + r\Big)e_l,$$

and

$$\frac{(1-r)(t_{k-1}-t_m)}{t_k-t_m} + r \geq \frac{t_{k-1}-t_m}{t_k-t_m} \geq \frac{1}{3}.$$

So, for any function f with $\|\nabla f\|_\infty \leq 1$, by (2.18) and the integration by parts formula for Malliavin derivative (see, for instance, [18, Section 1.3] for more details), we then have

$$\begin{aligned}
&\big|\mathbb{E}[\nabla_{\tilde{\Delta}_k}\nabla_{\Delta_k}P_{t_k,t}f(\Xi_r)]\big| = \left|\sum_{j,l=1}^{d}\mathbb{E}\left[(\Delta_k)_l\,(\tilde{\Delta}_k)_j\nabla_{e_l}\nabla_{e_j}P_{t_k,t}f\,(\Xi_r)\right]\right| \\
&\leq 3\left|\sum_{j,l=1}^{d}\mathbb{E}\left[(\Delta_k)_l\,(\tilde{\Delta}_k)_j\nabla_{D_{h_l}\Xi_r}\nabla_{e_j}P_{t_k,t}f\,(\Xi_r)\right]\right| \\
&= 3\left|\sum_{j,l=1}^{d}\mathbb{E}\left[(\Delta_k)_l\,(\tilde{\Delta}_k)_j D_{h_l}\left\{\nabla_{e_j}P_{t_k,t}f\,(\Xi_r)\right\}\right]\right| \\
&= 3\left|\sum_{j,l=1}^{d}\mathbb{E}\left[\delta\left((\Delta_k)_l\,(\tilde{\Delta}_k)_j h_l\right)\nabla_{e_j}P_{t_k,t}f\,(\Xi_r)\right]\right| \\
&= 3\left|\sum_{j,l=1}^{d}\mathbb{E}\left[\left(-D_{h_l}\{(\Delta_k)_l\}(\tilde{\Delta}_k)_j - (\Delta_k)_l\,D_{h_l}\{(\tilde{\Delta}_k)_j\} + (\Delta_k)_l\,(\tilde{\Delta}_k)_j\delta(h_l)\right)\nabla_{e_j}P_{t_k,t}f\,(\Xi_r)\right]\right| \\
&= 3\left|\mathbb{E}\left[\nabla_{\tilde{\Delta}_k}P_{t_k,t}f\,(\Xi_r)\left(\sum_{l=1}^{d}D_{h_l}\{(\Delta_k)_l\}\right)\right]\right| + 3\left|\mathbb{E}\left[\sum_{l=1}^{d}(\Delta_k)_l\left(\nabla_{D_{h_l}\{\tilde{\Delta}_k\}}P_{t_k,t}f\,(\Xi_r)\right)\right]\right| \\
&\quad + 3\left|\mathbb{E}\left[\nabla_{\tilde{\Delta}_k}P_{t_k,t}f\,(\Xi_r)\left(\sum_{l=1}^{d}(\Delta_k)_l\,\delta(h_l)\right)\right]\right| \\
&:= T_1 + T_2 + T_3,
\end{aligned}$$

where $(\Delta_k)_j$ denotes the j-th component of $\Delta_k \in \mathbb{R}^d$. It follows from Cauchy-Schwarz inequality, (2.18), (2.33), (2.38), (2.41) and (2.42) that

$$\begin{aligned}
T_1 &\leq 3\kappa_1 \mathrm{e}^{-\kappa_2(t-t_k)}\left(\mathbb{E}|\tilde{\Delta}_k|^2\right)^{\frac{1}{2}}\left(\mathbb{E}\left(\sum_{l=1}^{d}D_{h_l}\{(\Delta_k)_l\}\right)^2\right)^{\frac{1}{2}} \\
&\leq c_6 \mathrm{e}^{-\kappa_2(t-t_k)} d^{\frac{\alpha+1}{2}}\left(1+|x|^2\right)^{\frac{\alpha}{2}}\eta_k^{\frac{5}{2}\wedge\frac{3\alpha}{2}},
\end{aligned}$$

$$\begin{aligned}
T_2 &\leq 3\kappa_1 \mathrm{e}^{-\kappa_2(t-t_k)}\left(\mathbb{E}|\Delta_k|^2\right)^{\frac{1}{2}}\left(\mathbb{E}\sum_{l=1}^{d}\left(D_{h_l}\{\tilde{\Delta}_k\}\right)^2\right)^{\frac{1}{2}} \\
&\leq c_6 \mathrm{e}^{-\kappa_2(t-t_k)} d(1+|x|^2)^{\frac{1}{2}}\eta_k^{\frac{5}{2}},
\end{aligned}$$

and

$$T_3 \le 3\kappa_1 \mathrm{e}^{-\kappa_2(t-t_k)} \left(\mathbb{E}|\tilde{\Delta}_k|^2\right)^{\frac{1}{2}} \left(\mathbb{E}|\Delta_k|^4\right)^{\frac{1}{4}} \left(d\sum_{l=1}^{d} \mathbb{E}\left[\delta(h_l)^4\right]\right)^{\frac{1}{4}}$$
$$\le c_6 \mathrm{e}^{-\kappa_2(t-t_k)} d^{\frac{3}{2}}(1+|x|^2)\eta_k^2,$$

for some constant $c_6 > 0$. Combining these estimates, we derive (2.36). □

Proof of Theorem 2.1. For $p \in (0,1)$, according to [13, Lemma 2.1] for $\psi(r) = r^p$ and $\mathbb{W}_0(\mu_1,\mu_2) = \frac{1}{2}\|\mu_1-\mu_2\|_{\mathrm{TV}}$, it holds for any two probability measures μ and ν that

$$\mathbb{W}_p(\mu,\nu) \le \inf_{t>0}\left\{2\sqrt{d}t^{\frac{p}{2}}\mathbb{W}_0(\mu,\nu) + dt^{\frac{p-1}{2}}\mathbb{W}_1(\mu,\nu)\right\}$$
$$\le 2^{2-p} d^{\frac{1+p}{2}} \mathbb{W}_0(\mu,\nu)^{1-p}\mathbb{W}_1(\mu,\nu)^p.$$

So, we only need to prove statements (1) and (2).

It is well known that P_t is ergodic and admits a unique invariant measure μ under $(A1)$ (see for instance [12]). By (2.18) and (2.19), we have

$$\begin{aligned}
\mathbb{W}_1(\mathscr{L}(X_t^x),\mu) &= \sup_{\|\nabla f\|_\infty<1}\left|P_t f(x) - \int_{\mathbb{R}^d} P_t f(y)\mu(\mathrm{d}y)\right| \\
&\le \sup_{\|\nabla f\|_\infty<1}\left|\int_{\mathbb{R}^d} \left(P_t f(x) - P_t f(y)\right)\mu(\mathrm{d}y)\right| \\
&\le \sup_{\|\nabla f\|_\infty<1} \|\nabla P_t f\|_\infty \int_{\mathbb{R}^d} |x-y|\mu(\mathrm{d}y) \\
&\le \kappa_1 \mathrm{e}^{-\kappa_2 t}\mu(|x-\cdot|) \le c_1(d^{\frac{1}{2}}+|x|)\mathrm{e}^{-\kappa_2 t},
\end{aligned} \tag{2.43}$$

where the last inequality is a consequence of the ergodicity and Lemma 2.2 and similarly,

$$\begin{aligned}
\mathbb{W}_0(\mathscr{L}(X_t^x),\mu) &= \sup_{\|f\|_\infty<1}\left|P_t f(x) - \int_{\mathbb{R}^d} P_t f(y)\mu(\mathrm{d}y)\right| \\
&\le \min\left\{1, \kappa_1 \mathrm{e}^{-\kappa_2 t} t^{-\frac{1}{2}}\mu(|x-\cdot|)\right\} \\
&\le c_1(d^{\frac{1}{2}}+|x|)\mathrm{e}^{-\kappa_2 t}, \qquad t>0,
\end{aligned}$$

for some constant $c_1 = c_1(K_1,K_2) \in (0,\infty)$. So we only need to prove the desired upper bound for $\mathbb{W}_i(\mathscr{L}(X_{t_n}^x),\mathscr{L}(Y_{t_n}^x)), i=1,0$. For any $z\in\mathbb{R}^d$, let $\mu_n^z := \mathscr{L}(X_{t_{n-1},t_n}(z)), \nu_n^z := \mathscr{L}(X_{t_{n-1},t_n}(z))$.

(a) Estimate on $\mathbb{W}_1$. When $\alpha \in (0,1]$, (2.1) follows from (2.17) and Lemma 2.5(1). When $\alpha \in (1,2]$, by (2.17) and Lemma 2.5(2) for $t = t_n$, we find a constant $\kappa = \kappa(K_1,K_2,\eta,\alpha) \in (0,\infty)$ such that for any $1 \le k \le n-1$,

$$\mathbb{W}_1(\mathscr{L}(X_{t_n}^x,\mathscr{L}(Y_{t_n}^x)) \le \kappa d^{\frac{3}{2}}(1+|x|^2)\sum_{k=1}^{n-1}\mathrm{e}^{-\kappa_2(t_n-t_k)}\eta_k^{1+\frac{\alpha}{2}}$$
$$+ \sup_{\|\nabla f\|_\infty\le 1}|Q_{0,t_{n-1}}(P_{t_{n-1},t_n} - Q_{t_{n-1},t_n})f(x)|.$$

Notice that

$$(2.44)\qquad \sum_{k=1}^{n-1} \mathrm{e}^{-\kappa_2(t_n-t_k)}\eta_k^{1+\frac{\alpha}{2}} \geq \eta_n^{\frac{\alpha}{2}} \sum_{k=1}^{n-1} \mathrm{e}^{-\kappa_2(t_n-t_k)}\eta_k \geq \eta_n^{\frac{\alpha}{2}} \int_0^{t_{n-1}} e^{-\kappa_2(t_n-t)}\mathrm{d}t \geq \frac{1}{\kappa_2}\eta_n^{\frac{\alpha}{2}}.$$

At the same time, (2.34) implies that, for any f satisfying $\|\nabla f\|_\infty \leq 1$,

$$\begin{aligned}\left|Q_{0,t_{n-1}}\left(P_{t_{n-1},t_n} - Q_{t_{n-1},t_n}\right) f(x)\right| &= \left|\mathbb{E}\left[f\left(X_{t_{n-1},t_n}(Y^x_{t_{n-1}})\right) - f(Y^x_{t_n})\right]\right| \\ &\leq \mathbb{E}\,|\Delta_n| \leq C_1 d^{\frac{1}{2}}(1+|x|^2)^{\frac{1}{2}}\eta_n^{\frac{3}{2}},\end{aligned}$$

where Δ_n is defined as in (2.29). Thus, according to (2.44), the last term can be absorbed by the sum term and (2.4) holds for a possibly larger constant $\kappa \in (0,\infty)$.

(b) Estimate on $\mathbb{W}_0$. By (2.19) and Lemma 2.5 for $t = t_n$, we can find a constant $k_0 = k_0(K_1, K_2, \eta, \alpha) \in (0,\infty)$ such that for any $1 \leq k \leq n-1$,

$$\begin{aligned}&\sup_{\|f\|_\infty\leq 1} |Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,t_n} f(x)| \\ &= \sup_{\|f\|_\infty\leq 1} |Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,\frac{t_k+t_n}{2}}\{P_{\frac{t_k+t_n}{2},t_n} f\}(x)| \\ &\leq \left(\sup_{\|f\|_\infty\leq 1} \|\nabla P_{\frac{t_k+t_n}{2},t_n} f\|_\infty\right) \sup_{\|\nabla g\|_\infty\leq 1} |Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,\frac{(t_k+t_n)}{2}} g(x)| \\ &\leq 2\kappa_1 \mathrm{e}^{-\kappa_2(t_n-t_k)/2}(t_n-t_k)^{-\frac{1}{2}} \sup_{\|\nabla g\|_\infty\leq 1} |Q_{0,t_{k-1}}(P_{t_{k-1},t_k} - Q_{t_{k-1},t_k})P_{t_k,\frac{(t_k+t_n)}{2}} g(x)| \\ &\leq 1_{(0,1]}(\alpha)\frac{k_0 d^{\frac{1}{2}}(1+|x|)}{\sqrt{t_n-t_k}}\mathrm{e}^{-\kappa_2(t_n-t_k)}\eta_k^{1+\frac{\alpha}{2}} + 1_{(1,2]}(\alpha)\frac{k_0 d^{\frac{3}{2}}(1+|x|^2)}{\sqrt{t_n-t_k}}\mathrm{e}^{-\kappa_2(t_n-t_k)}\eta_k^{1+\frac{\alpha}{2}}.\end{aligned}$$

Combining this with (2.17), we derive (2.2) and (2.5) provided there exists a positive constant $k_1 = k_1(K_1, K_2, \eta, \alpha)$ such that

$$(2.45)\qquad \sup_{\|f\|_\infty\leq 1} |Q_{0,t_{n-1}}(P_{t_{n-1},t_n} - Q_{t_{n-1},t_n})f(x)| \leq k_1(1+|x|)d^{\frac{1}{2}}\eta_n^{\frac{1+(\alpha\wedge 1)}{2}}, \quad n \geq 2.$$

To this end, for any $z \in \mathbb{R}^d$, let

$$\mu_n^z = \mathscr{L}(X_{t_{n-1},t_n}(z)), \quad \nu_n^z = \mathscr{L}(Y_{t_{n-1},t_n}(z)).$$

By Lemma 2.2, (2.45) follows if we can find constants $k_2 = k_2(K_1, K_2, \eta, \alpha) \in (0,\infty)$ such that

$$(2.46)\qquad \mathbb{W}_0(\mu_n^z, \nu_n^z) \leq k_2(d^{\frac{1}{2}} + |z|)\eta_n^{\frac{1+(\alpha\wedge 1)}{2}}, \quad n \geq 2, z \in \mathbb{R}^d.$$

Let us now show (2.45). Write

$$\mathrm{d}X_{t_{n-1},t}(z) = b(z)\mathrm{d}t + \sigma\mathrm{d}\tilde{B}_t, \quad X_{t_{n-1},t_{n-1}}(z) = z, \ t \in [t_{n-1}, t_n],$$

where

$$\tilde{B}_t = B_t - \int_{t_{n-1}}^t \left\{b(z) - b(X_{t_{n-1},s}(z))\right\}\mathrm{d}s, \quad t \in [t_{n-1}, t_n].$$

Let $R := \exp\left(\int_{t_{n-1}}^{t_n}\langle b(z)-b\left(X_{t_{n-1},s}(z)\right),\mathrm{d}B_s\rangle-\frac{1}{2}\int_{t_{n-1}}^{t_n}\left|b(z)-b\left(X_{t_{n-1},s}(z)\right)\right|^2\mathrm{d}s\right)$. In order to apply the Girsanov's theorem, we first show that $\mathbb{E}[R]=1$. By the continuity assumption of b in $(A1)$ and Young's inequality, there exist a positive constant $k_3=k_3(K_1,\alpha)$ such that

$$\left|b(X_{t_{n-1},s}(z))-b(z)\right|^2\le 4K_1^2\left|X_{t_{n-1},s}(z)-z\right|^{2\wedge 2\alpha}\le k_3\left(\left|X_{t_{n-1},s}(z)-z\right|^2+1\right). \tag{2.47}$$

So, according to [23, Ch.8,Exercise 1.40](another Novikov's type criterion), it suffices to show that,

$$\mathbb{E}\left[\exp\left(a\left|X_{t_{n-1},t}(z)-z\right|^2\right)\right]\le c, \tag{2.48}$$

for any $t\in[t_{n-1},t_n]$ and two constants a and c. It follows from Itô's formula, $(A1)$ and Young's inequality that, for any $t\in[t_{n-1},t_n]$,

$$\begin{aligned}\left|X_{t_{n-1},t}(z)-z\right|^2 &= 2\int_{t_{n-1}}^t\langle X_{t_{n-1},s}(z)-z,b(X_{t_{n-1},s}(z))\rangle\mathrm{d}s+2\int_{t_{n-1}}^t\langle X_{t_{n-1},s}(z)-z,\sigma\mathrm{d}B_s\rangle\\ &\quad+\|\sigma\|_{\mathrm{HS}}^2(t-t_{n-1})\\ &\le -2K_2\int_{t_{n-1}}^t\left|X_{t_{n-1},s}(z)-z\right|^2\mathrm{d}s+2\int_{t_{n-1}}^t\langle X_{t_{n-1},s}(z)-z,\sigma\mathrm{d}B_s\rangle\\ &\quad+2\int_{t_{n-1}}^t\left|X_{t_{n-1},s}(z)-z\right||b(z)|\,\mathrm{d}s+K_1d\eta_n\\ &\le -2\tilde{K}_2\int_{t_{n-1}}^t\left|X_{t_{n-1},s}(z)-z\right|^2\mathrm{d}s+2\int_{t_{n-1}}^t\langle X_{t_{n-1},s}(z)-z,\sigma\mathrm{d}B_s\rangle+k_4d\eta_n\end{aligned}$$

for some positive constant $k_4=k_4(K_1,K_2,b(z))$ and $\tilde{K}_2\in(0,K_2)$. As a consequence, for any $\gamma>0$, we have

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\gamma\left|X_{t_{n-1},t}(z)-z\right|^2+2\tilde{K}_2\gamma\int_{t_{n-1}}^t\left|X_{t_{n-1},s}(z)-z\right|^2\mathrm{d}s\right)\right]\\ &\le e^{k_4d\eta_n}\mathbb{E}\left[\exp\left(2\gamma\int_{t_{n-1}}^t\langle X_{t_{n-1},s}(z)-z,\sigma\mathrm{d}B_s\rangle\right)\right].\end{aligned}$$

Further more, Hölder's inequality , $(A1)$ and the local exponential martingale property implies that, for any fixed $\gamma\in(0,\frac{\tilde{K}_2}{4K_1^2})$ and $t\in[t_{n-1},t_n]$,

$$\begin{aligned}&\mathbb{E}\left[\exp\left(2\gamma\int_{t_{n-1}}^t\langle X_{t_{n-1},s}(z)-z,\sigma\mathrm{d}B_s\rangle\right)\right]\\ &\le\left(\mathbb{E}\left[\exp\left(4\gamma\int_{t_{n-1}}^t\langle X_{t_{n-1},s}(z)-z,\sigma\mathrm{d}B_s\rangle-8\gamma^2\int_{t_{n-1}}^t\left|\sigma\left(X_{t_{n-1},s}(z)-z\right)\right|^2\mathrm{d}s\right)\right]\right)^{\frac{1}{2}}\\ &\quad\times\left(\mathbb{E}\left[\exp\left(8\gamma^2\int_{t_{n-1}}^t\left|\sigma\left(X_{t_{n-1},s}(z)-z\right)\right|^2\mathrm{d}s\right)\right]\right)^{\frac{1}{2}}\\ &\le\left(\mathbb{E}\left[\exp\left(8\gamma^2\int_{t_{n-1}}^t\left|\sigma\left(X_{t_{n-1},s}(z)-z\right)\right|^2\mathrm{d}s\right)\right]\right)^{\frac{1}{2}}\end{aligned}$$

$$\leq \left(\mathbb{E}\left[\exp\left(8K_1^2\gamma^2\int_{t_{n-1}}^t \left|X_{t_{n-1},s}(z)-z\right|^2 \mathrm{d}s\right)\right]\right)^{\frac{1}{2}}$$

$$\leq \left(\mathbb{E}\left[\exp\left(\gamma\left|X_{t_{n-1},t}(z)-z\right|^2+2\tilde{K}_2\gamma\int_{t_{n-1}}^t \left|X_{t_{n-1},s}(z)-z\right|^2 \mathrm{d}s\right)\right]\right)^{\frac{1}{2}}.$$

Hence, (2.48) follows with $a=\gamma$, $c=e^{2k_4 d\eta_n}$ and R is a martingale.

Now, by Girsanov's theorem, $(\tilde{B}_t)_{t\in[t_{n-1},t_n]}$ is a Brownian motion under the probability measure $R\mathrm{d}\mathbb{P}$. Clearly, $\mathbb{E}[R-1]=0$. Notice that, by Cauchy-Schwarz inequality,

$$(\mathbb{E}|R-1|)^2=\left(\mathbb{E}\left[\frac{|R-1|}{\sqrt{1+\frac{R-1}{3}}}\sqrt{1+\frac{R-1}{3}}\right]\right)^2\leq \mathbb{E}\left[\frac{(R-1)^2}{1+\frac{R-1}{3}}\right]\mathbb{E}\left[1+\frac{R-1}{3}\right].$$

Combining this with the fact that

$$(1+x)\log(1+x)-x\geq\frac{1}{2}\left(\frac{x^2}{1+\frac{x}{3}}\right),\ \forall x>-1,$$

yields

$$\frac{1}{2}\left(\mathbb{E}|R-1|\right)^2=\frac{1}{2}\frac{(\mathbb{E}|R-1|)^2}{\mathbb{E}\left[1+\frac{R-1}{3}\right]}\leq\frac{1}{2}\mathbb{E}\left[\frac{(R-1)^2}{1+\frac{R-1}{3}}\right]\leq\mathbb{E}[R\log R],$$

which is also known as a Pinsker type inequality. Hence, combining this with the definition of $\mathbb{W}_0$, Girsanov's theorem and Lemma 2.3, we find $k_5=k_5(K_1,K_2,\eta,\alpha)\in(0,\infty)$ such that

$$\begin{aligned}
&\mathbb{W}_0(\mu_n^z,\nu_n^z)^2=\frac{1}{4}\sup_{\|f\|_\infty\leq 1}|\mu_n^z(f)-\nu_n^z(f)|^2\\
&=\frac{1}{4}\sup_{\|f\|_\infty\leq 1}\left|\mathbb{E}[f(X_{t_{n-1},t_n}(z))-Rf(X_{t_{n-1},t_n}(z))]\right|^2\leq\frac{1}{4}\left[\mathbb{E}|R-1|\right]^2\leq\frac{1}{2}\mathbb{E}[R\log R]\\
&=\frac{1}{4}\int_{t_{n-1}}^{t_n}\mathbb{E}[R|b(z)-b\left(X_{t_{n-1},s}(z)\right)|^2]\mathrm{d}s=\frac{1}{4}\int_{t_{n-1}}^{t_n}\mathbb{E}[|b(z)-b\left(Y_{t_{n-1},s}(z)\right)|^2]\mathrm{d}s\\
&\leq K_1^2\int_{t_{n-1}}^{t_n}\mathbb{E}\left[|Y_{t_{n-1},s}(z)-z|^2+|Y_{t_{n-1},s}(z)-z|^{2(1\wedge\alpha)}\right]\mathrm{d}s\leq c_2(d+|z|^2)\eta_n^{1+(1\wedge\alpha)}.
\end{aligned}$$

So, (2.46) holds.

(c) For the particular case, let $\eta_k=\frac{\theta}{k}$ for some $\theta>\frac{\alpha}{2c_2}$. Since $\theta\left(\log(n)-\log(k+1)\right)\leq t_n-t_k\leq\theta\left(\log(n-1)-\log(k)\right)$, we can find a positive constant $k_6=k_6(\theta,\alpha,c_2)$ (may vary from line to line) such that

$$\begin{aligned}
\sum_{k=[\frac{n-1}{2}]+1}^{n-1}\mathrm{e}^{-c_2(t_n-t_k)}(t_n-t_k)^{-\frac{1}{2}}\eta_k^{1+\frac{\alpha}{2}}&\leq k_4\sum_{k=[\frac{n-1}{2}]+1}^{n-1}\left(\frac{k}{n}\right)^{\theta c_2}\frac{1}{\sqrt{\log n-\log k}}k^{-1-\frac{\alpha}{2}}\\
&\leq k_4 n^{-\frac{\alpha}{2}}\sum_{k=1}^{n-1}\frac{1}{\sqrt{\log n-\log(k)}}k^{-1}
\end{aligned}$$

$$
\begin{aligned}
&\leq k_4 n^{-\frac{\alpha}{2}} \int_{\frac{n-1}{2}}^{n-1} \frac{1}{t\sqrt{\log n - \log(t)}} \mathrm{d}t \\
&\leq k_4 n^{-\frac{\alpha}{2}},
\end{aligned}
$$

and

$$
\sum_{k=1}^{[\frac{n-1}{2}]} \mathrm{e}^{-c_2(t_n-t_k)}(t_n-t_k)^{-\frac{1}{2}}\eta_k^{1+\frac{\alpha}{2}} \leq k_4 n^{-\theta c_2} \sum_{k=1}^{[\frac{n-1}{2}]} k^{-1+\theta c_2-\frac{\alpha}{2}} \leq k_4 n^{-\frac{\alpha}{2}}.
$$

So (2.6) follows by combining above estimates with (2.43). (2.3) can be proved through the same argument and the proof is complete. □

3. $\mathbb{W}_p$-ESTIMATE FOR $p>1$: THE UNIFORM DISSIPATION CASE

To cover typical time dependent models, see Example 3.1 below, we consider the following time in-homogenous SDE:

$$
\mathrm{d}X_t = b_t(X_t)\mathrm{d}t + \sigma_t(X_t)\mathrm{d}B_t, \quad t \geq 0, \tag{3.1}
$$

where $b: \mathbb{R}^d \times [0,\infty) \mapsto \mathbb{R}^d$, $\sigma: \mathbb{R}^d \times [0,\infty) \to \mathbb{R}^{d\times d}$ are measurable. The associated continuous time Euler Scheme is defined by

$$
Y_t = Y_{t_k} + (t-t_k)b_{t_k}(Y_{t_k}) + \sigma_{t_k}(Y_{t_k})\left(B_t - B_{t_k}\right), \qquad t \in [t_k, t_{k+1}), k \geq 0, Y_0 = X_0. \tag{3.2}
$$

3.1. Main result and an example of bridge regression.

($A2$) Let $\alpha \in (0,1]$ and $p \in (1,\infty)$. There exist positive constants K_1, K_2 such that for any $x, y \in \mathbb{R}^d$, $s, t \in [0,\infty)$

$$
|b_t(x) - b_s(y)| \leq K_1\left(|x-y|^\alpha + |x-y| + |t-s|\right), \tag{3.3}
$$

$$
\|\sigma_t(x) - \sigma_s(y)\|_{\mathrm{HS}} \leq K_1\left(|x-y| + |t-s|\right), \quad \|\sigma_t(x)\|_{\mathrm{op}} \leq K_1, \tag{3.4}
$$

$$
\begin{aligned}
&p\langle b_t(x) - b_t(y), x-y\rangle + \frac{p}{2}\|\sigma_t(x)-\sigma_t(y)\|_{\mathrm{HS}}^2 \\
&\quad + \frac{p(p-2)}{2|x-y|^2}\left|(\sigma_t(x)-\sigma_t(y))^*(x-y)\right|^2 \leq -K_2|x-y|^2.
\end{aligned} \tag{3.5}
$$

We call (3.5) the uniform dissipation condition.

Theorem 3.1. *Assume ($A0$) and ($A2$). Then (3.1) is well-posed . Moreover, for any $K_2' \in (0, K_2)$, there exists a positive constant $c = c(K_1, K_2, \eta, K_2', p, \alpha)$ such that*

$$
\mathbb{W}_p(\mathscr{L}(X_{t_n}^x), \mathscr{L}(Y_{t_n}^x))^p \leq c d^{\frac{p}{2}}(1+|x|^p) \sum_{k=1}^n \eta_k^{1+\frac{\alpha p}{2}} \mathrm{e}^{-K_2'(t_n-t_k)}, \quad n \geq 1, \forall x \in \mathbb{R}^d. \tag{3.6}
$$

Consequently, when $\eta_k = \frac{\theta}{k}$ for some constant $\theta \in (0,\infty)$, there exists a positive constant $c' = c'(K_1, K_2, \theta, K_2', p, \alpha)$ such that

$$
\mathbb{W}_p(\mathscr{L}(X_{t_n}^x), \mathscr{L}(Y_{t_n}^x))^p \leq c' d^{\frac{p}{2}}(1+|x|^p) n^{-\left((\theta K_2')\wedge\frac{\alpha p}{2}\right)}, \quad n \geq 1, \forall x \in \mathbb{R}^d. \tag{3.7}
$$

In particular, if $(\sigma_t, b_t) = (\sigma, b)$ does not depend on t, then the solution of (3.1) is exponentially ergodic with unique invariant probability measure μ, and

$$\mathbb{W}_p(\mathscr{L}(X_{t_n}^x), \mathscr{L}(Y_{t_n}^x))^p + \mathbb{W}_p(\mu, \mathscr{L}(Y_{t_n}^x))^p$$
$$\le c d^{\frac{p}{2}} (1 + |x|^p) \sum_{k=1}^n \eta_k^{1+\frac{\alpha p}{2}} \mathrm{e}^{-K_2'(t_n - t_k)}, \quad n \ge 1, \forall x \in \mathbb{R}^d.$$

As an application of Theorem 3.1, we consider the following optimization problem that arises in the Bridge regression with the shrinkage parameter $\gamma \in (1, 2]$ and the tuning parameter $\lambda \ge 0$ (see, i.e., [9] for more details):

$$\tilde{\beta} = \mathrm{argmin}_{\beta \in \mathbb{R}^d} \{L(\beta)\}, L(\beta) := \sum_{i=1}^N \left(y_i - x_i^T \beta\right)^2 + \lambda \sum_{j=1}^d |\beta_j|^\gamma \tag{3.8}$$

where $(x_i, y_i)_{\{1 \le i \le N\}}$ are $\mathbb{R}^d \times \mathbb{R}$-valued data points. In particular, when $\gamma = 1, 2$, $\tilde{\beta}$ corresponds to the estimators of the well-known Lasso and Ridge regression, respectively. In practical applications, such optimization problems are usually solved by gradient descent algorithm and its variants. One common variant is the gradient descent algorithm with slowly decreasing Gaussian noise given by the following iterative formula

$$\beta_{k+1} = \beta_k - \eta_{k+1} \nabla L(\beta_k) + \sqrt{\eta_{k+1}} \sigma_{k+1} \zeta_{k+1}, \quad k \ge 0, \tag{3.9}$$

where $\sigma_k \downarrow 0$ as $k \uparrow \infty$, and $\{\eta_k\}_{k \ge 1}$ are i.i.d. d-dimensional standard Gaussian random variables. Similarly, we may consider it as an approximation of the following SDE:

$$\mathrm{d}\bar{\beta}_t = -\nabla L(\bar{\beta}_t)\mathrm{d}t + \bar{\sigma}_t \mathrm{d}B_t, \quad \bar{\beta}_0 = \beta_0. \tag{3.10}$$

It has been shown in [11] that, under appropriate assumptions on the drift coefficient, if we set $\sigma_k = \frac{1}{\sqrt{\log \log k}}$ for large k and $\eta_k = \frac{\theta}{k}$ (correspondingly, $\bar{\sigma}_t \sim \frac{1}{\sqrt{\log t}}$ for large t), then this algorithm will converge in probability to $\tilde{\beta}$.

It is easy to verify that L is strongly convex, i.e.,

$$\langle \nabla L(\beta_1) - \nabla L(\beta_2), \beta_1 - \beta_2 \rangle \ge K|\beta_1 - \beta_2|^2, \quad \forall \beta_1, \beta_2 \in \mathbb{R}^d \tag{3.11}$$

holds for some constant $K = K(x, \gamma, \lambda) > 0$. Below we use Theorem 3.1 to analyze the convergence rate of this algorithm.

Example 3.1. Let $\tilde{\beta}, \beta_k$ and $\bar{\beta}_t$ be defined as in (3.8), (3.9) and (3.10) respectively. For any $p \ge 1$ and $K' \in (0, K)$, let $\sigma_k = \{k^{-\frac{K'\theta}{p}} (\theta \log k)^{-\frac{2}{p}} \wedge 1\}$ and $\eta_k = \frac{\theta}{k}$ for some $\theta > \frac{(\gamma-1)p}{2K'}$ with $\gamma > 1$. Then there exists $c > 0$ such that for large n,

$$\mathbb{E}|\beta_n - \tilde{\beta}|^p \le c d^{\frac{p}{2}} \left[|\beta_0 - \tilde{\beta}|^p + (1 + |\beta_0|^p) \right] n^{-\frac{(\gamma-1)p}{2}}.$$

Proof. By Jensen's inequality, it suffice to consider $p \ge 2$. To apply Theorem 3.1, let $\bar{\sigma}_t = \{\mathrm{e}^{-\frac{K'}{p}t} t^{-\frac{2}{p}} \wedge 1\}$. Consider the gradient flow $\tilde{\beta}_t$ defined by

$$\mathrm{d}\tilde{\beta}_t = -\nabla L(\tilde{\beta})\mathrm{d}t, \quad \tilde{\beta}_0 = \beta_0. \tag{3.12}$$

By Itô's formula, (3.11) and Young's inequality, for any $K' \in (0, K)$, we can find a positive constant c such that

$$\mathrm{d}|\bar{\beta}_t - \tilde{\beta}_t|^p - \mathrm{d}M_t$$

$$= \left[p|\bar{\beta}_t - \tilde{\beta}_t|^{p-2}\langle\bar{\beta}_t - \tilde{\beta}_t, -\nabla L(\bar{\beta}_t) + \nabla L(\tilde{\beta}_t)\rangle + \frac{1}{2}p(p-2+d)\bar{\sigma}_t^2|\bar{\beta}_t - \tilde{\beta}_t|^{p-2}\right] \mathrm{d}t$$
$$\leq \left[-K'|\bar{\beta}_t - \tilde{\beta}_t|^p + cd^{\frac{p}{2}}\bar{\sigma}_t^p\right] \mathrm{d}t,$$

which yields that

$$\begin{aligned}
\mathbb{E}|\bar{\beta}_t - \tilde{\beta}_t|^p &\leq e^{-K't}\left[|\bar{\beta}_0 - \tilde{\beta}|^p + cd^{\frac{p}{2}}\int_0^t e^{K_2's}\sigma_s^p \mathrm{d}s\right] \\
&\leq e^{-K't}\left[|\bar{\beta}_0 - \tilde{\beta}|^p + cd^{\frac{p}{2}}\left(\int_0^1 e^{K_2's}\mathrm{d}s + \int_1^t \frac{1}{s^2}\mathrm{d}s\right)\right] \\
&\leq e^{-K't}\left[|\bar{\beta}_0 - \tilde{\beta}|^p + c'd^{\frac{p}{2}}\right].
\end{aligned}$$

As $t_n \sim \theta\log(n)$ for large n and $K'\theta > \frac{(\gamma-1)p}{2}$, it follows that

$$\mathbb{W}_p(\mathscr{L}(\bar{\beta}_{t_n}), \delta_{\tilde{\beta}})^p = \mathbb{E}|\bar{\beta}_{t_n} - \tilde{\beta}|^p \leq \left[|\bar{\beta}_0 - \tilde{\beta}|^p + c'd^{\frac{p}{2}}\right] n^{-K_2'\theta} \leq \left[|\bar{\beta}_0 - \tilde{\beta}|^p + c'd^{\frac{p}{2}}\right] n^{-\frac{(\gamma-1)p}{2}}.$$

On the other hand, It can be verified that L and $\bar{\sigma}_t$ satisfy $(A2)$ with $\alpha = \gamma - 1$. Hence, Theorem 3.1 implies that, for any $K' \in (0, K), n \geq 1$,

$$\mathbb{W}_p(\mathscr{L}(\bar{\beta}_{t_n}), \mathscr{L}(\beta_n))^p \leq cd^{\frac{p}{2}}(1 + |\beta_0|^p)n^{-\frac{(\gamma-1)p}{2}}.$$

Using triangle inequality to combining above two upper bounds gives us the desired result. □

3.2. **Proof of Theorem 3.1.**

Similar to the previous section, we first present the following two lemmas regarding the moment estimates which can be proved through the same way as Lemma 2.2 and 2.3.

Lemma 3.2. *Assume that $(A0)$ holds, (3.1) is well-posed and there exist positive constants κ_1, κ_2 such that*

(3.13) $$\langle x, b_t(x)\rangle \leq \kappa_1 - \kappa_2|x|^2, \ \ |b_t(x)| \leq \kappa_2(1 + |x|), \ \ \|\sigma_t(x)\|_{\mathrm{op}} \leq \kappa_2, \ \ \forall x \in \mathbb{R}^d.$$

Then for any $p \in (0, \infty)$, there exists a constant $\kappa = \kappa(\kappa_1, \kappa_2, p) \in (0, \infty)$ such that

(3.14) $$\sup_{t\geq 0}\mathbb{E}|X_t^x|^p \leq \kappa(d^{\frac{p}{2}} + |x|^p), \ \ \forall x \in \mathbb{R}^d.$$

If moreover

(3.15) $$|b_t(x) - b_s(y)| \leq K(|x-y| + |x-y|^\alpha + |t-s|), \ \ \forall x, y \in \mathbb{R}^d$$

holds for some constant $K \in (0, \infty)$, then there exists $\kappa' = \kappa'(\kappa_1, \kappa_2, K, \eta, \alpha, p) \in (0, \infty)$ such that

(3.16) $$\sup_{t\geq 0}\mathbb{E}|Y_t^x|^p \leq \kappa'(d^{\frac{p}{2}} + |x|^p), \ \ \forall x \in \mathbb{R}^d.$$

Lemma 3.3. *Assume that the conditions in Lemma 3.2 hold. Then, for any $p > 0$ there exists a constant $\kappa = \kappa(\kappa_1, \kappa_2, p) \in (0, \infty)$ such that*

$$\mathbb{E}|X_t^x - x|^p \leq \kappa(d^{\frac{p}{2}} + |x|^p)(1 \wedge t)^{\frac{p}{2}\wedge 1}, \ \ x \in \mathbb{R}^d, \ \ t \geq 0.$$

If moreover (3.15) *holds, then there exists* $\kappa' = \kappa'(\kappa_1, \kappa_2, \eta, K, p, \alpha) \in (0, \infty)$ *such that*

$$\mathbb{E}|Y_t^x - Y_{t_{k-1}}^x|^p \le \kappa' d^{\frac{p}{2}}(1+|x|^p)\eta_k^{\frac{p}{2}}, \ \ x \in \mathbb{R}^d, \ k \ge 1, \ t \in [t_{k-1}, t_k].$$

Proof of Theorem 3.1. All constants below depend only on $K_1, K_2, K_2', \eta, \alpha$ and p.

(a) In the uniform dissipation case, the well-posedness is well known, see for instance [2]. Let $k \ge 0$. By [24, Theorem 4.1], we can choose $\mathscr{F}_{t_k}$-measurable random variables $\bar{X}_{t_k}$ and $\bar{Y}_{t_k}$ such that

$$\mathscr{L}(X_{t_k}^x) = \mathscr{L}(\bar{X}_{t_k}), \ \ \mathscr{L}(Y_{t_k}^x) = \mathscr{L}(\bar{Y}_{t_k}), \ \ \mathbb{W}_p(\mathscr{L}(X_{t_k}^x), \mathscr{L}(Y_{t_k}^x))^p = \mathbb{E}|\bar{X}_{t_k} - \bar{Y}_{t_k}|^p. \tag{3.17}$$

We consider the SDEs for $t \in [t_k, t_{k+1}]$,

$$\begin{aligned} \mathrm{d}\bar{X}_t &= b_t(\bar{X}_t)\mathrm{d}t + \sigma_t(\bar{X}_t)\mathrm{d}B_t, \\ \mathrm{d}\bar{Y}_t &= b_{t_k}(\bar{Y}_{t_k})\mathrm{d}t + \sigma_{t_k}(\bar{Y}_{t_k})\mathrm{d}B_t. \end{aligned}$$

By the weak uniqueness and the definition of $\mathbb{W}_p$, we obtain

$$\mathbb{W}_p\big(\mathscr{L}(X_{t_{k+1}}^x), \mathscr{L}(Y_{t_{k+1}}^x)\big)^p \le \mathbb{E}|\bar{X}_{t_{k+1}} - \bar{Y}_{t_{k+1}}|^p. \tag{3.18}$$

Let $Z_t = \bar{X}_t - \bar{Y}_t$. By $(A2)$ and Itô's formula, we find a martingale $(M_t)_{t\in[t_k, t_{k+1}]}$ such that

$$\begin{aligned}
&\mathrm{d}|Z_t|^p - \mathrm{d}M_t \\
&= |Z_t|^{p-2}\Big[p\langle Z_t, b_t(\bar{X}_t) - b_{t_k}(\bar{Y}_{t_k})\rangle + \frac{p}{2}\|\sigma_t(\bar{X}_t) - \sigma_{t_k}(\bar{Y}_{t_k})\|_{\mathrm{HS}}^2 \\
&\qquad + \frac{1}{2}p(p-2)\frac{|\{\sigma_t(\bar{X}_t) - \sigma_{t_k}(\bar{Y}_{t_k})\}^* Z_t|^2}{|Z_t|^2}\Big]\mathrm{d}t \\
&\le |Z_t|^{p-2}\Big[p\langle Z_t, b_t(\bar{X}_t) - b_t(\bar{Y}_t)\rangle + \frac{p}{2}\|\sigma_t(\bar{X}_t) - \sigma_t(\bar{Y}_t)\|_{\mathrm{HS}}^2 \\
&\qquad + \frac{1}{2}p(p-2)\frac{|\{\sigma_t(\bar{X}_t) - \sigma_t(\bar{Y}_t)\}^* Z_t|^2}{|Z_t|^2} \\
&\qquad + p|Z_t||b_t(\bar{Y}_t) - b_{t_k}(\bar{Y}_{t_k})| + \frac{1}{2}p(p-1)\big\{\|\sigma_t(\bar{Y}_t) - \sigma_{t_k}(\bar{Y}_{t_k})\|_{\mathrm{HS}}^2\big\}\Big]\mathrm{d}t, \\
&\le -K_2|Z_t|^p\mathrm{d}t + pK_1|Z_t|^{p-1}\Big(|\bar{Y}_t - \bar{Y}_{t_k}|^\alpha + |\bar{Y}_t - \bar{Y}_{t_k}| + |t - t_k|\Big)\mathrm{d}t \\
&\quad + p(p-1)K_1^2|Z_t|^{p-2}\left[|\bar{Y}_t - \bar{Y}_{t_k}|^2 + (t-t_k)^2\right]\mathrm{d}t, \ \ t \in [t_k, t_{k+1}].
\end{aligned}$$

By Young's inequality, for any fixed $K_2' \in (0, K_2)$, we find a constant $c_1 > 0$ such that

$$\mathrm{d}|Z_t|^p - \mathrm{d}M_t \le -K_2'|Z_t|^p\mathrm{d}t + c_1\Big(|\bar{Y}_t - \bar{Y}_{t_k}|^{p\alpha} + |\bar{Y}_t - \bar{Y}_{t_k}|^p + \eta_{k+1}^p\Big)\mathrm{d}t, \ \ t \in [t_k, t_{k+1}],$$

which further implies that, for $t \in [t_k, t_{k+1}]$

$$\mathbb{E}|Z_t|^p \le \mathrm{e}^{-K_2'(t-t_k)}\mathbb{E}|Z_{t_k}|^p + c_1\mathbb{E}\int_{t_k}^t \Big(|\bar{Y}_s - \bar{Y}_{t_k}|^{\alpha p} + |\bar{Y}_s - \bar{Y}_{t_k}|^p + \eta_{k+1}^p\Big)\mathrm{d}s. \tag{3.19}$$

By Lemma 3.3, Lemma 3.2 and $\mathscr{L}(\bar{Y}_{t_k}) = \mathscr{L}(Y_{t_k}^x)$, we can find a constant $c_2 > 0$ such that

$$\mathbb{E}|\bar{Y}_s - \bar{Y}_{t_k}|^p = \mathbb{E}\left[|Y_{s-t_k}^z - z|^p\big|z = Y_{t_k}^x\right] \le c_2(d\eta_{k+1})^{\frac{p}{2}}(1+|x|^p), \ \ s \in [t_k, t_{k+1}], \tag{3.20}$$

so that by Jensen's inequality and (3.19), we find a constant $c_3 > 0$ such that

$$\mathbb{E}|Z_t|^p \le \mathrm{e}^{-K_2'(t-t_k)}\mathbb{E}|Z_{t_k}|^p + c_3 d^{\frac{p}{2}}\eta_{k+1}^{1+\frac{p\alpha}{2}}(1+|x|^p),\ \ t\in[t_k,t_{k+1}].$$

Combining this with (3.17) and (3.18), we derive

$$\mathbb{W}_p\big(\mathscr{L}(X^x_{t_{k+1}}),\mathscr{L}(Y^x_{t_{k+1}})\big)^p \le \mathrm{e}^{-K_2'\eta_{k+1}}\mathbb{W}_p\big(\mathscr{L}_{X^x_{t_k}},\mathscr{L}_{Y^x_{t_k}}\big)^p + c_3 d^{\frac{p}{2}}\eta_{k+1}^{1+\frac{p\alpha}{2}}(1+|x|^p),\ \ k\ge 0, x\in\mathbb{R}^d.$$

Iterating in k we prove the desired upper bound estimate in (3.6).

(b) Let $\eta_k = \frac{\theta}{k}(k\ge 1)$ for some constant $\theta\in(0,\infty)$. Then there exists a constant $c_4\in(0,\infty)$ such that

$$\theta\log\frac{n}{k} - c_4\theta \le t_n - t_k \le \theta\log\frac{n}{k} + \theta c_4,\ \ 1\le k\le n.$$

By this and (3.6), when $\theta\ne\frac{\alpha p}{2K_2'}$, we can find constants $c_5, c_6\in(0,\infty)$ such that

$$\mathbb{W}_p\big(\mathscr{L}(X^x_{t_n}),\mathscr{L}(Y^x_{t_n})\big)^p \le c_4(1+|x|^p)n^{-\theta K_2'}\sum_{k=1}^n k^{-1-\frac{\alpha p}{2}+\theta K_2'}$$

$$\le c_5(1+|x|^p)n^{-\left((\theta K_2')\wedge\frac{\alpha p}{2}\right)},\ \ n\ge 1, x\in\mathbb{R}^d.$$

This implies (3.7) when $\theta\ne\frac{\alpha p}{2K_2'}$. If $\theta=\frac{\alpha p}{2K_2'}$, we may apply the above estimate for $K_2''\in(K_2',K_2)$ replacing K_2', so that (3.7) holds as well.

(c) Let $(\sigma_t,b_t)=(\sigma,b)$ does not depend on t. The exponential ergodicity can be proved in a standard way [12]. By $(A2)$, synchronuous coupling and Itô's formula, for any $x,y\in\mathbb{R}^d$ we can find a martingale M_t such that

$$\mathrm{d}|X^x_t - X^y_t|^p \le -K_2|X^x_t - X^y_t|^p\mathrm{d}t + \mathrm{d}M_t,\ \ t\ge 0.$$

So, $\mathbb{E}|X^x_t - X^y_t|^p \le |x-y|^p\mathrm{e}^{-K_2 t}, t\ge 0$. By the Markov property, this implies, for any $\mu_1,\mu_2\in\mathscr{P}$,

$$\begin{aligned}\mathbb{W}_p(P_t^*\mu_1,P_t^*\mu_2)^p &\le \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)}\int_{\mathbb{R}^d\times\mathbb{R}^d}\mathbb{E}\left[|X^x_t - X^y_t|^p\right]\pi(\mathrm{d}x,\mathrm{d}y)\\ &\le \inf_{\pi\in\mathscr{C}(\mu_1,\mu_2)}\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^p\, e^{-K_2 t}\pi(\mathrm{d}x,\mathrm{d}y) = \mathbb{W}_p(\mu_1,\mu_2)^p\mathrm{e}^{-K_2 t},\end{aligned}\tag{3.21}$$

where the first inequality is by $P_t^*\nu=\mathscr{L}(X_t)$ for X_t solving (1.1) with initial distribution $\nu\in\mathbb{P}$. Moreover, let L be the generator associated with (1.1) given by

$$Lf = \langle b,\nabla f\rangle + \frac{1}{2}\langle\sigma\sigma^T,\nabla^2 f\rangle_{\mathrm{HS}},\quad \forall f\in C^2\left(\mathbb{R}^d;\mathbb{R}\right).\tag{3.22}$$

Then $(A2)$ implies that

$$L|\cdot|^{p\vee 2} \le c_1 d^{\frac{p\vee 2}{2}} - c_2|\cdot|^p$$

for some constants $c_1,c_2>0$. By a standard tightness argument, this implies that P_t^* has an invariant probability measure μ with $\mu(|\cdot|^p)\le\frac{c_1 d^{\frac{p\vee 2}{2}}}{c_2}<\infty$. Combining this with (3.21) we conclude that μ is the unique invariant probability measure of P_t^*, and there exists a constant $c_0'\in(0,\infty)$ such that

$$\mathbb{W}_p(\mathscr{L}(X^x_t),\mu)^p = \mathbb{W}_p(P_t^*\delta_x,\mu)^p \le \mu(|x-\cdot|^p)\mathrm{e}^{-K_2 t} \le c_0'(1+|x|^p)\mathrm{e}^{-K_2 t},\ \ x\in\mathbb{R}^d, t\ge 1.\tag{3.23}$$

Combining this together with (3.6) and the triangle inequality, implies the desired upper bound for $\mathbb{W}_p(\mathscr{L}(Y_{t_n}^x), \mu)^p$. $\square$

References

[1] Alfonsi, A., Jourdain, B., and Kohatsu-Higa, A. (2015). Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme. *Electron. J. Probab. 20* (2015), no. 70, 1–31. 2

[2] Cerrai, S. (Ed.). (2001). Second order PDE's in finite and infinite dimension: a probabilistic approach. Berlin, Heidelberg: Springer Berlin Heidelberg. 25

[3] Peng Chen, Qi-Man Shao, and Lihu Xu. A probability approximation framework: Markov process approach. *The Annals of Applied Probability, 33*(2), 1619-1659, April 2023. 3

[4] Dalalyan, A. S. (2017). Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 651-676. 1, 2

[5] Durmus, A., and Moulines, E. (2017). Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability,27*(3), 1551–1587. 1, 2, 3

[6] Durmus, A., and Moulines, E. (2019). High-dimensional Bayesian inference via the unadjusted Langevin algorithm. *Bernoulli 25*(4A), 2019, 2854–2882. 1, 2, 3

[7] Eberle, A. (2016). Reflection couplings and contraction rates for diffusions. *Probability theory and related fields, 166*(3), 851-886. 10

[8] E, W., Han, J. and Jentzen, A. Deep Learning-Based Numerical Methods for High-Dimensional Parabolic Partial Differential Equations and Backward Stochastic Differential Equations. *Commun. Math. Stat*. **5**, 349–380 (2017). https://doi.org/10.1007/s40304-017-0117-6 1

[9] Fu, W. J. (1998). Penalized regressions: the bridge versus the lasso. *Journal of computational and graphical statistics, 7*(3), 397-416. 23

[10] Gyöngy, I., and Martínez, T. (2001). On stochastic differential equations with locally unbounded drift. *Czechoslovak Mathematical Journal, 51*(4), 763-783. 1

[11] Gelfand, S. B., and Mitter, S. K. (1993). Metropolis-type annealing algorithms for global optimization in $\mathbb{R}^d$. *SIAM Journal on Control and Optimization, 31*(1), 111-131. 23

[12] Hairer, M. (2010). Convergence of Markov processes. *Lecture notes*. 18, 26

[13] X. Huang, P. Ren, F.-Y. Wang, *Probability distance estimates between diffusion processes and applications to singular McKean-Vlasov SDEs,* arXiv:2304.07562. 18

[14] Lamberton, D., and Pages, G. (2002). Recursive computation of the invariant distribution of a diffusion. *Bernoulli*, 367-405.

[15] Lamberton, D., and Pages, G. (2003). Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. *Stochastics and dynamics, 3*(04), 435-451.

[16] Mei, S., Montanari, A., and Nguyen, P. M. (2018). A mean field view of the landscape of two-layer neural networks. *Proceedings of the National Academy of Sciences, 115*(33), E7665-E7671. 1

[17] Mou, W., Flammarion, N., Wainwright, M. J., and Bartlett, P. L. (2022). Improved bounds for discretization of Langevin diffusions: Near-optimal rates without convexity. *Bernoulli, 28*(3), 1577-1601. 2

[18] Nualart, D. (2006). *The Malliavin calculus and related topics* (Vol. 1995, p. 317). Berlin: Springer. 15, 17

[19] Pages, G., and Panloup, F. (2023). Unadjusted Langevin algorithm with multiplicative noise: Total variation and Wasserstein bounds. *The Annals of Applied Probability, 33*(1), 726-779. 1, 3, 4

[20] Priola, E., and Wang, F. Y. (2006). Gradient estimates for diffusion semigroups with singular coefficients. *Journal of Functional Analysis, 236*(1), 244-264. 10

[21] Ren, P. (2023). Singular McKean–Vlasov SDEs: Well-posedness, regularities and Wang's Harnack inequality. *Stochastic Processes and their Applications, 156*, 291-311. 1, 4

[22] Roberts, G. O., and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 341-363. 2

[23] Revuz, D., and Yor, M. (2013). *Continuous martingales and Brownian motion* (Vol. 293). Springer Science and Business Media. 20

[24] Villani, C. (2009). *Optimal transport: old and new* (Vol. 338, p. 23). Berlin: springer. 25

[25] F.-Y. Wang, *Exponential Ergodicity for singular reflecting McKean-Vlasov SDEs,* Stoch. Proc. Apll. 160(2023), 265–293. 1

[26] Xie, L., and Zhang, X. (2016). Sobolev differentiable flows of SDEs with local Sobolev and super-linear growth coefficients. *The Annals of Probability, 44*(6), 3661-3687. 1

[27] L. Xie, X. Zhang, *Ergodicity of stochastic differential equations with jumps and singular coefficients,* Ann. Inst. Henri Poincaré Probab. Stat. 56(2020), 175–229. 1

[28] Zhang, X. (2005). Strong solutions of SDES with singular drift and Sobolev diffusion coefficients. *Stochastic Processes and their Applications, 115*(11), 1805-1818. 1

[29] X. Zhang, *Stochastic homeomorphism flows of SDEs with singular drifts and Sobolev diffusion coefficients,* Electr. J. Probab. 16(2011), 1096–1116. 4

(X. Li) 1. Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau S.A.R., China. 2. Zhuhai UM Science & Technology Research Institute, Zhuhai, China

Email address: yc07904@um.edu.mo

(F. Y. Wang) Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

Email address: wangfy@tju.edu.cn

(L. Xu) 1. Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau S.A.R., China. 2. Zhuhai UM Science & Technology Research Institute, Zhuhai, China

Email address: lihuxu@um.edu.mo