STRONG AND WEAK CONVERGENCE FOR AVERAGING PRINCIPLE OF DDSDE WITH SINGULAR DRIFT

MENGYU CHENG, ZIMO HAO, MICHAEL RÖCKNER

Abstract

In this paper, we study the averaging principle for distribution dependent stochastic differential equations with drift in localized L^p spaces. Using Zvonkin's transformation and estimates for solutions to Kolmogorov equations, we prove that the solutions of the original system strongly and weakly converge to the solution of the averaged system as the time scale ε goes to zero. Moreover, we obtain rates of the strong and weak convergence that depend on p respectively.

Contents

1. Introduction

Consider the following distribution dependent stochastic differential equation (in short, DDSDE):

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \Sigma(t, X_t, \mu_t)\mathrm{d}W_t, \tag{1.1}$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$ and $\Sigma : \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d \otimes \mathbb{R}^d$ are measurable functions and $\mu_t := \mathcal{L}(X_t)$ is the time marginal law of X_t. Here $\mathcal{P}(\mathbb{R}^d)$ denotes the space of all probability

Keywords: Averaging principle, Distribution dependent SDE, Heat kernel.

This work is supported by the China Scholarship Council (CSC 202006060083), NNSFC grant of China (No. 12131019, 11731009) and the German Research Foundation (DFG) through the Collaborative Research Centre (CRC) 1283/2 2021 - 317210226 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications".

measures over $\mathbb{R}^d$ and $(W_t)_{t\geqslant 0}$ is a d-dimensional standard Brownian motion on some stochastic basis $(\Omega, \mathscr{F}, \mathbf{P}; (\mathscr{F}_t)_{t\geqslant 0})$. The Kolmogorov operator of (1.1) is given by

$$\mathscr{L}_t\varphi(x) = A_{ij}(t, x, \mu_t)\partial_i\partial_j\varphi(x) + b(t, x, \mu_t) \cdot \nabla\varphi(x),$$

where $A_{ij}(t, x, \mu_t) := \frac{1}{2}\Sigma_{ik}\Sigma_{jk}(t, x, \mu_t)$. Here and below we use the usual Einstein convention for summation.

Note that DDSDE (1.1) is also called mean-field SDE or McKean-Vlasov SDE in the literature if

$$b(t, x, \mu) := \int_{\mathbb{R}^d} \hat{b}(t, x, y)\mu(\mathrm{d}y), \quad \Sigma(t, x, \mu) := \int_{\mathbb{R}^d} \hat{\Sigma}(t, x, y)\mu(\mathrm{d}y) \tag{1.2}$$

for all $(t, x, \mu) \in \mathbb{R}_+ \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, where $\hat{b} : \mathbb{R}_+ \times \mathbb{R}^{2d} \to \mathbb{R}^d$ and $\hat{\Sigma} : \mathbb{R}_+ \times \mathbb{R}^{2d} \to \mathbb{R}^d \otimes \mathbb{R}^d$ are measurable functions. It naturally appears in the studies on the limiting behavior of interacting particle systems and mean-field games. Roughly speaking, DDSDE (or McKean-Vlasov SDE) describe stochastic systems whose evolution is determined not only by the microcosmic location of the particle, but also by the macrocosmic distribution. So it has attracted wide attention (see [7, 8, 9, 23, 34, 44] and references therein).

When b and Σ satisfy some continuity assumptions, there are numerous results devoted to studying the existence and uniqueness of this type of DDSDE (1.1) (see [21, 46] for examples). For the case that b is only measurable, of at most linear growth and Lipschitz continuous with respect to μ, when Σ does not depend on μ, is uniformly non-degenerate and Lipschitz continuous, by using the classical Krylov estimates, Mishura and Veretenikov [35] obtained the strong well-posedness of (1.1). After that, the strong well-posedness was extended to local $L^q_t L^p_x$ drift by Röckner and Zhang in [40]. Moreover, by the relative entropy method and Girsanov's theorem, Lacker [31] also obtained some well-posedness results for linear growth cases (see also [30]). Then, in the special case (1.2), Han obtained the well-posedness for $L^q_t L^p_x$ drift based on the relative entropy method in [22]. In [50], by some heat kernel estimates and Schauder-Tychonoff fixed point theorem, Zhao established the well-posedness results for DDSDE in a more general case (1.1) when b satisfies some $L^q_t L^p_x$ condition and the derivatives of Σ with respect to μ are Hölder continuous. It should be noted that by Zvonkin's transformation of [53] (see also [48] and [49]) and entropy method, the last two authors together with Zhang in [24] obtained the strong well-posedness for (1.1), where Σ does not dependent on μ and b is in mixed $L^q_t L^p_x$ space. For existence and uniqueness results on weak solutions to Nemytskii-type (=density dependent) DDSDE with merely measurable coefficients we refer to [2, 3, 4, 5].

In this work, we are interested in investigating the averaging principle of the following DDSDE with highly oscillating time component

$$\mathrm{d}X^\varepsilon_t = b\left(\frac{t}{\varepsilon}, X^\varepsilon_t, \mu^\varepsilon_t\right)\mathrm{d}t + \sigma(X^\varepsilon_t)\mathrm{d}W_t, \quad X^\varepsilon_0 = \xi \in \mathscr{F}_0, \tag{1.3}$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is a measurable function, $\mu^\varepsilon_t := \mathcal{L}(X^\varepsilon_t)$ is the time marginal law of X^ε_t and the time scale $0 < \varepsilon \ll 1$.

Usually, solving the original system (1.3) is relatively difficult because of the highly oscillating time component. Therefore, it is desirable to find a simplified system which simulates and predicts the evolution of the original system over a long time scale. As is well known, the highly oscillating time component can be "averaged" out to produce such a simplified system under some suitable conditions, which is called averaging principle.

More exactly, let

$$\bar{b}(x, \mu) := \lim_{T\to\infty} \frac{1}{T} \int_0^T b(t, x, \mu)\mathrm{d}t. \tag{1.4}$$

If b is independent of μ, b is called a *KBM-vector feld* (KBM stands for Krylov, Bogolyubov and Mitropolsky) if the convergence (1.4) is uniformly with respect to x on bounded subsets of $\mathbb{R}^d$ (see e.g. [41]). The averaging principle states, as the time scale ε goes to zero, that the solution of the original equation (1.3) converges to that of the following averaged equation on any finite time interval

$$\mathrm{d}X_t = \bar{b}(X_t, \mu_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = \xi, \tag{1.5}$$

where μ_t stands for the distribution of X_t.

The averaging principle was firstly established for deterministic systems by Krylov, Bogolyubov and Mitropolsky [6, 28]. Then it was extended to stochastic differential equations by Khasminskii [26]. After that, extensive work on the averaging principle for finite and infinite dimensional stochastic differential equations was done; see e.g. [1, 10, 11, 12, 15, 18, 19, 20, 27, 33, 36, 45, 47] and the references therein.

Recall that the strong convergence rate of the averaging principle for slow-fast McKean-Vlasov SDE was established by the techniques of time discretization and Poisson equation in [38]. Furthermore, as discussed in [25], the strong convergence rate of the averaging principle for slow-fast McKean-Vlasov SPDE was studied, based on the variational approach and the technique of time discretization. Note that the coefficients of the slow equation with fast variables were globally Lipschitz continuous with respect to the slow variable in the above results. Recently, the strong convergence without a rate for DDSDE with highly oscillating time component driven by fractional Brownian motion and standard Brownian motion was shown in [43], which requires that the drift term is continuous in the slow variable.

Obviously, Lipschitz or continuity assumptions on b are too strong for some applications. There are a lot of interesting models from physics, only having bounded measurable or even singular L^p interaction kernels b and $\hat{b}$. For example, the rank-based interaction studied in [30] has an indicator kernel, which is discontinuous; the Biot-Savart law appearing in the vortex description of 2-dimensional incompressible Navier-Stokes equations has a singular kernel like $x^\perp/|x|^2$ (see e.g. [52]). However, to the best of our knowledge, there is no result concerning the averaging principle both of DDSDE and SDE with L^p drift.

Following the above motivations, we consider the strong and weak convergence of the averaging principle for DDSDE with L^p drift in the present paper. Moreover, we obtain the rate of the strong and weak convergence, which is important for functional limit theorems in probability and homogenization in PDEs. Throughout this paper we need the following conditions.

($\mathbf{H}_b^1$) Let $p_0 \in (d \vee 2, \infty)$ and assume that there is a nonnegative constant κ_0 such that for all $t \geqslant 0$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$

$$|\!|\!| b(t,\cdot,\mu)|\!|\!|_{p_0} + \frac{|\!|\!| b(t,\cdot,\mu) - b(t,\cdot,\nu)|\!|\!|_{p_0}}{\|\mu-\nu\|_{var}} \leqslant \kappa_0,$$

where $\|\mu-\nu\|_{var} := \sup_{A \in \mathscr{B}(\mathbb{R}^d)} |\mu(A) - \nu(A)|$ is total variation.

($\mathbf{H}_b^2$) There are functions $\bar{b} : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^d$, $\omega : \mathbb{R}_+ \to \mathbb{R}_+$ and $H : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}_+$ such that for all $(T, t, x, \mu) \in \mathbb{R}_+^2 \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$

$$\left| \frac{1}{T} \int_t^{T+t} (b(s,x,\mu) - \bar{b}(x,\mu))\mathrm{d}s \right| \leqslant \omega(T) H(x,\mu), \tag{1.6}$$

where $\lim_{t\to\infty} \omega(t) = 0$ and $\sup_\mu |\!|\!| H(\cdot,\mu)|\!|\!|_{p_0} < \kappa_0$. Here p_0 and κ_0 are as in **($\mathbf{H}_b^1$)**.

($\mathbf{H}_\sigma$) There are constants $p > d \vee 2$, $\kappa_1 > 1$ and $\beta \in (0,1)$ such that for all $x, y, \xi \in \mathbb{R}^d$,

$$\kappa_1^{-1}|\xi| \leqslant |\sigma(x)\xi| \leqslant \kappa_1|\xi|, \quad |\!|\!| \nabla\sigma|\!|\!|_p \leqslant \kappa_1,$$

and

$$\|\sigma(x) - \sigma(y)\|_{HS} \leqslant \kappa_1 |x-y|^\beta,$$

where $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm.

Please see the definition of the localized L^p norm $|\!|\!|\cdot|\!|\!|_p$ in Section 2.1.

Remark 1.1. We Note that

$$\begin{aligned}\|\bar{b}(\cdot,\mu)\|_{p_0} &\leqslant \frac{1}{T}\int_0^T \|b(s,\cdot,\mu)\|_{p_0}\mathrm{d}s + \|\frac{1}{T}\int_0^T \big(b(s,\cdot,\mu)-\bar{b}(\cdot,\mu)\mathrm{d}s\big)\|_{p_0}\\ &\leqslant \kappa_0+\omega(T)\|H(\cdot,\mu)\|_{p_0},\end{aligned}$$

provided conditions $(\mathbf{H}_b^1)$ and $(\mathbf{H}_b^2)$ hold. Taking $T\to\infty$, we have

$$\|\bar{b}(\cdot,\mu)\|_{p_0}\leqslant \kappa_0.$$

Similarly, we have

$$\|\bar{b}(\cdot,\mu)-\bar{b}(\cdot,\nu)\|_{p_0}\leqslant \kappa_0\|\mu-\nu\|_{var}.$$

Thus, the coefficient $\bar{b}$ satisfies the condition $(\mathbf{H}_b^1)$ with the same constant κ_0.

Under assumptions $(\mathbf{H}_b^1)$ and $(\mathbf{H}_\sigma)$, for any initial value $\xi\in\mathscr{F}_0$, it is well-known that there is a unique strong solution to DDSDE (1.3) (respectively, (1.5)); see [50] and [24]. The aim of this work is to show the following strong and weak convergence of the averaging principle for DDSDE and SDE with L^p drift.

Theorem 1.2. *Under $(\mathbf{H}_b^1)$, $(\mathbf{H}_b^2)$ and $(\mathbf{H}_\sigma)$, for any $T>0$ and $\ell\in(0,1)$, there is a constant C, depending only on $\kappa_0,\kappa_1,T,d,\beta,p_0,p,\ell$, such that for any $\varepsilon>0$*

$$\sup_{t\in[0,T]}\|\mu_t^\varepsilon-\mu_t\|_{var}\leqslant C\inf_{h>0}\left(h^{\frac{1}{2}-\frac{d}{2p_0}}+\omega\left(\frac{h}{\varepsilon}\right)\right) \tag{1.7}$$

and

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\varepsilon-X_t|^{2\ell}\right)\leqslant C\inf_{h>0}\left((\omega(h/\varepsilon))^2+h^{1-\frac{d}{p_0}}\right)^{\ell}. \tag{1.8}$$

When the drift b is independent of the distribution, we have the following results, where the convergence rate is independent of p_0.

Theorem 1.3. *Assume that*

$$b(t,x,\mu)\equiv b(t,x).$$

Under $(\mathbf{H}_b^1)$, $(\mathbf{H}_b^2)$ and $(\mathbf{H}_\sigma)$, for any $T>0$, $\delta>0$ and $\ell\in(0,1)$, there is a constant C, depending only on $\kappa_0,\kappa_1,T,d,p_0,p,\delta,\ell$, such that for any $\varepsilon>0$

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^\varepsilon-X_t|^{2\ell}\right)\leqslant C\inf_{h>0}\left(\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2+h^{1-\delta}\right)^{\ell}.$$

Remark 1.4.
(i) Since we use the Zvonkin transformation using the parabolic equation, when $\bar{b}=\bar{b}(t)=\bar{b}(\cdot,\mu_t)$ depends on the time variable t, the time regularity for solutions to this parabolic equation affects the convergence rate (see (5.2) and Lemma 3.15 for more details). When $\bar{b}$ is independent of time, we can construct the Zvonkin transformation using the elliptic equation. Hence, the convergence rates in Theorems 1.2 and 1.3 are different.
(ii) Noting that $\|f\|_{p_0}\lesssim\|f\|_\infty$ for all $p_0\in(1,\infty)$, all results in our paper are valid for $p_0=\infty$, in which case the rate of convergence in (1.8) is

$$\inf_{h>0}\left(\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2+h^{1-\delta}\right)^{\ell}$$

for any $\delta>0$. In particular, we obtain the convergence rate $\varepsilon^{\frac{1}{3}-\delta}$ for a large number of examples (see e.g. Example 1.7 below), which is faster than $\varepsilon^{\frac{1}{6}}$ in [25].

Remark 1.5. These averaging principle results are also applicable to the following system

$$\mathrm{d}X_t = \varepsilon b\,(t, X_t, \mathcal{L}(X_t))\,\mathrm{d}t + \sqrt{\varepsilon}\sigma(X_t)\mathrm{d}W_t. \tag{1.9}$$

Define $Z_t^\varepsilon := X_{t/\varepsilon}$ and $W_t^\varepsilon := \sqrt{\varepsilon}W_{t/\varepsilon}$ for all $t \in \mathbb{R}_+$. We rewrite (1.9) as

$$\mathrm{d}Z_t^\varepsilon = b\,(t/\varepsilon, Z_t^\varepsilon, \mathcal{L}(Z_t^\varepsilon))\,\mathrm{d}t + \sigma(Z_t^\varepsilon)\mathrm{d}W_t^\varepsilon.$$

Then we can consider the following system

$$\mathrm{d}\tilde{X}_t^\varepsilon = b\left(t/\varepsilon, \tilde{X}_t^\varepsilon, \mathcal{L}(\tilde{X}_t^\varepsilon)\right)\mathrm{d}t + \sigma(\tilde{X}_t^\varepsilon)\mathrm{d}W_t.$$

Note that since the drift of both the DDSDE and SDE in this paper is locally $L_x^{p_0}$ integrable, we cannot use Gronwall's lemma or the generalized Gronwall lemma directly to prove the convergence of X^ε to X as in [25, 43]. On the other hand, our system (1.3) can be rewritten in the following slow-fast system:

$$\begin{cases} \mathrm{d}X_t^\varepsilon = b\,(Y_t^\varepsilon, X_t^\varepsilon, \mu_t^\varepsilon)\,\mathrm{d}t + \sigma(X_t^\varepsilon)\mathrm{d}W_t, \\ \mathrm{d}Y_t^\varepsilon = \dfrac{1}{\varepsilon}\mathrm{d}t. \end{cases}$$

Since the Kolmogorov operator of the fast process $Y_t^\varepsilon = \frac{t}{\varepsilon}$, $t \geqslant 0$, does not have a second order elliptic part, we cannot use the technique based on the Poisson equation as in [38]. To overcome these difficulties, we use Zvonkin's transformation to remove the drift b and employ the classical technique of time discretization.

More precisely, consider the following backward PDE for $t \in [0, T]$ related to (1.5)

$$\partial_t u + a_{ij}\partial_i\partial_j u - \lambda u + B\cdot\nabla u + B = 0, \quad u(T) = 0,$$

where $B(t, x) := \bar{b}(x, \mu_t)$, $\lambda \geqslant 0$, is the dissipative term. Under $(\mathbf{H}_\sigma)$ and $(\mathbf{H}_b^1)$, Xia et al [48] proved that for a sufficiently large number λ, there is a solution u such that

$$|\nabla u(t, x)| \leqslant \frac{1}{2}, \quad t \in [0, T], x \in \mathbb{R}^d.$$

Hence, if we define $\Phi_t(x) := x + u(t, x)$, then $x \to \Phi_t(x)$ is a C^1 diffeomorphism of $\mathbb{R}^d$. By Itô's formula, $Y_t^\varepsilon := \Phi_t(X_t^\varepsilon)$ and $Y_t := \Phi_t(X_t)$ solve the following new SDEs:

$$\begin{aligned} \mathrm{d}Y_t^\varepsilon =& \lambda u(t, \Phi_t^{-1}(Y_t^\varepsilon))\mathrm{d}t + (\sigma^*\nabla\Phi_t(\Phi_t^{-1}(Y_t^\varepsilon)))\mathrm{d}W_t \\ &+ \big(b(t/\varepsilon, X_t^\varepsilon, \mu_t^\varepsilon) - \bar{b}(X_t^\varepsilon, \mu_t)\big)\cdot\nabla\Phi_t\big(X_t^\varepsilon\big)\mathrm{d}t \end{aligned}$$

and

$$\mathrm{d}Y_t = \lambda u(t, \Phi_t^{-1}(Y_t))\mathrm{d}t + (\sigma^*\nabla\Phi_t)(\Phi_t^{-1}(Y_t))\mathrm{d}W_t,$$

where σ^* is the transpose of σ and Φ_t^{-1} is the inverse of $x \to \Phi_t(x)$. Since these new systems have differentiable diffusion coefficients and the drifts are Lipshitz continuous, we can use the stochastic Gronwall inequality. Please see the complete formulation in Section 5.

The remaining part of the proof is about how to use the technique of time discretization to estimate the following crucial term

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\Big|\int_0^t \big(b(\frac{s}{\varepsilon}, X_s^\varepsilon, \mu_s^\varepsilon) - \bar{b}(X_s^\varepsilon, \mu_s)\big)\cdot\nabla\Phi_s\big(X_s^\varepsilon\big)\mathrm{d}s\Big|^2\Big]. \tag{1.10}$$

In particular, we need to estimate

$$\|\mu_t - \mu_t^\varepsilon\|_{var} \tag{1.11}$$

and the following difference for $B(t) = b(s/\varepsilon)\cdot\nabla\Phi_t$ and $B(t) = \bar{b}\cdot\nabla\Phi_t$:

$$\mathbb{E}\Big|\int_0^t \big(B(s, X_s^\varepsilon, \mu_s^\varepsilon) - B(s, X_{\pi_h(s)}^\varepsilon, \mu_{\pi_h(s)}^\varepsilon)\big)\mathrm{d}s\Big|^2, \tag{1.12}$$

where $\pi_h(s)$ is defined by $\pi_h(s)=s$ for $s\in[0,h)$ and

$$\pi_h(s):=kh,\quad s\in[kh,(k+1)h),\ \forall k\in\mathbb{N}.$$

Since b is not continuous, it is not an easy problem to give an estimate for (1.12). Thanks to the important observation in [17, Lemma 2.1], we deal with this problem by considering the time regularity for the semi-group generated by the solutions to (1.3). Basically, if $X_t^\varepsilon=W_t$ is a standard Brownian motion and f is a discontinuous function, we have

$$\begin{aligned}&\mathbb{E}\Big|\int_0^t\big(f(W_s)-f(W_{\pi_h(s)})\big)\mathrm{d}s\Big|^2\\&=2\mathbb{E}\int_0^t\int_s^t\big(f(W_s)-f(W_{\pi_h(s)})\big)\big(f(W_r)-f(W_{\pi_h(r)})\big)\mathrm{d}r\mathrm{d}s\\&=2\mathbb{E}\int_0^t\big(f(W_s)-f(W_{\pi_h(s)})\big)\Big(\mathbb{E}^{\mathscr{F}_s}\int_s^t\big(f(W_r)-f(W_{\pi_h(r)})\big)\mathrm{d}r\Big)\mathrm{d}s,\end{aligned}$$

where $\mathscr{F}_s:=\sigma(W_r,r\leqslant s)$ and $\mathbb{E}^{\mathscr{F}_s}$ is the conditional expectation with respect to $\mathscr{F}_s$. We note that by the Markov property,

$$\begin{aligned}&\mathbb{E}^{\mathscr{F}_s}\int_s^t\big(f(W_r)-f(W_{\pi_h(r)})\big)\mathrm{d}r\\&=\mathbb{E}^{\mathscr{F}_s}\int_s^{s+h}\big(f(W_r)-f(W_{\pi_h(r)})\big)\mathrm{d}r+\int_{s+h}^t\big(P_{r-s}f(W_s)-P_{\pi_h(r)-s}f(W_s)\big)\mathrm{d}r,\end{aligned}$$

where $P_tf(x):=\mathbb{E}f(x+W_t)$ is the semi-group of Brownian motion. Note that as $h\to0$ the first term converges to zero. And the second term also converges to zero as $h\to0$ since $s\to P_sf$ is continuous. Please see Section 3 below for more details. We mention that similar estimates are obtained in [32] by the stochastic sewing techniques.

To estimate (1.11), we employ a method based on the Kolmogorov equation which is also used in [37]. Then, again by time discretization, we estimate the difference (1.11) and obtain (1.7) (see Section 4).

Now we illustrate our results by some examples. Firstly, the following example shows that the function $\omega(t)$ can be of the form $t^{-\alpha}$ with any power $\alpha\in(0,1]$, which can also be used in some systems with singular interactions.

Example 1.6. Consider the following DDSDE in $\mathbb{R}^d$

$$\begin{aligned}\mathrm{d}X_t^\varepsilon&=\left(\Big[(1+t/\varepsilon)^{-\alpha_1}+1\Big]\int_{\mathbb{R}^d}\frac{X_t^\varepsilon-y}{|X_t^\varepsilon-y|^{\alpha_2}}\mu_t^\varepsilon(\mathrm{d}y)\right)\mathrm{d}t+\mathrm{d}W_t\\&=:b(t/\varepsilon,X_t^\varepsilon,\mu_t^\varepsilon)\mathrm{d}t+\mathrm{d}W_t,\end{aligned}$$

where $\alpha_1>0$, $1<\alpha_2<2\wedge(1+\frac{d}{2})$ and μ_t^ε is the distribution of X_t^ε. It is clear that the averaged equation is

$$\begin{aligned}\mathrm{d}X_t&=\left(\int_{\mathbb{R}^d}\frac{X_t-y}{|X_t-y|^{\alpha_2}}\mu_t(\mathrm{d}y)\right)\mathrm{d}t+\mathrm{d}W_t\\&=:\bar{b}(X_t,\mu_t)\mathrm{d}t+\mathrm{d}W_t,\end{aligned}$$

where μ_t is the distribution of X_t, and

$$\left|\frac{1}{T}\int_t^{t+T}\big(b(s,x,\mu)-\bar{b}(x,\mu)\big)\,\mathrm{d}s\right|\leqslant\omega\,(T)\,(1-\alpha_1)^{-1}\int_{\mathbb{R}^d}\frac{|x-y|}{|x-y|^{\alpha_2}}\mu(\mathrm{d}y)$$

for all $(T,t,x,\mu)\in\mathbb{R}^2\times\mathbb{R}\times\mathcal{P}(\mathbb{R}^d)$, where

$$\omega(t)=\begin{cases}t^{-(\alpha_1\wedge1)}&\text{for }\alpha_1\in(0,1)\cup(1,\infty)\\t^{-1}\ln t&\text{for }\alpha_1=1.\end{cases}$$

Then we have for any $\delta > 0$

$$\sup_{t\in[0,T]} \|\mu_t^\varepsilon - \mu_t\|_{var} \leqslant C\varepsilon^{\frac{\alpha(2-\alpha_2)}{2+2\alpha-\alpha_2}-\delta}$$

and

$$\left[\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\varepsilon - X_t|^{2\ell}\right)\right]^{\frac{1}{\ell}} \leqslant C\varepsilon^{\frac{4\alpha-2\alpha\alpha_2}{2+2\alpha-\alpha_2}-\delta}$$

for any $0<\ell<1$, where $\alpha = \alpha_1 \wedge 1$ for $\alpha_1 \in (0,\infty)$.

Next we give a more general example, where the function $\omega(t) \asymp t^{-1}$, i.e. there exists a constant C such that $C^{-1}t^{-1} \leqslant \omega(t) \leqslant Ct^{-1}$.

Example 1.7. Let $p_0 \in (d\vee 2, \infty)$. Consider the following DDSDE

$$\mathrm{d}X_t^\varepsilon = \left[\int_{\mathbb{R}} F\left(\sin(\xi t/\varepsilon), \int_{\mathbb{R}^d} \phi(X_t^\varepsilon, y)\mu_t^\varepsilon(\mathrm{d}y)\right)\nu(\mathrm{d}\xi)\right]\mathrm{d}t + \mathrm{d}W_t, \tag{1.13}$$

where μ_t^ε is the time marginal law of X_t, $F : [-1,1]\times\mathbb{R}^m \to \mathbb{R}^d$ is measurable and satisfies for some constant $L_F > 0$

$$|F(t,0)| \leqslant L_F,\ |F(t,x) - F(t,y)| \leqslant L_F|x-y| \quad \text{for all } (t,x,y) \in [-1,1]\times\mathbb{R}^{2m}, \tag{1.14}$$

ν is some finite measure on $\mathbb{R}$ satisfying

$$\int_{\mathbb{R}\setminus\{0\}} \frac{1}{\xi}\nu(\mathrm{d}\xi) < \infty$$

and $\phi : \mathbb{R}^d \to \mathbb{R}^m$ is measurable and satisfies

$$\sup_y \|\phi(\cdot, y)\|_{p_0} < \infty.$$

Set

$$b(t,x,\mu) := \int_{\mathbb{R}} F\left(\sin(\xi t), \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\right)\nu(\mathrm{d}\xi)$$

and

$$\begin{aligned}\bar{b}(x,\mu) :=& \frac{1}{2\pi}\int_0^{2\pi} F\left(\sin\tau, \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\right)\nu(\mathbb{R}\setminus\{0\})\mathrm{d}\tau \\ &+ F\left(0, \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\right)\nu(\{0\}).\end{aligned}$$

We claim that

$$(b,\bar{b}) \text{ satisfies conditions } (\mathbf{H}_b^1) \text{ and } (\mathbf{H}_b^2) \text{ with } \omega(T) := \frac{4\pi L_F}{T}\int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|}, \tag{1.15}$$

which is proved in the Appendix.

Hence, based on Theorem 1.2 and Lemma A.1, we have

$$\sup_{t\in[0,T]} \|\mu_t^\varepsilon - \mu_t\|_{var} \leqslant C\varepsilon^{\frac{1}{3}-\frac{2d}{9p_0-3d}}$$

and

$$\left[\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\varepsilon - X_t|^{2\ell}\right)\right]^{\frac{1}{\ell}} \leqslant C\varepsilon^{\frac{2}{3}\left(1-\frac{2d}{3p_0-d}\right)} \tag{1.16}$$

for any $0<\ell<1$ and $T>0$, where X_t is the solution of

$$\mathrm{d}X_t = \bar{b}(X_t,\mu_t)\mathrm{d}t + \mathrm{d}W_t.$$

Here μ_t is the distribution of X_t for all $t \geqslant 0$.

Remark 1.8. In this paper, we only consider the case where σ does not depend on time t and the measure μ. We hope that in future work we can use our framework here to study the time-inhomogeneous case $\sigma = \sigma(t,x)$ or the even more general case $\sigma = \sigma(t,x,\mu)$ under the following standard condition on σ

$(\mathbf{H}_\sigma^{'})$ There are functions ω_σ and H_σ such that

$$\frac{1}{T}\int_t^{T+t} \|\sigma(t,x,\mu) - \bar{\sigma}(x,\mu)\|_{HS}^2 \mathrm{d}t \leqslant \omega_\sigma(T) H_\sigma(x,\mu)$$

for all $(T,t,x) \in (\mathbb{R}_+^2 \times \mathbb{R}^d)$, where $\lim_{t\to\infty} \omega_\sigma(t) = 0$ and $\sup_\mu \|H_\sigma(\cdot,\mu)\|_{\frac{p_0}{2}} < \infty$.

Structure of the paper.

In Section 2, we first introduce the localized Bessel potential space, the embedding lemma and the local Hardy-Littlewood maximal function. Then we show some well-posedness results and priori estimates about parabolic and elliptic PDEs for later establishing the time regularity of solutions in Section 3 and performing Zvonkin's transformation in Section 5.

In Section 3, we study the time regularity of solutions to both SDE and PDE in the L^p framework. As usual, we use the technique of time discretization to obtain the convergence for the averaging principle. That is we need to show a time discretization's type estimate of (1.12), and then estimate (1.10) by condition $(\mathbf{H}_b^2)$ (see Section 3.2). For (1.12), we first asume that Z_t^ε is a solution to the SDE (1.3) without drift, that is to say $b \equiv 0$. By the semi-group property from condition $(\mathbf{H}_\sigma)$ and heat kernel estimates we estimate (1.12) for any localized L^p integrable B with $p > (d \vee 2)$. Then it follows from the Girsanov transform that (1.12) holds for X_t^ε, where X_t^ε is the solution to (1.3) (see Section 3.1). We also prove the time regularity of the gradient of solutions for parabolic PDEs in Section 3.3, which is used in Section 4 and Section 5 respectively.

In Section 4, after realizing that (4.8) holds, we prove the weak convergence rate of (1.11), where the key point is the estimates for time regularity obtained in Section 3.3.

In Section 5, we give the proofs to our main results Theorems 1.2 and 1.3 by Zvonkin's transformation.

Notations.

Throughout this paper, $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d, d \in \mathbb{N}$. For $j \in \mathbb{N} \cup \{0\}$, we use the notation ∇^j to denote the jth order derivative. Moreover, let C_0^∞ denote the function space of all smooth functions with compact support. We write C_b^k (respectively, C_b^∞) to mean the space of all smooth functions with bounded jth derivatives for all integer $j \in [0,k]$ (respectively, $j \in \mathbb{N} \cup \{0\}$). Let $[\alpha]$ be the largest integer which is smaller than α for any constant $\alpha > 0$. We use C with or without subscripts to denote an unimportant constant, whose value may change from line to line. Writing " $:=$ " we mean equal by definition. By $A \lesssim B$ we mean that for some unimportant constant $C \geqslant 1$,

$$A \leqslant CB.$$

2. Preliminaries

In this section we show some auxiliary results for later use.

2.1. Localized Bessel potential space.

For any $(\alpha,p) \in \mathbb{R} \times (1,\infty)$, we write

$$H^{\alpha,p} := (\mathbb{I} - \Delta)^{-\alpha/2} \left(L^p(\mathbb{R}^d)\right)$$

for the usual Bessel potential space with the norm given by $\|f\|_{\alpha,p} := \|(\mathbb{I}-\Delta)^{\alpha/2} f\|_p$, where $\|\cdot\|_p$ is the usual L^p-norm. Here $(\mathbb{I}-\Delta)^{\alpha/2} f$ is defined through Fourier's transform

$$(\mathbb{I}-\Delta)^{\alpha/2} f := \mathcal{F}^{-1}\left((1+|\cdot|^2)^{\alpha/2}\mathcal{F}f\right).$$

We note that if $\alpha = n \in \mathbb{N}$ and $p \in (1, \infty)$, an equivalent norm in $H^{n,p}$ is given by

$$\|f\|_{n,p} = \|f\|_p + \|\nabla^n f\|_p.$$

Let $\chi \in C_0^\infty(\mathbb{R}^d)$ such that $\chi(x) = 1$ for $|x| \leqslant 1$, $\chi(x) = 0$ for $|x| > 2$ and $|\chi(x)| \leqslant 1, \forall x \in \mathbb{R}^d$. Define

$$\chi_r(x) := \chi(x/r), \quad \chi_r^z(x) := \chi_r(x - z)$$

for all $r > 0$ and $z \in \mathbb{R}^d$. Given $r > 0$, we introduce the following localized $H^{\alpha,p}$-space:

$$\widetilde{H}^{\alpha,p} := \left\{ f \in H^{\alpha,p}_{\text{loc}}(\mathbb{R}^d) : \|\!|f|\!\|_{\alpha,p} := \sup_z \|\chi_r^z f\|_{\alpha,p} < \infty \right\}.$$

Clearly, this space does not depend on r and the corresponding norms are equivalent. When $\alpha = 0$, we simply write

$$\widetilde{L}^p := \widetilde{H}^{0,p} \quad \text{and} \quad \|\!|f|\!\|_p := \|\!|f|\!\|_{0,p}.$$

It follows from Hölder's inequality that for any $1 \leqslant p_2 \leqslant p_1 \leqslant \infty$

$$L^{p_1} \subset \widetilde{L}^{p_1} \subset \widetilde{L}^{p_2}.$$

For $T > 0$, $p, q \in (1, \infty)$ and $\alpha \in \mathbb{R}$, we set

$$\mathbb{L}^p_q(T) := L^q\big([0,T]; L^p\big), \quad \mathbb{H}^{\alpha,p}_q(T) := L^q\big([0,T]; H^{\alpha,p}\big).$$

Now we introduce the localized space

$$\widetilde{\mathbb{H}}^{\alpha,p}_q(T) := \left\{ f \in \mathbb{H}^{\alpha,p}_q(T) : \|\!|f|\!\|_{\widetilde{\mathbb{H}}^{\alpha,p}_q(T)} := \sup_{z \in \mathbb{R}^d} \|\chi_r^z f\|_{\mathbb{H}^{\alpha,p}_q(T)} < \infty \right\}.$$

By a finite covering technique, it can be verified that also the definition of $\widetilde{\mathbb{H}}^{\alpha,p}_q$ does not depend on the choice of r (see [48, Section 2]). We note that all these spaces are Banach spaces and that

$$L^q\big([0,T]; \widetilde{H}^{\alpha,p}\big) \subset \widetilde{\mathbb{H}}^{\alpha,p}_q(T).$$

For $\alpha = 0$, set

$$\widetilde{\mathbb{L}}^p_q(T) := \widetilde{\mathbb{H}}^{0,p}_q(T).$$

If $q = \infty$, for simplicity, we define

$$\widetilde{\mathbb{H}}^{\alpha,p}(T) := L^\infty([0,T]; \widetilde{H}^{\alpha,p}), \quad \widetilde{\mathbb{L}}^p(T) := \widetilde{\mathbb{L}}^p_\infty(T), \quad \text{and} \quad \mathbb{L}^\infty_T := L^\infty([0,T] \times \mathbb{R}^d).$$

Let C^α denote the Hölder space of order α, which consists of all functions g for which the norm

$$\|g\|_{C^\alpha} := \sum_{|\beta| \leqslant [\alpha]} \|\nabla^\beta g\|_{L^\infty} + \sum_{|\beta| = [\alpha]} \big[\nabla^\beta g\big]_{C^{\alpha - [\alpha]}}$$

is finite, where

$$[g]_{C^{\alpha-[\alpha]}} := \sup_{\substack{x,y \in \mathbb{R}^d \\ x \neq y}} \frac{|g(x) - g(y)|}{|x - y|^{\alpha - [\alpha]}}.$$

Lemma 2.1 (Embedding lemma)**.** *Let $1 < p < \infty$. Then we have*

$$\widetilde{H}^{\alpha,p} \subset C^{\alpha - d/p}$$

and

$$\widetilde{\mathbb{H}}^{\alpha,p}(T) \subset L^\infty\left([0,T]; C^{\alpha - d/p}\right)$$

provided $\alpha > d/p$.

Proof. It follows from Sobolev's embedding theorem that $H^{\alpha,p}\subset C^{\alpha-d/p}$ if $\alpha>d/p$. Note that

$$\|g\|_{C^{\alpha-d/p}}\leqslant\sup_z\|\chi_r^z g\|_{C^{\alpha-d/p}}$$

for all $g\in C^{\alpha-d/p}$ and $r>0$. Therefore, we have

$$\|g\|_{C^{\alpha-d/p}}\leqslant\sup_z\|\chi_r^z g\|_{C^{\alpha-d/p}}\lesssim\sup_z\|\chi_r^z g\|_{H^{\alpha,p}}=\|g\|_{\widetilde{H}^{\alpha,p}}.$$

Moreover, for all $f\in\widetilde{\mathbb{H}}^{\alpha,p}$ one sees that

$$\sup_{t\in[0,T]}\|f(t)\|_{C^{\alpha-d/p}}\leqslant\sup_{t\in[0,T]}\sup_z\|\chi_r^z f(t)\|_{C^{\alpha-d/p}}\lesssim\sup_{t\in[0,T]}\sup_z\|\chi_r^z f(t)\|_{H^{\alpha,p}}=\|f\|_{\widetilde{\mathbb{H}}^{\alpha,p}}.$$

□

Now we introduce the local Hardy-Littlewood maximal function, which is defined by

$$\mathcal{M}f(x):=\sup_{r\in(0,1)}\frac{1}{|B_r|}\int_{B_r}f(x+y)\mathrm{d}y.$$

The following result is taken from Lemma 2.1 in [48] (see also [16, Appendix A]).

Lemma 2.2. *(i) There is a constant $C=C(d)>0$, such that for any $f\in L^\infty(\mathbb{R}^d)$ with $\nabla f\in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ and for Lebesgue-almost all $x,y\in\mathbb{R}^d$,*

$$|f(x)-f(y)|\leqslant C|x-y|\big(\mathcal{M}|\nabla f|(x)+\mathcal{M}|\nabla f|(y)+\|f\|_\infty\big).\tag{2.1}$$

(ii) For any $p\in(1,\infty)$, there is a constant $C=C(d,p)>0$ such that for all $f\in\widetilde{L}^p$,

$$\|\mathcal{M}f\|_p\leqslant C\|f\|_p.\tag{2.2}$$

2.2. **Parabolic equation.** In order to study DDSDE, we consider the following second order parabolic PDE in $\mathbb{R}_+\times\mathbb{R}^d$:

$$\partial_t u=a_{ij}\partial_i\partial_j u-\lambda u+b\cdot\nabla u+f,\quad u(0)=\varphi,\tag{2.3}$$

where $\lambda\geqslant 0$, $a=(a_{ij}):\mathbb{R}^d\to\mathbb{R}^d\otimes\mathbb{R}^d$ is a symmetric matrix-valued Borel measurable function satisfying $(\mathbf{H}_a)$, i.e.,

$(\mathbf{H}_a)$ there exist constants $c_0>0$ and $\theta\in(0,1)$ such that

$$c_0^{-1}|\xi|\leqslant|a(x)\xi|\leqslant c_0|\xi|,\quad\|a(x)-a(y)\|_{HS}\leqslant c_0|x-y|^\theta$$

for all $\xi\in\mathbb{R}^d$ and $x,y\in\mathbb{R}^d$,

and $b:\mathbb{R}\times\mathbb{R}^d\to\mathbb{R}^d$ is a vector-valued Borel measurable function. Firstly, we introduce the definition of a solution to PDE (2.3).

Definition 2.3. Let $T>0$, $p,q\in(1,\infty)$, $\lambda\geqslant 0$, $b,f\in\widetilde{\mathbb{L}}^p_q(T)$ and $\varphi\in C_b^\infty$. We call a function u with $\partial_t u\in\widetilde{\mathbb{L}}^p_q(T)$ and $u\in\widetilde{\mathbb{H}}^{2,p}_q(T)$ a *solution* of PDE (2.3) if for Lebesgue almost all $(t,x)\in\mathbb{R}_+\times\mathbb{R}^d$,

$$u(t,x)=\int_0^t\big(a_{ij}\partial_i\partial_j u(s,x)-\lambda u(s,x)+b\cdot\nabla u(s,x)+f(s,x)\big)\mathrm{d}s+\varphi(x).$$

Remark 2.4. For any $\chi\in C_0^\infty(\mathbb{R}^d)$ and $f\in\widetilde{\mathbb{H}}^{\alpha,p}_q(T)$, by the definition of the localized spaces $\widetilde{\mathbb{H}}^{\alpha,p}_q(T)$, we have $\chi f\in\mathbb{H}^{\alpha,p}_q(T)$. Hence for any solution u of PDE (2.3) in the sense of Definition 2.3, χu is Hölder continuous on $[0,T]\times\mathbb{R}^d$ if $d/p+2/q<2$ according to [29, Lemma 10.2]. Moreover, $\nabla(\chi u)$ is Hölder continuous on $[0,T]\times\mathbb{R}^d$ if $d/p+2/q<1$. In view of the arbitrariness of the cut-off function χ, u (respectively, ∇u) are locally Hölder continuous on $[0,T]\times\mathbb{R}^d$ if $d/p+2/q<2$ (respectively, $d/p+2/q<1$).

The following property of the solution u comes from [48, Theorem 3.2].

Lemma 2.5. *Let $T>0$, $p,q\in(1,\infty)$ with $2/q+d/p<1$, $\lambda\geqslant 0$, $b\in\widetilde{\mathbb{L}}^p_q(T)$. Set*

$$\Theta:=(d,T,p,q,\|b\|_{\widetilde{\mathbb{L}}^p_q(T)},c_0,\theta).$$

Then there is a constant $\lambda_0=\lambda_0(\Theta)$ such that for all $\lambda\geqslant\lambda_0$ and $f\in\widetilde{\mathbb{L}}^p_q(T)$ and $\varphi\in\widetilde{H}^{2,p}$, there is a unique solution u to PDE (2.3) on $[0,T]$ in the sense of Definition 2.3 such that for any $\alpha\in[0,2)$, $p'\in[p,\infty]$, $q'\in[q,\infty]$ with

$$\beta:=2-\alpha+\frac{2}{q'}+\frac{d}{p'}-\big(\frac{2}{q}+\frac{d}{p}\big)>0,$$

there is a constant $C=C(\Theta,\alpha,p',q')>0$ such that for all $\lambda\geqslant\lambda_0$

$$\lambda^{\frac{\beta}{2}}\|u\|_{\widetilde{\mathbb{H}}^{\alpha,p'}_{q'}(T)}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\nabla^2 u\|_{\widetilde{\mathbb{L}}^p_q(T)}\leqslant C\big(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\big). \tag{2.4}$$

Remark 2.6. By Lemma 2.1, we have $u\in L^\infty([0,T];\mathbf{C}^\gamma)$ for any $\gamma\in(1,2-2/q-d/p)$.

Proof. We note that u is a solution to PDE (2.3) with $u(0)=\varphi$ in the sense of Definition 2.3 if and only if $\bar u:=u-\varphi$ is a solution to PDE (2.3) with $\bar u(0)=0$ and $f=f+a_{ij}\partial_i\partial_j\varphi-\lambda\varphi-b\cdot\nabla\varphi$. Based on Lemma 2.1, we have

$$\begin{aligned}\|f+a_{ij}\partial_i\partial_j\varphi-\lambda\varphi-b\cdot\nabla\varphi\|_{\widetilde{\mathbb{L}}^p_q(T)}&\lesssim\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}+\|b\|_{\widetilde{\mathbb{L}}^p_q(T)}\|\nabla\varphi\|_\infty\\&\lesssim\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+(1+\|b\|_{\widetilde{\mathbb{L}}^p_q(T)})\|\varphi\|_{2,p},\end{aligned}$$

and complete the proof by [48, Theorem 3.2]. □

With the help of a priori estimate (2.4), we obtain the well-posedness of PDE (2.3) for any $\lambda\geqslant 0$.

Proposition 2.7. *Let $T>0$, $p,q\in(1,\infty)$ with $2/q+d/p<1$, $\lambda\geqslant0$, $b\in\widetilde{\mathbb{L}}^p_q(T)$. Then for all $f\in\widetilde{\mathbb{L}}^p_q(T)$ and $\varphi\in\widetilde{H}^{2,p}$ there is a unique solution u to PDE (2.3) on $[0,T]$ in the sense of Definition 2.3 such that*

$$\|\nabla u\|_{\widetilde{\mathbb{L}}^\infty_T}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|u\|_{\widetilde{\mathbb{H}}^{2,p}_q(T)}\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{\widetilde{H}^{2,p}}\right), \tag{2.5}$$

where $C=C(\Theta,\lambda)$.

Proof. By the standard continuity method, it suffices to show a priori estimate (2.5) for (2.3). To this end, we rewrite (2.3) as

$$\partial_t u=a_{ij}\partial_i\partial_j u-(\lambda+\lambda_0)u+b\cdot\nabla u+f+\lambda_0 u,$$

where λ_0 is as in Lemma 2.5. In view of Lemma 2.5, we have

$$(\lambda+\lambda_0)^{\frac{\beta}{2}}\|u\|_{\widetilde{\mathbb{H}}^{\alpha,p}_{q'}}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\nabla^2u\|_{\widetilde{\mathbb{L}}^p_q(T)}\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\lambda_0\|u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\right), \tag{2.6}$$

where $C=C(\Theta,\alpha,q')>0$, $\beta=2-\alpha+\frac{2}{q'}-\frac{2}{q}>0$. Taking $q'=\infty$ in (2.6), then we have

$$(\lambda+\lambda_0)^{\frac{\beta}{2}}\sup_{t\in[0,T]}\|u(t)\|_p\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\lambda_0\left(\int_0^T\|u(t)\|_p^q\mathrm{d}t\right)^{\frac1q}+\|\varphi\|_{2,p}\right).$$

Now it follows from Gronwall's Lemma that

$$\sup_{t\in[0,T]}\|u(t)\|_p\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\right), \tag{2.7}$$

where C depends on Θ,α,λ. Combining (2.6) and (2.7), we obtain for $1<\alpha<2-2/q$

$$\|u\|_{\widetilde{\mathbb{H}}^{\alpha,p}}+\|\partial_t u\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|u\|_{\widetilde{\mathbb{H}}^{2,p}_q(T)}\leqslant C\left(\|f\|_{\widetilde{\mathbb{L}}^p_q(T)}+\|\varphi\|_{2,p}\right).$$

and complete the proof by Lemma 2.1. □

2.3. **Elliptic equation.** Now we consider the following second order elliptic PDE in $\mathbb{R}^d$:

$$a_{ij}\partial_i\partial_j u - \lambda u + b\cdot\nabla u = f, \tag{2.8}$$

where $\lambda \geqslant 0$, $a = (a_{ij}) : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ is a symmetric matrix-valued Borel measurable function satisfying $(\mathbf{H}_a)$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ is a vector-valued Borel measurable function. Firstly, we introduce the definition of a solution to PDE (2.8).

Definition 2.8. Let $p \in (1,\infty)$, $\lambda, T \geqslant 0$ and $b, f \in \widetilde{L}^p$. We call $u \in \widetilde{H}^{2,p}$ a *solution* of PDE (2.8) if for Lebesgue almost all $x \in \mathbb{R}^d$,

$$a_{ij}\partial_i\partial_j u(x) - \lambda u(x) + b(x)\cdot\nabla u(x) = f(x).$$

As a corollary of Lemma 2.5, we have the following results.

Lemma 2.9. *Assume $b \in \widetilde{L}^p$ for some $p > d$. Then there are constants $\lambda_0 = \lambda_0(d, \|b\|_p, p, c_0, \theta)$ and $C = C(d, \|b\|_p, p, p', c_0, \theta)$ such that for any $\lambda \geqslant \lambda_0$ and $f \in \widetilde{L}^p$, there exists a unique solution u to PDE (2.8) in the sense of Definition 2.8 such that*

$$\lambda^{\frac{\beta}{2}}\|u\|_{\widetilde{H}^{\alpha,p'}} + \|\nabla^2 u\|_p \leqslant C\|f\|_p, \tag{2.9}$$

where $\alpha \in [0,2), p \in [p,\infty], p' \in [p,\infty]$ with $\beta := 2 - \alpha + \frac{d}{p'} - \frac{d}{p} > 0$.

Proof. As usual, it suffices to show the a priori estimate (2.9). Let $T > 0$, u be a solution to (2.8) and ϕ be a nonnegative and nonzero smooth function on $[0,\infty)$ with $\phi(0) = 0$. Define $\tilde{u}(t,x) := \phi(t)u(x)$. Then, one sees that $\tilde{u}$ is a solution to the following parabolic equation in the sense of Definition 2.3:

$$\partial_t\tilde{u} = a_{ij}\partial_i\partial_j\tilde{u} - \lambda\tilde{u} + b\cdot\nabla\tilde{u} - \phi f + \phi' u, \quad \tilde{u}(0) = 0.$$

By (2.4), we have for any $\alpha \in [0,2), p' \in [p,\infty]$ with $\beta := 2 - \alpha + \frac{d}{p'} - \frac{d}{p} > 0$,

$$\lambda^{\frac{\beta}{2}}\|\tilde{u}\|_{\widetilde{\mathbb{H}}^{\alpha,p'}(T)} + \|\nabla^2\tilde{u}\|_{\widetilde{\mathbb{L}}^p(T)} \leqslant C\|\phi' u - \phi f\|_{\widetilde{\mathbb{L}}^p(T)},$$

which implies that

$$\lambda^{\frac{\beta}{2}}\|u\|_{\widetilde{H}^{\alpha,p'}} + \|\nabla^2 u\|_p \leqslant C\|\phi\|_\infty^{-1}\left(\|\phi'\|_\infty\|u\|_p + \|\phi\|_\infty\|f\|_p\right),$$

where $\|\phi\|_\infty := \sup_{t\in[0,T]}|\phi(t)|$. Noting that $\|u\|_p \leqslant \|u\|_{p'}$, we obtain (2.9) and complete the proof. □

3. Analysis of time regularity

In this section, letting $T > 0$, we assume that

$$B \in \widetilde{\mathbb{L}}^{p_0}(T) \text{ and } \sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d \text{ satisfy } (\mathbf{H}_\sigma) \text{ for some } p_0 > d. \tag{3.1}$$

and consider the following SDE on a probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$:

$$X_{s,t}^x = x + \int_s^t B(r, X_{s,r}^x)\mathrm{d}r + \int_s^t \sigma(X_{s,r}^x)\mathrm{d}W_r, \tag{3.2}$$

where W_t is a standard d-dimensional Brownian motion. Furthermore, consider the PDE on $[0,T]\times\mathbb{R}^d$

$$\partial_t u = a_{ij}\partial_i\partial_j u - \lambda u + B\cdot\nabla u + f, \quad u(0) = \varphi, \tag{3.3}$$

where $\lambda \geqslant 0$, $f \in \widetilde{\mathbb{L}}^{p_0}(T)$, $\varphi \in C_b^\infty$ and $a_{ij} := \frac{1}{2}\sum_{k=1}^d \sigma_{ik}\sigma_{jk}$. Under condition (3.1), by [48] there is a unique strong solution $X_{s,\cdot}^x$ to (3.2) for any $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$. The purpose of this section is to obtain some moment estimates for the following functionals of $X_{0,t}^x$

$$\int_0^T f(s, X_{0,\pi_h(s)}^x)\mathrm{d}s \quad \text{and} \quad \int_0^T \left[f(s, X_{0,s}^x, \mu_s^x) - f(s, X_{0,\pi_h(s)}^x, \mu_{\pi_h(s)}^x)\right]\mathrm{d}s,$$

where μ_t^x is the distribution of $X_{0,t}^x$ and $f \in \mathbb{L}_q^p(T)$ for some $2/q + d/p < 2$. The first integral is estimated by Krylov's type estimates. Compared to the case of smooth coefficients in [43], f in the second integral has no regularity. To overcome this obstacle, we use the observation mentioned in the introduction to replace $f(X_{0,t}^x) - f(X_{0,\pi_h(t)}^x)$ by $P_{0,t}^X f - P_{0,\pi_h(s)}^X f$, where P^X is the transition semi-group of X. Hence, we only need to obtain some time regularity results for the semigroup.

In Subsection 3.1, we consider the time-homogeneous case with $B \equiv 0$. By Girsanov's Theorem, we extend the results in Subsection 3.1 to $X_{s,t}^x$ in Subsection 3.2. Moreover, we obtain additional time regularity estimates for P^X by Duhamel's formula which can not be gotten from Girsanov's theorem. In the light of Duhamel's formula again, we also have two time regularity estimates for ∇u in Subsection 3.3, where u is the solution to (3.3).

For simplicity, throughout this section we set

$$\Xi := (d, T, p_0, \|B\|_{\widetilde{\mathbb{L}}^{p_0}(T)}, \kappa_1, \beta).$$

3.1. **Time regularity for solutions to SDE with no drift.** First of all, we recall the following generalized Itô formula from [48, Lemma 4.1-iii)] (see also [29, Theorem 3.7] for the original one).

Lemma 3.1 (Generalized Itô formula)**.** *Let $p, q \in [2, \infty)$ with $2/q + d/p < 1$. For any $T > 0$ and any $u \in \widetilde{\mathbb{H}}_q^{2,p}(T)$ with $\partial_t u \in \widetilde{\mathbb{L}}_q^p(T)$, we have for any $t \in [s, T]$ and $x \in \mathbb{R}^d$,*

$$\begin{aligned} u(t, X_{s,t}^x) = u(s, x) &+ \int_s^t (\partial_r u + a_{ij}\partial_i\partial_j u + B \cdot \nabla u)(r, X_{s,r}(x))\mathrm{d}r \\ &+ \int_s^t \nabla u(r, X_{s,r}(x))\mathrm{d}W_r. \end{aligned} \tag{3.4}$$

In this subsection, we consider the following case where $B \equiv 0$:

$$Z_t^x = x + \int_0^t \sigma(Z_s^x)\mathrm{d}W_s. \tag{3.5}$$

Define $P_t^\sigma f(x) := \mathbb{E} f(Z_t^x)$. By Proposition 2.7, there is a unique solution to the following second order parabolic PDE on $[0, T] \times \mathbb{R}^d$:

$$\partial_t u = a_{ij}\partial_i\partial_j u, \quad u_0 = \varphi. \tag{3.6}$$

Lemma 3.2. *Assume $(\mathbf{H}_\sigma)$ holds. Let $0 \leqslant s \leqslant t$, $\varphi \in C_b^\infty(\mathbb{R}^d)$, u and Z_t^x be the solution to (3.6) and (3.5) respectively. Then we have $\mathbb{P}$-a.s.*

$$\mathbb{E}\left[\varphi(Z_t^x)|\mathscr{F}_s\right] = u(t-s, Z_s^x). \tag{3.7}$$

In particular,

$$\mathbb{E}\left[\varphi(Z_t^x)|\mathscr{F}_s\right] = P_{t-s}^\sigma \varphi(Z_s^x) \quad \mathbb{P}-\text{a.s.} \tag{3.8}$$

Moreover, for any $t \geqslant 0$ and $f \in C_b^2$,

$$P_t^\sigma f - f = \int_0^t P_r^\sigma(a_{ij}\partial_i\partial_j f)\mathrm{d}r. \tag{3.9}$$

Proof. For all $t > 0$, applying the generalized Itô formula (3.4) to $s \mapsto u(t-s, Z_s^x)$, we have

$$u(0, Z_t^x) = u(t-s, Z_s^x) + \int_s^t \left(-\partial_r u(t-r, Z_r^x) + a_{ij}\partial_i\partial_j u(t-r, Z_r^x)\right)\mathrm{d}r + \int_s^t \nabla u(t-r, Z_r^x)\mathrm{d}W_r.$$

Noting that $u(t-s, Z_s^x)$ is $\mathscr{F}_s$-measurable, and taking conditional expectation with respect to $\mathscr{F}_s$ on both sides, we have

$$\mathbb{E}\left[\varphi(Z_t^x)|\mathscr{F}_s\right] = u(t-s, Z_s^x) \quad \mathbb{P}-\text{a.s.},$$

which for $s = 0$ implies that

$$P_t^\sigma \varphi(x) = \mathbb{E}\varphi(Z_t^x) = u(t, x).$$

Then (3.8) is straightforward from (3.7). For (3.9), since $f \in C_b^2$, we use the classical Itô formula and have

$$f(Z_t^x) = f(x) + \int_0^t a_{ij}\partial_i\partial_j f(Z_r^x)\mathrm{d}r + \int_0^t \nabla f(Z_r^x)\mathrm{d}W_r.$$

Then, we have (3.9) by taking expectation and complete the proof. $\square$

Based on Lemma 3.2 and the uniqueness of (3.6), we have the following Chapman-Kolmogorov equations

$$P_s^\sigma P_t^\sigma = P_{s+t}^\sigma. \tag{3.10}$$

Set

$$\rho_t(x) := (2\pi t)^{-d/2} e^{-|x|^2/(2t)}.$$

Then the following lemma is from [13, Theorem 2.3].

Lemma 3.3. *Assume* $(\mathbf{H}_\sigma)$ *holds. Then there is a unique function* $p_\cdot^\sigma(\cdot,\cdot) : \mathbb{R}_+ \times \mathbb{R}^{2d} \to \mathbb{R}$ *such that for any* $j = 0, 1, 2$

$$|\nabla_x^j p_t^\sigma(x,y)| \leqslant c_1 t^{-\frac{j}{2}} \rho_{c_2 t}(x-y) \tag{3.11}$$

and

$$P_t^\sigma f(x) = \int_{\mathbb{R}^d} f(y) p_t^\sigma(x,y)\mathrm{d}y \tag{3.12}$$

for any $f \in C(\mathbb{R}^d)$*, where* c_1 *and* c_2 *are positive constants depending on* Ξ*.*

For any $h \in (0,1)$, recall that $\pi_h(t) := t$ for $t \in [0,h)$ and

$$\pi_h(t) := kh, \quad t \in [kh, (k+1)h), \ k \geqslant 1.$$

Remark 3.4. *The reason why we define* $\pi_h(t) = t$ *for* $t \in [0,h)$ *is that the function space here is* L^p*. If the initial data don't have an* L^q *density,* $\mathbb{E}f(Z_{\pi_h(t)}) = \mathbb{E}f(Z_0)$*,* $f \in L^p$*, will blow up for all* $t < h$*.*

Now we give the following Krylov estimate and Khasminskii estimate.

Lemma 3.5. *Assume* $(\mathbf{H}_\sigma)$ *holds. For any* $T > 0$*,* $k = 0, 1, 2$*,* $p \in [1,\infty]$ *and* $q \in [p,\infty]$*, there is a constant* $C = C(\Xi, p, q)$ *such that for all* $0 \leqslant s < t \leqslant T$*,* $x \in \mathbb{R}^d$ *and nonnegative functions* $f \in \widetilde{L}^p$

$$\|\nabla^k P_t^\sigma f\|_q \leqslant C t^{-k/2 - d/(2p) + d/(2q)} \|f\|_p \tag{3.13}$$

and for $2/q + d/p < 2$*,* $h > 0$ *and nonnegative functions* $f \in \widetilde{\mathbb{L}}_q^p(T)$

$$\mathbb{E}\int_s^t f(r, Z_r^x)\mathrm{d}r + \mathbb{E}\int_s^t f(r, Z_{\pi_h(r)}^x)\mathrm{d}r \leqslant C(t-s)^{1-\frac{1}{q}-\frac{d}{2p}} \|f\|_{\widetilde{\mathbb{L}}_q^p(T)}. \tag{3.14}$$

Moreover, for any $f \in \widetilde{\mathbb{L}}_q^p(T)$ *with* $d/p + 2/q < 2$*,*

$$\sup_{x \in \mathbb{R}^d} \mathbb{E}\exp\left(\int_0^T f(t, Z_t^x)\mathrm{d}t\right) < \infty. \tag{3.15}$$

Proof. Without loss of generality we assume that $c_2 = 1$ in (3.11). Combining Lemma 3.3 and Young's convolution inequality, one sees that

$$\begin{aligned}
\|\nabla^k P_t^\sigma f\|_q &\lesssim \Big\| \int_{\mathbb{R}^d} f(y) \nabla_x^k p_t^\sigma(\cdot, y)\mathrm{d}y \Big\|_q \\
&\lesssim t^{-k/2} \|\rho_t * f\|_q \lesssim t^{-k/2} \sup_w \Big\| \mathbf{1}_{|\cdot - w| \leqslant 1} \int_{\mathbb{R}^d} f(\cdot - y)\rho_t(y)\mathrm{d}y \Big\|_q \\
&\lesssim t^{-k/2} \sup_w \Big\| \mathbf{1}_{|\cdot - w| \leqslant 1} \frac{1}{|B_1|} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_{|y-z|\leqslant 1} f(\cdot - y) \mathbf{1}_{|y-z|\leqslant 1} \rho_t(y)\mathrm{d}y\mathrm{d}z \Big\|_q
\end{aligned}$$

$$
\begin{aligned}
&\lesssim t^{-k/2} \sup_w \| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_{|\cdot-y-w+z|\leqslant 2} f(\cdot - y) \mathbf{1}_{|y-z|\leqslant 1} \rho_t(y) \mathrm{d}y \mathrm{d}z \|_q \\
&\lesssim t^{-k/2} \sup_w \int_{\mathbb{R}^d} \| \int_{\mathbb{R}^d} \mathbf{1}_{|\cdot-y-w+z|\leqslant 2} f(\cdot - y) \mathbf{1}_{|y-z|\leqslant 1} \rho_t(y) \mathrm{d}y \|_q \mathrm{d}z \\
&\lesssim t^{-k/2} \int_{\mathbb{R}^d} \sup_w \| \mathbf{1}_{|\cdot-w+z|\leqslant 2} f(\cdot) \|_p \| \mathbf{1}_{|\cdot-z|\leqslant 1} \rho_t(\cdot) \|_r \mathrm{d}z \\
&\lesssim t^{-k/2} \int_{\mathbb{R}^d} \left(\int_{|y-z|\leqslant 1} |\rho_t(y)|^r \mathrm{d}y \right)^{1/r} \mathrm{d}z \|\!| f |\!\|_p,
\end{aligned}
$$

where $1 + 1/q = 1/r + 1/p$. Next, one sees that

$$
\int_{\mathbb{R}^d} \left(\int_{|y-z|\leqslant 1} (\rho_t(y))^r \mathrm{d}y \right)^{1/r} \mathrm{d}z \lesssim \|\rho_t\|_r + \int_{|z|>2} \left(\int_{|y-z|\leqslant 1} (\rho_t(y))^r \mathrm{d}y \right)^{1/r} \mathrm{d}z.
$$

We note that

$$
|z| > 2, |y - z| \leqslant 1 \Rightarrow |y| \geqslant |z| - |y - z| \geqslant \frac{|z|}{2},
$$

which implies that $\rho_t(y) \lesssim \rho_t(z/2)$, and $\|\rho_t\|_r \lesssim t^{-d/2+d/(2r)} = t^{-d/(2p)+d/(2q)}$, we have

$$
\begin{aligned}
\int_{\mathbb{R}^d} \left(\int_{|y-z|\leqslant 1} (\rho_t(y))^r \mathrm{d}y \right)^{1/r} \mathrm{d}z &\lesssim \|\rho_t\|_r + \int_{|z|>2} \left(\int_{|y-z|\leqslant 1} (\rho_t(z/2))^r \mathrm{d}y \right)^{1/r} \mathrm{d}z \\
&\lesssim t^{-d/(2p)+d/(2q)} + \int_{|z|>2} \rho_t(z/2) \mathrm{d}z \lesssim t^{-d/(2p)+d/(2q)} + 1
\end{aligned}
$$

and obtain (3.13).

Now we show (3.14). Set $p' := \frac{p}{p-1}$ and $q' := \frac{q}{q-1}$. Without loss of generality, we take $s = 0$. By (3.13), for any $h > 0$,

$$
\mathbb{E} \int_0^t f(s, Z^x_{\pi_h(s)}) \mathrm{d}s \lesssim \int_{\mathbb{R}^d} \Big(\int_0^t \big(\int_{|y-z|\leqslant 1} |\rho_{\pi_h(s)}(y)|^{p'} \mathrm{d}y \big)^{q'/p'} \mathrm{d}s \Big)^{1/q'} \mathrm{d}z \|\!| f |\!\|_{\widetilde{\mathbb{L}}^p_q(T)}.
$$

Then, we have

$$
\begin{aligned}
\mathscr{I} &:= \int_{\mathbb{R}^d} \Big(\int_0^t \big(\int_{|y-z|\leqslant 1} |\rho_{\pi_h(s)}(y)|^{p'} \mathrm{d}y \big)^{q'/p'} \mathrm{d}s \Big)^{1/q'} \mathrm{d}z \\
&\lesssim \big(\int_0^t \|\rho_{\pi_h(s)}\|^{q'}_{p'} \mathrm{d}s \big)^{1/q'} + \int_{|z|>2} \Big(\int_0^t |\rho_{\pi_h(s)}(z/2)|^{q'} \mathrm{d}s \Big)^{1/q'} \mathrm{d}z \\
&\lesssim \big(\int_0^t (\pi_h(s))^{-dq'/(2p)} \mathrm{d}s \big)^{1/q'} + t \int_{|z|>2} |z|^{-d} \exp(-|z|^2/(16T)) \mathrm{d}z \\
&\lesssim t^{1-1/q-d/(2p)} + t,
\end{aligned}
$$

since $dq'/(2p) < 1$ and $\rho_s(z) \leqslant C|z|^{-d} \exp(-\frac{|z|^2}{4T})$ for all $s < T$. Similarly, we obtain

$$
\mathbb{E} \int_0^t f(s, Z^x_s) \mathrm{d}s \lesssim t^{1-\frac{1}{q}-\frac{d}{2p}} + t.
$$

Finally, noting that by (3.8)

$$
\mathbb{E} \Big[\int_s^t f(s, Z^x_r) \mathrm{d}r \Big| \mathscr{F}_s \Big] = \mathbb{E} \int_s^t f(s, Z^y_{r-s}) \mathrm{d}r \Big|_{y=Z^x_s},
$$

(3.15) is direct from (3.14) (see [49, Corollary 3.5] for example) and we complete the proof.

∎

Remark 3.6. *We note that for any fixed $h>0$ and $x\in\mathbb{R}^d$*

$$\mathbb{E}\exp\big(\int_0^T f(t,Z^x_{\pi_h(t)})\mathrm{d}t\big)<\infty$$

is not true. For example, letting $Z^x_t=W_t$, when $d\geqslant 2$ and $f(t,x)=|x|^{-1/2}\in\widetilde{L}^{d+1}(\mathbb{R}^d)$, we have

$$\mathbb{E}\exp\left(\int_h^{2h} f(t,W_{\pi_h(t)})\mathrm{d}t\right)=\mathbb{E}\exp(h|W_h|^{-1/2})$$

$$=\int_{\mathbb{R}^d} e^{\frac{h}{\sqrt{|x|}}}\rho_h(x)\mathrm{d}x\geqslant\frac{h^{2d}}{(2d)!}\int_{\mathbb{R}^d}\frac{1}{|x|^d}\rho_h(x)\mathrm{d}x=\infty.$$

Lemma 3.7. *Let $T>0$, $k=0,1$, and $1\leqslant p\leqslant q\leqslant\infty$. Then there is a constant $C=C(\Xi,p,q)$ such that for all $0<s\leqslant t\leqslant T$,*

$$\|\nabla^k(P^\sigma_t\varphi-P^\sigma_s\varphi)\|_q\leqslant C\Big([(t-s)^{\frac{2-k}{2}}s^{\frac{k-2}{2}}]\wedge 1\Big)s^{-\frac{k}{2}-\frac{d}{2p}+\frac{d}{2q}}\|\varphi\|_p. \tag{3.16}$$

Proof. Based on (3.10) and (3.9), one sees that

$$P^\sigma_t\varphi-P^\sigma_s\varphi=P^\sigma_{t-s}(P^\sigma_s\varphi)-P^\sigma_s\varphi=\int_0^{t-s}P^\sigma_r(a_{ij}\partial_i\partial_jP^\sigma_s\varphi)\mathrm{d}r.$$

By (3.13), we have

$$\begin{aligned}\|\nabla^k(P^\sigma_t\varphi-P^\sigma_s\varphi)\|_q&\lesssim\int_0^{t-s}r^{-\frac{k}{2}}\|\nabla^2P^\sigma_s\varphi\|_q\mathrm{d}r\\&\lesssim\int_0^{t-s}r^{-\frac{k}{2}}s^{-1+\frac{d}{2q}-\frac{d}{2p}}\|\varphi\|_p\mathrm{d}r\\&\lesssim\Big[(t-s)^{\frac{2-k}{2}}s^{\frac{k-2}{2}}\Big]s^{-\frac{k}{2}-\frac{d}{2p}+\frac{d}{2q}}\|\varphi\|_p.\end{aligned}$$

Moreover, noting that by (3.13)

$$\begin{aligned}\|\nabla^k(P^\sigma_t\varphi-P^\sigma_s\varphi)\|_q&\leqslant\|\nabla^kP^\sigma_t\varphi\|_q+\|\nabla^kP^\sigma_s\varphi\|_q\\&\lesssim s^{-\frac{k}{2}-\frac{d}{2p}+\frac{d}{2q}}\|\varphi\|_p\end{aligned}$$

for $s\leqslant t$, we complete the proof. □

When $p=\infty$ and $\sigma\equiv\mathbb{I}$, the following lemma has been proved in [17, Lemma 2.1] for Brownian motion. For classical L^p spaces, Lê and Ling obtained these results by the stochastic sewing lemma in [32]. For the localized L^p space, we provide a different proof here, which is based on (3.16) and (3.13).

Lemma 3.8. *Assume $(\mathbf{H}_\sigma)$ holds. Then for any $T>0$ and $p\in(d\vee 2,\infty)$, there is a constant $C=C(\Xi,p)$ such that for any stopping time $\tau\leqslant T$, $h\in(0,1)$, $x\in\mathbb{R}^d$ and $f\in\widetilde{\mathbb{L}}^p(T)$,*

$$\mathbb{E}\Big|\int_0^\tau(f(t,Z^x_t)-f(t,Z^x_{\pi_h(t)}))\mathrm{d}t\Big|^2\leqslant Ch\ln h^{-1}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}. \tag{3.17}$$

Proof. We divide the proof into four steps. In Step 1, we prove that

$$\mathbb{E}\Big|\int_s^t(f(r,Z^x_r)-f(r,Z^x_{\pi_h(r)}))\mathrm{d}r\Big|^2\leqslant C\big(hs^{-\frac{d}{2p}}(t-s)^{1-\frac{d}{2p}}+h\ln h^{-1}(t-s)^{1-\frac{d}{p}}\big)\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}; \tag{3.18}$$

In Step 2, we show (3.17) for $\tau=T$; In Step 3, we show

$$\sup_{a,b\in[0,T]}\mathbb{E}\left(\int_a^b(f(t,Z^x_t)-f(t,Z^x_{\pi_h(t)}))\mathrm{d}t\right)^2\leqslant Ch\ln h^{-1}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)}; \tag{3.19}$$

In Step 4, we show (3.17) for any stopping time τ with $\tau\leqslant T$.

(Step 1) For simplicity of notation, we drop the time t in $f(t,x)$. First, we note that by Hölder's inequality and (3.14),

$$\mathbb{E}\left(\int_0^{2h} f(Z_t^x)-f(Z_{\pi_h(t)}^x)\mathrm{d}t\right)^2 \leqslant 2h\mathbb{E}\int_0^{2h}\Big|f(Z_t^x)-f(Z_{\pi_h(t)}^x)\Big|^2\mathrm{d}t \lesssim h|\!|\!|f|\!|\!|_p.$$

Hence, without loss of generality, we may assume $s>2h$. The symmetry implies

$$\begin{aligned}
&\mathbb{E}\Big|\int_s^t (f(Z_r^x)-f(Z_{\pi_h(r)}^x))\mathrm{d}r\Big|^2\\
=&2\int_s^t\int_{r_1}^t \mathbb{E}\Big[(f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x))(f(Z_{r_2}^x)-f(Z_{\pi_h(r_2)}^x))\Big]\mathrm{d}r_2\mathrm{d}r_1\\
=&2\int_s^t\int_{r_1}^{r_1+h} \mathbb{E}\Big[(f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x))(f(Z_{r_2}^x)-f(Z_{\pi_h(r_2)}^x))\Big]\mathrm{d}r_2\mathrm{d}r_1\\
&+2\int_s^t\int_{r_1+h}^{t} \mathbb{E}\Big[(f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x))(f(Z_{r_2}^x)-f(Z_{\pi_h(r_2)}^x))\Big]\mathrm{d}r_2\mathrm{d}r_1\\
:=&2\mathscr{I}_1+2\mathscr{I}_2.
\end{aligned}$$

By Hölder's inequality and (3.13), one sees that

$$\begin{aligned}
&\mathbb{E}\Big[(f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x))(f(Z_{r_2}^x)-f(Z_{\pi_h(r_2)}^x))\Big]\\
\leqslant&\left(\mathbb{E}\Big|f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x)\Big|^2\right)^{1/2}\left(\mathbb{E}\Big|f(Z_{r_2}^x)-f(Z_{\pi_h(r_2)}^x)\Big|^2\right)^{1/2}\\
\lesssim&(r_1^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})(r_2^{-\frac{d}{2p}}+\pi_h(r_2)^{-\frac{d}{2p}})|\!|\!|f|\!|\!|_p^2,
\end{aligned}$$

which implies that

$$\begin{aligned}
\mathscr{I}_1&\lesssim|\!|\!|f|\!|\!|_p^2\int_s^t\int_{r_1}^{r_1+h}(r_1^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})(r_2^{-\frac{d}{2p}}+\pi_h(r_2)^{-\frac{d}{2p}})\mathrm{d}r_2\mathrm{d}r_1\\
&\lesssim|\!|\!|f|\!|\!|_p^2(s-h)^{-\frac{d}{2p}}h\int_s^t(r_1^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})\mathrm{d}r_1\lesssim hs^{-\frac{d}{2p}}(t-s)^{1-\frac{d}{2p}}|\!|\!|f|\!|\!|_p^2.
\end{aligned}$$

For $\mathscr{I}_2$, we use conditional expectation and the Markov property (3.8). For simplicity, let

$$\mathbb{E}^{\mathscr{G}}[\cdot]:=\mathbb{E}[\cdot|\mathscr{G}].$$

Then, noting that $r_1\leqslant r_2-h<r_2$, by the Markov property (3.8), we have

$$\begin{aligned}
\mathscr{I}_2&=\int_s^t\int_{r_1+h}^t\mathbb{E}\Big((f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x))\mathbb{E}^{\mathscr{F}_{r_1}}[f(Z_{r_2}^x)-f(Z_{\pi_h(r_2)}^x)]\Big)\mathrm{d}r_2\mathrm{d}r_1\\
&=\int_s^t\int_{r_1+h}^t\mathbb{E}\Big(\Big[f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x)\Big]\Big[P_{r_2-r_1}^\sigma f(Z_{r_1}^x)-P_{\pi_h(r_2)-r_1}^\sigma f(Z_{r_1}^x)\Big]\Big)\mathrm{d}r_2\mathrm{d}r_1.
\end{aligned}$$

Setting $G:=P_{r_2-r_1}^\sigma f-P_{\pi_h(r_2)-r_1}^\sigma f$, in view of Hölder's inequality and (3.13), one sees that

$$\begin{aligned}
\mathscr{I}_2&\leqslant\int_s^t\int_{r_1+h}^t\left(\mathbb{E}\Big|f(Z_{r_1}^x)-f(Z_{\pi_h(r_1)}^x)\Big|^2\right)^{1/2}\left(\mathbb{E}|G(Z_{r_1}^x)|^2\right)^{1/2}\mathrm{d}r_2\mathrm{d}r_1\\
&\lesssim|\!|\!||f|^2|\!|\!|_{p/2}^{1/2}\int_s^t\int_{r_1+h}^t(r_1{}^{-\frac{d}{2p}}+\pi_h(r_1)^{-\frac{d}{2p}})r_1{}^{-\frac{d}{2p}}|\!|\!||G|^2|\!|\!|_{p/2}^{1/2}\mathrm{d}r_2\mathrm{d}r_1\\
&\lesssim|\!|\!|f|\!|\!|_p\int_s^t\int_{r_1+h}^t(r_1-h)^{-\frac{d}{p}}|\!|\!|G|\!|\!|_p\mathrm{d}r_2\mathrm{d}r_1.
\end{aligned}$$

We note that by (3.16),

$$|\!|\!|G|\!|\!|_p=|\!|\!|P_{r_2-r_1}^\sigma f-P_{\pi_h(r_2)-r_1}^\sigma f|\!|\!|_p$$

$$\lesssim \Big([(r_2-\pi_h(r_2))(\pi_h(r_2)-r_1)^{-1}]\wedge 1\Big)\|\!|f|\!\|_p$$
$$\lesssim \Big[(h(r_2-r_1-h)^{-1})\wedge 1\Big]\|\!|f|\!\|_p,$$

By a change of variables we have

$$\begin{aligned}\mathscr{I}_2 &\lesssim \|\!|f|\!\|_p^2\int_s^t\int_{r_1+h}^t (r_1-h)^{-\frac{d}{p}}\Big[(h(r_2-r_1-h)^{-1})\wedge 1\Big]\mathrm{d}r_2\mathrm{d}r_1\\ &\lesssim \|\!|f|\!\|_p^2\int_{s-h}^{t-h}\int_0^t (r_1)^{-\frac{d}{p}}\Big[(h(r_2)^{-1})\wedge 1\Big]\mathrm{d}r_2\mathrm{d}r_1\\ &\lesssim h(t-s)^{1-\frac{d}{p}}\|\!|f|\!\|_p^2\int_0^{t/h}\Big[(r_2)^{-1}\wedge 1\Big]\mathrm{d}r_2 \lesssim h\ln h^{-1}(t-s)^{1-\frac{d}{p}}\|\!|f|\!\|_p^2,\end{aligned}$$

and we obtain (3.18).

(Step 2) In this step, we use the method in [32, Lemma 3.4] to show (3.17) with $\tau=T$. Let $\alpha:=\frac{d}{4p}$, $\|\cdot\|:=\big(\mathbb{E}|\cdot|^2\big)^{\frac{1}{2}}$,

$$F_t:=\int_0^t(f(r,Z_r^x)-f(r,Z^x_{\pi_h(r)}))\mathrm{d}r \quad\text{and}\quad \mathcal{H}:=\big(Ch\ln h^{-1}\big)^{1/2}\|\!|f|\!\|_{\widetilde{\mathbb{L}}^p(T)},$$

where C is the constant in (3.18). By (3.18), we have

$$\|F_t-F_s\|\leqslant \mathcal{H}(s^{-\alpha}(t-s)^{1/2-\alpha}+(t-s)^{1/2-2\alpha}). \tag{3.20}$$

For any $n\in\mathbb{N}$, set

$$t_n:=2^{-n}T.$$

Then, by (3.20)

$$\|F_T\|\leqslant\sum_{n=0}^\infty\|F_{t_n}-F_{t_{n+1}}\|\leqslant\mathcal{H}\sum_{n=0}^\infty(t_{n+1}^{-\alpha}(t_n-t_{n+1})^{1/2-\alpha}+(t_n-t_{n+1})^{1/2-2\alpha}).$$

Noting that $t_{n+1}=T2^{-n-1}$ and $t_n-t_{n+1}=T2^{-n-1}$, we have

$$\begin{aligned}\|F_T\|&\leqslant\mathcal{H}\sum_{n=0}^\infty(T^{-\alpha}2^{\alpha(n+1)}(t_n-t_{n+1})^{1/2-\alpha}+(t_n-t_{n+1})^{1/2-2\alpha})\\ &\leqslant 2\mathcal{H}T^{1/2-2\alpha}\sum_{n=0}^\infty 2^{-(1/2-2\alpha)(n+1)}\leqslant 2\mathcal{H}T^{1/2-2\alpha}\end{aligned}$$

and obtain (3.17) for $\tau=T$.

(Step 3) For any $a,b\in[0,T]$, define

$$f_{a,b}(t,x):=\mathbf{1}_{t\in[a,b]}f(t,x).$$

Then, by Step 1, one sees that

$$\begin{aligned}\mathbb{E}\left(\int_a^b(f(t,Z_t^x)-f(t,Z^x_{\pi_h(t)}))\mathrm{d}t\right)^2&=\mathbb{E}\big(\int_0^T(f_{a,b}(t,Z_t^x)-f_{a,b}(t,Z^x_{\pi_h(t)}))\mathrm{d}t\big)^2\\ &\leqslant Ch\ln h^{-1}\|\!|f_{a,b}|\!\|^2_{\widetilde{\mathbb{L}}^p(T)}.\end{aligned}$$

Noting that $\|\!|f_{a,b}|\!\|_{\widetilde{\mathbb{L}}^p(T)}\leqslant\|\!|f|\!\|_{\widetilde{\mathbb{L}}^p(T)}$, we get (3.19).

(Step 4) Without loss of generality, we assume that $h<T/2$ and that τ only takes finite values $a_1,a_2,...,a_n\in[0,T]$. Otherwise, for any stopping time, we choose $\tau_n,n\in\mathbb{N}$, which only take finite

values to approximate τ and (3.17) follows from (3.14) and the dominated convergence theorem. First, we have

$$\mathbb{E}\left|\int_\tau^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\right|^2 \leqslant 2\mathbb{E}\left|\int_\tau^{(\tau+2h)\wedge T}(f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\right|^2 + 2\mathbb{E}\left|\int_{(\tau+2h)\wedge T}^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\right|^2.$$

We note that by Hölder's inequality and (3.14) for $p/2>d/2$,

$$\begin{aligned}&\mathbb{E}\Big|\int_\tau^{(\tau+2h)\wedge T}(f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\Big|^2\\ &\leqslant \mathbb{E}\Big|\int_0^T \mathbf{1}_{t\in[\tau,\tau+2h]}\mathrm{d}t\int_0^T |f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x)|^2\mathrm{d}t\Big|\\ &\lesssim h|\!|\!| |f|^2|\!|\!|_{\widetilde{\mathbb{L}}^{p/2}(T)} \lesssim h|\!|\!| f|\!|\!|^2_{\widetilde{\mathbb{L}}^p(T)}.\end{aligned}$$

Now we estimate the second term. In fact,

$$\begin{aligned}&\mathbb{E}\left(\int_{(\tau+2h)\wedge T}^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\right)^2\\ &=\sum_{i=1}^n \mathbb{E}\left[\mathbf{1}_{\tau=a_i}\left(\int_{(a_i+2h)\wedge T}^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\right)^2\right].\end{aligned}$$

We note that $\mathbf{1}_{\tau=a_i}\in\mathscr{F}_{a_i}\subset\mathscr{F}_{([a_i/h]+1)h}$, without loss of generality, assuming $a_i+2h<T$, one sees that

$$\begin{aligned}&\mathbb{E}\left[\mathbf{1}_{\tau=a_i}\left(\int_{a_i+2h}^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\right)^2\right]\\ &=\mathbb{E}\Big(\mathbf{1}_{\tau=a_i}\mathbb{E}\Big[\Big(\int_{a_i+2h}^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\Big)^2\Big|\mathscr{F}_{a_i}\Big]\Big)\\ &:=\mathbb{E}(\mathbf{1}_{\tau=a_i}\mathscr{A}_i),\end{aligned}$$

where

$$\mathscr{A}_i=\mathbb{E}\left[\mathbb{E}\left[\left(\int_{a_i+2h}^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\right)^2\Big|\mathscr{F}_{([a_i/h]+1)h}\right]\Big|\mathscr{F}_{a_i}\right].$$

Moreover, by the Markov property (3.8), we have

$$\begin{aligned}&\mathbb{E}\Big[\Big(\int_{a_i+2h}^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\Big)^2\Big|\mathscr{F}_{([a_i/h]+1)h}\Big]\\ &=\mathbb{E}\Big(\int_{a_i+2h}^T (f(t,Z^y_{t-([a_i/h]+1)h})-f(t,Z^y_{\pi_h(t)-([a_i/h]+1)h}))\mathrm{d}t\Big)^2\Big|_{y=Z^x_{([a_i/h]+1)h}}\\ &=\mathbb{E}\Big(\int_{a_i+2h-([a_i/h]+1)h}^{T-([a_i/h]+1)h}(f(t+([a_i/h]+1)h,Z_t^y)-f(t+([a_i/h]+1)h,Z^y_{\pi_h(t)}))\mathrm{d}t\Big)^2\Big|_{y=Z^x_{([a_i/h]+1)h}}\\ &\lesssim h\ln h^{-1}|\!|\!| f|\!|\!|^2_{\widetilde{\mathbb{L}}^p(T)}\end{aligned}$$

by (3.19). Therefore, we have

$$\mathbb{E}\Big|\int_\tau^T (f(t,Z_t^x)-f(t,Z_{\pi_h(t)}^x))\mathrm{d}t\Big|^2 \lesssim h|\!|\!| f|\!|\!|^2_{\widetilde{\mathbb{L}}^p(T)}+\sum_{i=1}^n \mathbb{E}(\mathbf{1}_{\tau=a_i}\mathscr{A}_i)$$

$$\lesssim h\|\!|f|\!\|^2_{\widetilde{\mathbb{L}}^p(T)} + h\ln h^{-1}\|\!|f|\!\|^2_{\widetilde{\mathbb{L}}^p(T)}\sum_{i=1}^n \mathbb{E}\mathbf{1}_{\tau=a_i}$$
$$\lesssim h\ln h^{-1}\|\!|f|\!\|^2_{\widetilde{\mathbb{L}}^p(T)},$$

which implies that

$$\begin{aligned}&\mathbb{E}\left|\int_0^\tau (f(t,Z^x_t)-f(t,Z^x_{\pi_h(t)}))\mathrm{d}t\right|^2\\ &\leqslant 2\mathbb{E}\left|\int_0^T (f(t,Z^x_t)-f(t,Z^x_{\pi_h(t)}))\mathrm{d}t\right|^2 + 2\mathbb{E}\left|\int_\tau^T (f(t,Z^x_t)-f(t,Z^x_{\pi_h(t)}))\mathrm{d}t\right|^2\\ &\lesssim h\ln h^{-1}\|\!|f|\!\|^2_{\widetilde{\mathbb{L}}^p(T)}\end{aligned}$$

and completes the proof.

□

Corollary 3.9. *Assume* $(\mathbf{H}_\sigma)$ *holds. For any* $T>0$*,* $p\in(d\vee 2,\infty)$ *and* $\delta>0$*, there is a constant* $C=C(\Xi,p)$ *such that for any* $x\in\mathbb{R}^d$ *and* $f\in\widetilde{\mathbb{L}}^p(T)$*,*

$$\mathbb{E}\left(\sup_{t\in[0,T]}\left|\int_0^t (f(s,Z^x_s)-f(s,Z^x_{\pi_h(s)}))\mathrm{d}s\right|^2\right)\leqslant Ch^{1-\delta}\|\!|f|\!\|^2_{\widetilde{\mathbb{L}}^p(T)}. \tag{3.21}$$

Proof. Let

$$\eta_t := \left|\int_0^t (f(s,Z^x_s)-f(s,Z^x_{\pi_h(s)}))\mathrm{d}s\right|^2$$

and $\eta^*_t := \sup_{s\in[0,t]}\eta_s$. First of all, it follows from Hölder's inequality and (3.14) that for any $\gamma\in(1,p/(d\vee 2))$,

$$\begin{aligned}\mathbb{E}(\eta^*_T)^\gamma &\lesssim \mathbb{E}\int_0^T \left(|f|^{2\gamma}(s,Z^x_s)+|f|^{2\gamma}(s,Z^x_{\pi_h(s)})\right)\mathrm{d}s\\ &\lesssim \|\!|\,|f|^{2\gamma}|\!\|_{\widetilde{\mathbb{L}}^{p/(2\gamma)}(T)} \lesssim \|\!|f|\!\|^{2\gamma}_{\widetilde{\mathbb{L}}^p(T)}.\end{aligned} \tag{3.22}$$

For any $\lambda>0$, let

$$\tau_\lambda := \inf\{t\geqslant 0,\ \eta_t>\lambda\}.$$

We note that $\eta_{\tau_\lambda}=\lambda$, since η is a continuous process. Then,

$$\lambda\mathbb{P}(\eta^*_T>\lambda)\leqslant\lambda\mathbb{P}(\tau_\lambda\leqslant T)\leqslant\mathbb{E}\big(\eta_{\tau_\lambda}\mathbf{1}_{\{\tau_\lambda\leqslant T\}}\big)\leqslant\mathbb{E}\eta_{\tau_\lambda\wedge T}.$$

In view of (3.17), we have

$$\lambda\mathbb{P}(\eta^*_T>\lambda)\lesssim h\ln h^{-1}\|\!|f|\!\|^2_{\widetilde{\mathbb{L}}^p(T)}.$$

Set $\Xi_h := h\ln h^{-1}\|\!|f|\!\|^2_{\widetilde{\mathbb{L}}^p(T)}$. Then, for any $\delta\in(0,1)$, by a change of variables,

$$\begin{aligned}\mathbb{E}\big(\eta^*_T\big)^{1-\delta} &= (1-\delta)\int_0^\infty \lambda^{-\delta}\mathbb{P}(\eta^*_T>\lambda)\mathrm{d}\lambda\\ &\lesssim \int_0^\infty \lambda^{-\delta}\big(1\wedge(\Xi_h\lambda^{-1})\big)\mathrm{d}\lambda\\ &\lesssim \Xi_h^{1-\delta}\int_0^\infty \lambda^{-\delta}(1\wedge\lambda^{-1})\mathrm{d}\lambda \lesssim (h\ln h^{-1})^{(1-\delta)}\|\!|f|\!\|^{2(1-\delta)}_{\widetilde{\mathbb{L}}^p(T)}.\end{aligned} \tag{3.23}$$

Combining (3.22) and (3.23), in view of Hölder's inequality, for any $\delta>0$ small enough, we have

$$\mathbb{E}\eta^*_T = \mathbb{E}\left[\big(\eta^*_T\big)^{1-\delta-\sqrt{\delta}}\big(\eta^*_T\big)^{\delta+\sqrt{\delta}}\right]$$

$$\lesssim \left[\mathbb{E}\big(\eta_T^*\big)^{1-\delta}\right]^{\frac{1-\delta-\sqrt{\delta}}{1-\delta}} \left[\mathbb{E}\big(\eta_T^*\big)^{(\sqrt{\delta}+1)(1-\delta)}\right]^{\frac{\sqrt{\delta}}{1-\delta}}$$
$$\lesssim h^{1-\delta-\sqrt{\delta}}(\ln h^{-1})^{1-\delta-\sqrt{\delta}}|\!|\!| f|\!|\!|^2_{\widetilde{\mathbb{L}}^p(T)}$$

and complete the proof.

□

3.2. **Time discretization for SDEs with $\widetilde{\mathbb{L}}^p$ drift.** Now, let us extend estimate (3.21) from Z^x_{t-s} to the solution $X^x_{s,t}$ of SDE (3.2) both in the sense of paths and distributions (see (3.28) and (3.37) below). Recall that

$$X^x_{s,t} = x + \int_s^t B(r, X^x_{s,r})\mathrm{d}r + \int_s^t \sigma(X^x_{s,r})\mathrm{d}W_r. \tag{3.24}$$

Let $\mu^x_{s,t}$ denote the distribution of $X^x_{s,t}$. For simplicity, we also set

$$X^x_t := X^x_{0,t} \quad \text{and} \quad \mu^x_t := \mu^x_{0,t}.$$

The following estimates follow from Girsanov's transform and estimates for Z^x_t (see [48, Lemma 4.1] for (ii) and (iii)).

Lemma 3.10. *Assume (3.1) with $p_0 > d \vee 2$.*
(i) For any $T > 0$ and $p \in (1,\infty)$, there is a constant $C = C(\Xi, p)$ such that for any $x \in \mathbb{R}^d$, $0 \leqslant s < t \leqslant T$ and nonnegative $f \in \widetilde{L}^p$,

$$\mathbb{E} f(X^x_t) \leqslant C t^{-d/(2p)} |\!|\!| f|\!|\!|_p. \tag{3.25}$$

(ii)For any $T > 0$ and $2/q + d/p < 2$, there is a constant $C = C(\Xi, p, q)$ such that for any $x \in \mathbb{R}^d$, $0 \leqslant s < t \leqslant T$ and nonnegative $f \in \widetilde{\mathbb{L}}^p_q(T)$,

$$\mathbb{E}\int_s^t f(r, X^x_r)\mathrm{d}r + \mathbb{E}\int_s^t f(r, X^x_{\pi_h(r)})\mathrm{d}r \leqslant C(t-s)^{1-\frac{1}{q}-\frac{d}{2p}} |\!|\!| f|\!|\!|_{\widetilde{\mathbb{L}}^p_q(T)}. \tag{3.26}$$

(iii)For any $T > 0$, $d/p + 2/q < 2$ and $f \in \widetilde{\mathbb{L}}^p_q(T)$,

$$\sup_{x\in\mathbb{R}^d} \mathbb{E}\exp\left(\int_0^T f(t, X^x_t)\mathrm{d}t\right) < \infty. \tag{3.27}$$

(iv)For any $T > 0$, $\delta > 0$ and $p > d \vee 2$, there is a constant $C = C(\Xi, p)$ such that for any $x \in \mathbb{R}^d$, $h \in (0,1)$ and $f \in \widetilde{\mathbb{L}}^p(T)$,

$$\mathbb{E}\left(\sup_{t\in[0,T]}\left|\int_0^t (f(s, X^x_s) - f(s, X^x_{\pi_h(s)}))\mathrm{d}s\right|^2\right) \leqslant C h^{1-\delta} |\!|\!| f|\!|\!|^2_{\widetilde{\mathbb{L}}^p(T)}. \tag{3.28}$$

Proof. Let $\widetilde{Z}^x$ be a solution on a probability space $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, (\widetilde{\mathscr{F}}_t)_{t\geqslant 0}, \widetilde{\mathbb{P}})$ to the following SDE

$$\widetilde{Z}^x_t = x + \int_0^t \sigma(\widetilde{Z}^x_r)\mathrm{d}\widetilde{W}_r,$$

where $\widetilde{W}_t$ is a standard d-dimensional Brownian motion. Since $p_0 > 2 \vee d$, $B^2 \in \widetilde{\mathbb{L}}^{p_0/2}(T)$. By (3.15), one sees that for any $\gamma > 0$

$$\sup_x \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}} \exp\left(\gamma \int_0^T |\sigma^{-1}B(t, \widetilde{Z}^x_t)|^2\mathrm{d}t\right) < \infty.$$

Hence, by Novikov's criterion,

$$\mathbb{Z}_T := \exp\left(-\int_0^T \sigma^{-1}B(t, \widetilde{Z}^x_t)\mathrm{d}\widetilde{W}_t - \frac{1}{2}\int_0^T |\sigma^{-1}B(t, \widetilde{Z}^x_t)|^2\mathrm{d}t\right)$$

is integrable and for any $q>0$

$$\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|\mathbb{Z}_T|^q \leqslant C(\Xi,q). \tag{3.29}$$

Define $\mathrm{d}\mathbb{Q}:=\mathbb{Z}_T\mathrm{d}\widetilde{\mathbb{P}}$. Then, by Girsanov's theorem,

$$\bar{W}_t := \widetilde{W}_t - \int_0^t \sigma^{-1}B(s,\widetilde{Z}^x_s)\mathrm{d}s \text{ is a } \mathbb{Q}\text{-martingale.}$$

In other words,

$$\widetilde{Z}^x_t = x + \int_0^t B(s,\widetilde{Z}_s)\mathrm{d}s + \int_0^t \sigma(\widetilde{Z}_s)\mathrm{d}\bar{W}_s, \quad \mathbb{Q}-a.e.$$

Therefore, by the uniqueness of (3.24), we have

$$\mathbb{Q}\circ(\widetilde{Z}^x_\cdot)^{-1} = \mathbb{P}\circ(X^x_\cdot)^{-1}. \tag{3.30}$$

Now, we show (i)-(iv) one by one.

(i): In view of (3.30), one sees that

$$\mathbb{E}f(X^x_t) = \widetilde{\mathbb{E}}_{\mathbb{Q}} f(\widetilde{Z}^x_t) = \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\Big[\mathbb{Z}_T f(\widetilde{Z}_t)\Big].$$

By Hölder's inequality, (3.29) and (3.13), we have for any $r\in(1,p)$ and $1/r'+1/r=1$,

$$\mathbb{E}f(X^x_t) \leqslant \Big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|\mathbb{Z}_T|^{r'}\Big)^{1/r'}\Big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|f(\widetilde{Z}_t)|^r\Big)^{1/r} \lesssim \Big(t^{-dr/(2p)}\||f|^r\|_{p/r}\Big)^{1/r} \lesssim t^{-d/(2p)}\|f\|_p,$$

which is (3.25).

(ii): Similarly to (i), by Hölder's inequality and (3.14), we have

$$\begin{aligned}
\mathbb{E}\int_s^t f(u,X^x_u)\mathrm{d}u &\leqslant \Big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|\mathbb{Z}_T|^{r'}\Big)^{1/r'}\left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\left[\int_s^t f(u,\widetilde{Z}^x_u)\mathrm{d}u\right]^r\right)^{1/r}\\
&\lesssim \left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\left((t-s)^{r-1}\int_s^t |f(u,\widetilde{Z}^x_u)|^r\mathrm{d}u\right)\right)^{1/r}\\
&\lesssim (t-s)^{1-1/q-d/(2p)}\left(\||f|^r\|_{\widetilde{\mathbb{L}}^{p/r}_{q/r}}\right)^{1/r} \lesssim (t-s)^{1-1/q-d/(2p)}\|f\|_{\widetilde{\mathbb{L}}^p_q}.
\end{aligned}$$

The term $\mathbb{E}\int_s^t f(u,X^x_{\pi_h(u)})\mathrm{d}u$ can be estimated the same way.

(iii): For (3.27), it again follows from Hölder's inequality that

$$\begin{aligned}
\mathbb{E}\exp\left(\int_0^T f(t,X^x_t)\mathrm{d}t\right) &= \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\left[\mathbb{Z}_T\exp\big(\int_0^T f(t,\widetilde{Z}^x_t)\mathrm{d}t\big)\right]\\
&\leqslant \big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|\mathbb{Z}_T|^{r'}\big)^{1/r'}\left(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\exp\big(r\int_0^T f(t,\widetilde{Z}^x_t)\mathrm{d}t\big)\right)^{1/r} < \infty
\end{aligned}$$

by (3.29) and (3.15).

(iv): Let

$$A^f(h,X) := \sup_{t\in[0,T]}\Big|\int_0^t (f(s,X^x_s)-f(s,X^x_{\pi_h(s)}))\mathrm{d}u\Big|^2$$

and

$$A^f(h,\widetilde{Z}) := \sup_{t\in[0,T]}\Big|\int_0^t (f(s,\widetilde{Z}^x_s)-f(s,\widetilde{Z}^x_{\pi_h(s)}))\mathrm{d}u\Big|^2.$$

For any $\delta\in(0,1)$, we note that

$$\mathbb{E}A^f(h,X) = \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\big(\mathbb{Z}_T A^f(h,\widetilde{Z})\big)$$

$$= \widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\big(\mathbb{Z}_T|A^f(h,\widetilde{Z})|^\delta|A^f(h,\widetilde{Z})|^{1-\delta}\big).$$

Based on Hölder's inequality, (3.29), (3.26) and (3.21), for $1/r'+1/r=1$ with some $r\in(1,p/2)$, we have

$$\begin{aligned}\mathbb{E}A^f(h,X) &\leqslant \Big[\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}\big(|\mathbb{Z}_T|^{1/\delta}A^f(h,\widetilde{Z})\big)\Big]^\delta\Big[\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}A^f(h,\widetilde{Z})\Big]^{1-\delta}\\ &\leqslant \big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|\mathbb{Z}_T|^{r'/\delta}\big)^{\delta/r'}\big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}|A^f(h,\widetilde{Z})|^r\big)^{\delta/r}\big(\widetilde{\mathbb{E}}_{\widetilde{\mathbb{P}}}A^f(h,\widetilde{Z})\big)^{1-\delta}\\ &\lesssim \||f|^{2r}\|^{\delta/r}_{\widetilde{\mathbb{L}}^{p/(2r)}(T)}h^{(1-\delta_0)}\|f\|^{2(1-\delta)}_{\widetilde{\mathbb{L}}^p(T)}\\ &\lesssim h^{1-\delta_0}\|f\|^2_{\widetilde{\mathbb{L}}^p(T)},\end{aligned}$$

where $\delta_0=\delta_0(\delta)\to 0$ as $\delta\to 0$, and complete the proof.

□

Next, we want to prove an estimate for $\sup_x\|\mu_s^x-\mu_t^x\|_{var}$. To this end, we will use the relation between the PDE and the SDE. For any $T>0$, consider the following backward PDE:

$$\partial_t u^T + B\cdot\nabla u^T + a_{ij}\partial_i\partial_j u^T = 0,\quad u^T(T)=\varphi, \tag{3.31}$$

where $\varphi\in C_b^\infty$. By Proposition 2.7, there exists a unique solution u^T to (3.31) in the sense of Definition 2.3. Set

$$P^X_{s,t}f(x) := \mathbb{E}f(X^x_{s,t}),\quad P^X_t := P^X_{0,t}.$$

By (3.25), the domain of $P^X_{s,t}$ includes $\widetilde{L}^p$ for any $p\in(1,\infty]$. Then we have the following probabilistic representation.

Proposition 3.11. *Let $T>0$, $\varphi\in C_b^\infty(\mathbb{R}^d)$, u^T and $X^x_{s,t}$ be the solutions to (3.31) and (3.24) respectively. Then,*

$$u^T(s,x) = \mathbb{E}\varphi(X^x_{s,T}) = P^X_{s,T}\varphi(x). \tag{3.32}$$

Proof. It is straightforward to obtain (3.32) by applying the generalized Itô formula (3.4) to the function $t\mapsto u^T(t,X^x_{s,t})$ and taking expectation. □

Apart from the probabilistic representation, by the generalized Itô formula, we have the following Duhamel formula.

Lemma 3.12 (Duhamel formula)**.** *For any $\varphi\in C_b^\infty(\mathbb{R}^d)$,*

$$P^X_{s,t}\varphi(x) = P^\sigma_{t-s}\varphi(x) + \int_s^t P^X_{s,r}\big(B(r)\cdot\nabla P^\sigma_{t-r}\varphi\big)(x)\mathrm{d}r. \tag{3.33}$$

Proof. For any $t\in[0,T]$, let $v^t=v^t(t,x)$ be the solution to the following backward PDE:

$$\partial_r v^t + a_{ij}\partial_i\partial_j v^t = 0,\quad v^t(t)=\varphi.$$

Based on (3.32), one sees that $v^t(r)=P^\sigma_{t-r}\varphi$. By the generalized Itô formula (3.4), we have

$$\mathbb{E}v^t(t,X^x_{s,t}) = v^t(s,x) + \mathbb{E}\int_s^t(\partial_r v^t + a_{ij}\partial_i\partial_j v^t + B\cdot\nabla v^t)(r,X^x_{s,r})\mathrm{d}r,$$

which implies that

$$\begin{aligned}P^X_{s,t}\varphi(x) &= P^\sigma_{t-s}\varphi + \mathbb{E}\int_s^t(B(r)\cdot\nabla P^\sigma_{t-r}\varphi)(X^x_{s,r})\mathrm{d}r\\ &= P^\sigma_{t-s}\varphi + \int_s^t P^X_{s,r}(B(r)\cdot\nabla P^\sigma_{t-r}\varphi)(x)\mathrm{d}r,\end{aligned}$$

and we complete the proof. □

Now we can prove the estimate for $\|\mu_t^x-\mu_s^x\|_{var}$. We note that similar results have been proved in [50, Lemma 3.8 (1)(ii)] based on heat kernel estimates. It should be mentioned that the order of time regularity in [50] depends on the Hölder index β of σ and thus this is not applicable to our case.

Lemma 3.13. *Assume (3.1). For any $T>0$ and $q\in[p_0,\infty)$, there is a constant $C=C(\Xi,q)$ such that for any $0<s\leqslant t\leqslant T$ and $\varphi\in C_b^\infty$,*

$$\|P_t^X\varphi-P_s^X\varphi\|_\infty\leqslant C\Big[[(t-s)^{\frac12-\frac{d}{2q}}s^{-\frac12+\frac{d}{2q}}]\wedge 1\Big]s^{-\frac{d}{2q}}|\!|\!|\varphi|\!|\!|_q. \tag{3.34}$$

In particular, when $q=\infty$, for any $\delta>0$, there is a constant $C=C(\Xi,\delta)$ such that for all $x\in\mathbb{R}^d$,

$$\|\mu_t^x-\mu_s^x\|_{var}\leqslant C\Big[[(t-s)^{1/2-\delta}s^{-1/2}]\wedge 1\Big]. \tag{3.35}$$

Proof. For simplicity, let

$$\alpha:=\frac12-\frac{d}{2q}.$$

From (3.33), one sees that

$$\begin{aligned}P_t^X\varphi-P_s^X\varphi=&\big(P_t^\sigma\varphi-P_s^\sigma\varphi\big)+\int_s^t P_r^X\big(B(r)\cdot\nabla P_{t-r}^\sigma\varphi\big)\mathrm{d}r\\&+\int_0^s P_r^X\Big[B(r)\cdot\nabla\big(P_{t-r}^\sigma-P_{s-r}^\sigma\big)\varphi\Big]\mathrm{d}r\\:=&\mathscr{I}_1+\mathscr{I}_2+\mathscr{I}_3.\end{aligned}$$

Based on (3.16), we have

$$\begin{aligned}\|\mathscr{I}_1\|_\infty&\lesssim\Big[[(t-s)s^{-1}]\wedge 1\Big]s^{-\frac{d}{2q}}|\!|\!|\varphi|\!|\!|_q\\&\lesssim\Big[(t-s)^\alpha s^{-\alpha}\Big]s^{-\frac{d}{2q}}|\!|\!|\varphi|\!|\!|_q.\end{aligned}$$

By (3.25) and (3.13), we have

$$\begin{aligned}\|\mathscr{I}_2\|_\infty&\lesssim\int_s^t r^{-d/(2p_0)}|\!|\!|B(r)\cdot\nabla P_{t-r}^\sigma\varphi|\!|\!|_{p_0}\mathrm{d}r\\&\lesssim\int_s^t r^{-d/(2p_0)}\|\nabla P_{t-r}^\sigma\varphi\|_\infty\mathrm{d}r\\&\lesssim s^{-d/(2q)}\int_s^t r^{-d/(2p_0)+d/(2q)}(t-r)^{-1/2-d/(2q)}\mathrm{d}r|\!|\!|\varphi|\!|\!|_q\\&:=s^{-d/(2q)}K(t,s)|\!|\!|\varphi|\!|\!|_q,\end{aligned}$$

where

$$\begin{aligned}K(t,s)&\leqslant\left(s^{-\frac{d}{2p_0}+\frac{d}{2q}}\int_s^t(t-r)^{\alpha-1}\mathrm{d}r\right)\wedge\left((t-s)^{\frac12-\frac{d}{2p_0}}\int_0^1 r^{-\frac{d}{2p_0}+\frac{d}{2q}}(1-r)^{-\frac12-\frac{d}{2q}}\mathrm{d}r\right)\\&\lesssim(s^{-\alpha}(t-s)^\alpha)\wedge 1\end{aligned}$$

since $q\geqslant p_0>d$.

It remains to estimate $\mathscr{I}_3$. By (3.25) and (3.16), we have

$$\begin{aligned}\|\mathscr{I}_3\|_\infty&\lesssim\int_0^s r^{-d/(2p_0)}|\!|\!|B(r)\cdot\nabla\big(P_{t-r}^\sigma-P_{s-r}^\sigma\big)\varphi|\!|\!|_{p_0}\mathrm{d}r\\&\lesssim\int_0^s r^{-d/(2p_0)}\|\nabla\big(P_{t-r}^\sigma-P_{s-r}^\sigma\big)\varphi\|_\infty\mathrm{d}r\\&\lesssim\int_0^s r^{-d/(2p_0)}\Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r|\!|\!|\varphi|\!|\!|_q.\end{aligned}$$

We note that

$$\begin{aligned}\mathscr{J} &:= \int_0^s r^{-d/(2p_0)}\Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r\\ &\leqslant \int_0^s r^{-d/(2p_0)}(s-r)^{-1/2-d/(2q)}\mathrm{d}r \lesssim s^{1/2-d/(2p_0)-d/(2q)} \lesssim s^{-d/(2q)},\end{aligned} \tag{3.36}$$

since $p_0>d$. In addition, when $r\in(0,\frac{s}{2}]$, one sees that

$$\Big[(t-s)^{\frac12}(s-r)^{-\frac12}\Big]\wedge 1 \leqslant (t-s)^{\frac12-\frac{d}{2q}}(s-r)^{-\frac12+\frac{d}{2q}} \leqslant (t-s)^{\frac12-\frac{d}{2q}}\big(\frac{s}{2}\big)^{-\frac12+\frac{d}{2q}}.$$

Hence,

$$\begin{aligned}\mathscr{J} \lesssim & s^{-1/2-d/(2q)}\int_0^{s/2} r^{-d/(2p_0)}\Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big]\mathrm{d}r\\ &+ s^{-d/(2p_0)}\int_{s/2}^s \Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r\\ \lesssim & s^{-1}(t-s)^{1/2-d/(2q)}\int_0^{s/2} r^{-d/(2p_0)}\mathrm{d}r\\ &+ s^{-d/(2p_0)}\int_{s/2}^s \Big[[(t-s)^{\frac12}(s-r)^{-\frac12}]\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r.\end{aligned}$$

By a change of variables, we have

$$\begin{aligned}\mathscr{J} \lesssim & s^{-d/(2p_0)}(t-s)^{1/2-d/(2q)}\\ &+ s^{-d/(2p_0)}(t-s)^{1-1/2-d/(2q)}\int_0^{s/(t-s)} \Big[(s-r)^{-\frac12}\wedge 1\Big](s-r)^{-1/2-d/(2q)}\mathrm{d}r\\ \lesssim & s^{-d/(2p_0)}(t-s)^{1-1/2-d/(2q)} \lesssim s^{-\frac12}(t-s)^{\frac12-\frac{d}{2q}}\end{aligned}$$

since $q<\infty$ and $p_0>d$, which combined with (3.36) implies that

$$\mathscr{J} \lesssim \big(s^{-\frac12}(t-s)^{\alpha}\big)\wedge s^{-\frac{d}{2q}} = \Big([(t-s)^{\alpha}s^{-\alpha}]\wedge 1\Big)s^{-\frac{d}{2q}}.$$

Thus, we obtain (3.34). In particular, noting that

$$|\!|\!|\varphi|\!|\!|_q \lesssim \|\varphi\|_\infty, \quad \forall q<\infty,$$

by Lusin's theorem and (3.34), we have for all $\delta>0$ and $q\in[p_0\vee(d/(2\delta)),\infty)$

$$\begin{aligned}\|\mu_t^x-\mu_s^x\|_{var} &= \sup_{\varphi\in C_b^\infty}\frac{|P_t^X\varphi(x)-P_s^X\varphi(x)|}{\|\varphi\|_\infty} \lesssim [(t-s)^{\frac12-\frac{d}{2q}}s^{-\frac12+\frac{d}{2q}}]s^{-\frac{d}{2q}}\\ &\lesssim (t-s)^{\frac12-\delta}s^{-\frac12}.\end{aligned}$$

Moreover, it is easy to see that

$$\|\mu_t^x-\mu_s^x\|_{var}\leqslant 2,$$

which completes the proof. □

The following lemma is the distribution dependent version of (3.28).

Lemma 3.14. *For any $T>0$, $p\in(d\vee 2,\infty)$, assume that $f:\mathbb{R}_+\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}$ such that*

$$\kappa_f := \sup_{t\in[0,T]}\sup_{\mu,\nu}\Big(|\!|\!|f(t,\cdot,\mu)|\!|\!|_p + \frac{|\!|\!|f(t,\cdot,\mu)-f(t,\cdot,\nu)|\!|\!|_p}{\|\mu-\nu\|_{var}}\Big)<\infty.$$

Then, for any $\delta>0$, there is a constant $C=C(\Xi,p,\delta)$ such that for any $x\in\mathbb{R}^d$ and $h\in(0,1)$

$$\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t (f(s,X_s^x,\mu_s^x)-f(s,X_{\pi_h(s)}^x,\mu_{\pi_h(s)}^x))\mathrm{d}s\Big|^2\Big)\leqslant C(\kappa_f)^2h^{1-\delta}. \tag{3.37}$$

Proof. For simplicity, we drop the superscript x from X^x and μ^x. First of all, we note that

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t (f(s,X_s,\mu_s)-f(s,X_{\pi_h(s)},\mu_{\pi_h(s)}))\mathrm{d}s\Big|^2\Big)\\
\lesssim\;&\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t (f(s,X_s,\mu_s)-f(s,X_{\pi_h(s)},\mu_s))\mathrm{d}s\Big|^2\Big)\\
&+\mathbb{E}\Big(\sup_{t\in[0,T]}\Big|\int_0^t (f(s,X_{\pi_h(s)},\mu_s)-f(s,X_{\pi_h(s)},\mu_{\pi_h(s)}))\mathrm{d}s\Big|^2\Big)\\
:=\;&\mathscr{I}_1^h+\mathscr{I}_2^h.
\end{aligned}$$

By (3.28), for any $\delta>0$, we have

$$\mathscr{I}_1^h\lesssim h^{1-\delta}\sup_{t\in[0,T]}|\!|\!|f(t,\cdot,\mu_t)|\!|\!|_p^2\lesssim h^{1-\delta}(\kappa_f)^2.$$

For $\mathscr{I}_2^h$, we use the same method as in Step 2 of the proof of Lemma 3.8. For any $0\leqslant s<t\leqslant T$, set

$$\mathscr{J}_{s,t}^h:=\int_{s+h}^{t+h}|f(r,X_{\pi_h(r)},\mu_r)-f(r,X_{\pi_h(r)},\mu_{\pi_h(r)})|\mathrm{d}r$$

and $\|\cdot\|:=\big(\mathbb{E}|\cdot|^2\big)^{1/2}$. Then

$$\mathscr{I}_2^h\leqslant\mathbb{E}\Big|\int_h^{T+h}|f(s,X_{\pi_h(s)},\mu_s)-f(s,X_{\pi_h(s)},\mu_{\pi_h(s)})|\mathrm{d}s\Big|^2=\|\mathscr{J}_{0,T}^h\|^2,$$

since $\pi_h(r)=r$, if $r<h$. Based on Hölder's inequality and (3.26), one sees that if $2/q+d/p<1$, then

$$\begin{aligned}
\mathbb{E}|\mathscr{J}_{s,t}^h|^2&\lesssim(t-s)\mathbb{E}\int_{s+h}^{t+h}\Big|f(r,X_{\pi_h(r)},\mu_r)-f(r,X_{\pi_h(r)},\mu_{\pi_h(r)})\Big|^2\mathrm{d}r\\
&\lesssim(t-s)\Big(\int_{s+h}^{t+h}|\!|\!|f(r,\cdot,\mu_r)-f(r,\cdot,\mu_{\pi_h(r)})|\!|\!|_p^q\mathrm{d}r\Big)^{2/q}\\
&\lesssim(\kappa_f)^2(t-s)\Big(\int_{s+h}^{t+h}\|\mu_r-\mu_{\pi_h(r)}\|_{var}^q\mathrm{d}r\Big)^{2/q}.
\end{aligned}$$

Then, by (3.35) and the fact that $q>2$, for any $\delta>0$ we have

$$\begin{aligned}
\|\mathscr{J}_{s,t}^h\|&\lesssim\kappa_f(t-s)^{1/2}\Big(\int_{s+h}^{t+h}h^{(\frac12-\delta)q}(\pi_h(r))^{-q/2}\mathrm{d}r\Big)^{1/q}\\
&\lesssim\kappa_f(t-s)^{1/2}h^{1/2-\delta}\Big(\int_s^t r^{-q/2}\mathrm{d}r\Big)^{1/q}\lesssim\kappa_f h^{1/2-\delta}(t-s)^{1/2+1/q}s^{-1/2}.
\end{aligned}$$

Taking $t_n:=2^{-n}T$, we have

$$\begin{aligned}
\|\mathscr{J}_{0,T}^h\|&\leqslant\sum_{n=0}^\infty\|\mathscr{J}_{t_{n+1},t_n}^h\|\lesssim\kappa_f h^{1/2-\delta}\sum_{n=0}^\infty(t_n-t_{n+1})^{1/2+1/q}(t_{n+1})^{-1/2}\\
&\lesssim\kappa_f h^{1/2-\delta}T^{1/q}\sum_{n=0}^\infty 2^{-\frac{n+1}{q}}\lesssim\kappa_f h^{1/2-\delta},
\end{aligned}$$

which implies $\mathscr{I}_2^h\lesssim(\kappa_f)^2h^{1-2\delta}$ and we complete the proof. $\square$

Now, we have the following strong fluctuation result, which is crucial in this paper.

Lemma 3.15. *Let $T>0$ and $g:\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}$ be a bounded function satisfying*

$$c_g:=\sup_{\substack{t\neq s\in[0,T]\\ x\in\mathbb{R}^d}}\Big(|g(t,x)|+\frac{|g(t,x)-g(s,x)|}{|t-s|^\alpha}\Big)<\infty \tag{3.38}$$

for some $\alpha>0$. Assume (3.1), $(\mathbf{H}_b^1)$ and $(\mathbf{H}_b^2)$ hold. Then for any $\delta>0$, there is a constant $C=C(\Xi,\kappa_0,\delta,c_g)$ such that for any $x\in\mathbb{R}^d$ and $\varepsilon>0$,

$$\begin{aligned}&\mathbb{E}\left(\sup_{t\in[0,T]}\Big|\int_0^t g(s,X_s^x)(b(\frac{s}{\varepsilon},X_s^x,\mu_s^x)-\bar b(X_s^x,\mu_s^x))\mathrm{d}s\Big|^2\right)\\ &\leqslant C\inf_{h>0}\left(h^{1-\delta}+h^{2\alpha}+\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2\right).\end{aligned}\tag{3.39}$$

Proof. For simplicity, we drop the superscript x from X^x and μ^x. Set $b_\varepsilon(t):=b(t/\varepsilon)$ and

$$\mathbf{X}_\cdot:=(X_\cdot,\mu_\cdot).$$

For any $h\in(0,1)$,

$$\begin{aligned}&\mathbb{E}\left(\sup_{t\in[0,T]}\Big|\int_0^t g(s,X_s^x)(b_\varepsilon(s,X_s^x,\mu_s^x)-\bar b(X_s^x,\mu_s^x))\mathrm{d}s\Big|^2\right)\\ &\lesssim\mathbb{E}\left(\sup_{t\in[0,T]}\Big|\int_0^t((gb_\varepsilon)(s,\mathbf{X}_s)-(gb_\varepsilon)(s,\mathbf{X}_{\pi_h(s)}))\mathrm{d}s\Big|^2\right)\\ &\quad+\mathbb{E}\left(\sup_{t\in[0,T]}\Big|\int_0^t g(s,X_{\pi_h(s)})(b_\varepsilon(s,\mathbf{X}_{\pi_h(s)})-\bar b(\mathbf{X}_{\pi_h(s)}))\mathrm{d}s\Big|^2\right)\\ &\quad+\mathbb{E}\left(\sup_{t\in[0,T]}\Big|\int_0^t((g\bar b)(s,\mathbf{X}_s)-(g\bar b)(s,\mathbf{X}_{\pi_h(s)}))\mathrm{d}s\Big|^2\right)\\ &:=\mathscr{I}_1^{\varepsilon,h}+\mathscr{I}_2^{\varepsilon,h}+\mathscr{I}_3^{\varepsilon,h}.\end{aligned}$$

By $(\mathbf{H}_b^1)$ and (3.37), for any $\delta>0$, we have

$$\mathscr{I}_1^{\varepsilon,h}+\mathscr{I}_3^{\varepsilon,h}\lesssim\|g\|_\infty^2\kappa_0^2h^{1-\delta}.$$

For $\mathscr{I}_2^{\varepsilon,h}$, we note that by (3.38) and (3.26) with $p_0>(d/2)\vee 1$,

$$\begin{aligned}\mathscr{I}_2^{\varepsilon,h}&\lesssim\mathbb{E}\left(\sup_{t\in[0,T]}\Big|\int_0^t g(\pi_h(s),X_{\pi_h(s)})(b_\varepsilon(s,\mathbf{X}_{\pi_h(s)})-\bar b(\mathbf{X}_{\pi_h(s)}))\mathrm{d}s\Big|^2\right)\\ &\quad+\mathbb{E}\left(\sup_{t\in[0,T]}\int_0^t|s-\pi_h(s)|^{2\alpha}\Big(|b_\varepsilon(s,\mathbf{X}_{\pi_h(s)})|^2+|\bar b(\mathbf{X}_{\pi_h(s)}))|^2\Big)\mathrm{d}s\right)\\ &\lesssim\mathscr{I}_{21}^{\varepsilon,h}+h^{2\alpha},\end{aligned}$$

where

$$\mathscr{I}_{21}^{\varepsilon,h}:=\mathbb{E}\left(\sup_{t\in[0,T]}\Big|\int_0^t g(\pi_h(s),X_{\pi_h(s)})(b_\varepsilon(s,\mathbf{X}_{\pi_h(s)})-\bar b(\mathbf{X}_{\pi_h(s)}))\mathrm{d}s\Big|^2\right).$$

It suffices to show

$$\mathscr{I}_{21}^{\varepsilon,h}\lesssim h+\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2.$$

Letting $M(t)=[t/h]$ and noting that $\pi_h(s)=s$ for $s\in[0,h)$, we have

$$\begin{aligned}\mathscr{I}_{21}^{\varepsilon,h}\lesssim&\mathbb{E}\left(\sup_{t\in[0,h)}\Big|\int_0^t g(s,X_s)(b_\varepsilon(s,\mathbf{X}_s)-\bar b(\mathbf{X}_s))\mathrm{d}s\Big|^2\right)\\&+\mathbb{E}\left(\sup_{t\in[h,T]}\Big|\int_0^h g(s,X_s)(b_\varepsilon(s,\mathbf{X}_s)-\bar b(\mathbf{X}_s))\mathrm{d}s\Big|^2\right)\\&+\mathbb{E}\left(\sup_{t\in[h,T]}\Big|\int_{M(t)h}^t g(\pi_h(s),X_{\pi_h(s)})(b_\varepsilon(s,\mathbf{X}_{\pi_h(s)})-\bar b(\mathbf{X}_{\pi_h(s)})\mathrm{d}s\Big|^2\right)\\&+\mathbb{E}\left(\sup_{t\in[h,T]}\Big|\int_h^{M(t)h} g(\pi_h(s),X_{\pi_h(s)})(b_\varepsilon(s,\mathbf{X}_{\pi_h(s)})-\bar b(\mathbf{X}_{\pi_h(s)}))\mathrm{d}s\Big|^2\right)\\:=&\mathscr{I}_{211}^{\varepsilon,h}+\mathscr{I}_{212}^{\varepsilon,h}+\mathscr{I}_{213}^{\varepsilon,h}+\mathscr{I}_{214}^{\varepsilon,h}.\end{aligned}$$

It follows from Hölder's inequality that

$$\begin{aligned}\sum_{i=1}^3\mathscr{I}_{21i}^{\varepsilon,h}&\lesssim h\|g\|_\infty^2\left(\mathbb{E}\left[\int_0^T|b_\varepsilon(s,\mathbf{X}_s)|^2+|\bar b(\mathbf{X}_s)|^2\mathrm{d}s\right]+\mathbb{E}\left[\int_0^T|b_\varepsilon(s,\mathbf{X}_{\pi_h(s)})|^2+|\bar b(\mathbf{X}_{\pi_h(s)})|^2\mathrm{d}s\right]\right)\\&\lesssim h\|g\|_\infty^2\kappa_0^2,\end{aligned}$$

because of (3.26). Thus, we only need to prove

$$\mathscr{I}_{124}^{\varepsilon,h}\lesssim\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2.$$

By the definition of π_h, it is easy to see that

$$\mathscr{I}_{214}^{\varepsilon,h}\leqslant\mathbb{E}\Big(\sup_{2\leqslant m\leqslant M(T)}\Big|\sum_{k=1}^{m-1}g(kh,X_{kh})\int_{kh}^{(k+1)h}(b_\varepsilon(s,\mathbf{X}_{kh})-\bar b(\mathbf{X}_{kh}))\mathrm{d}s\Big|^2\Big).$$

Based on the fact that $|\sum_{k=1}^{m-1}a_k|^2\leqslant(m-1)\sum_{k=1}^{m-1}|a_k|^2$, one sees that

$$\mathscr{I}_{214}^{\varepsilon,h}\leqslant M(T)c_g^2\sum_{k=1}^{M(T)-1}\mathbb{E}\Big|\int_{kh}^{(k+1)h}(b_\varepsilon(s,\mathbf{X}_{kh})-\bar b(\mathbf{X}_{kh}))\mathrm{d}s\Big|^2.$$

By a change of variables and (1.6), we have

$$\begin{aligned}\mathscr{I}_{214}^{\varepsilon,h}&\lesssim M(T)\sum_{k=1}^{M(T)-1}\mathbb{E}\left|\varepsilon\int_{kh/\varepsilon}^{(k+1)h/\varepsilon}\big(b(s,\mathbf{X}_{kh})-\bar b(\mathbf{X}_{kh})\big)\mathrm{d}s\right|^2\\&\lesssim[\frac{T}{h}]h^2\sum_{k=1}^{M(T)-1}\mathbb{E}\Big|\frac{\varepsilon}{h}\int_{kh/\varepsilon}^{(k+1)h/\varepsilon}(b(s,\mathbf{X}_{kh})-\bar b(\mathbf{X}_{kh}))\mathrm{d}s\Big|^2\\&\lesssim h\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2\sum_{k=1}^{M(T)-1}\mathbb{E}|H(\mathbf{X}_{kh})|^2.\end{aligned}$$

We note that

$$h\sum_{k=1}^{M(T)-1}\mathbb{E}|H(\mathbf{X}_{kh})|^2=\mathbb{E}\int_h^{M(T)h}|H(\mathbf{X}_{\pi_h(s)})|^2\mathrm{d}s\leqslant\mathbb{E}\int_h^T|H(\mathbf{X}_{\pi_h(s)})|^2\mathrm{d}s.$$

Again by (3.26), we have

$$\mathscr{I}_{214}^{\varepsilon,h}\lesssim\left(\omega\left(\frac{h}{\varepsilon}\right)\right)^2\sup_\mu|\!|\!|H(\cdot,\mu)|\!|\!|_{p_0}^2$$

and complete the proof.

□

3.3. **Time regularity for solutions to the parabolic equation.** In this section, we establish the time regularity for the solution of PDE (3.3). First, we give the following probabilistic representation of the solution to PDE if $B \equiv 0$.

Lemma 3.16. *Let $B \equiv 0$ and u be a solution to PDE (3.3). Then,*

$$u(t) = \int_0^t e^{-\lambda(t-s)} P^\sigma_{t-s} f(s) \mathrm{d}s + e^{-\lambda t} P^\sigma_t \varphi. \tag{3.40}$$

Proof. Applying the generalized Itô formula (3.4) to

$$s \to e^{-\lambda s} u(t-s, Z^x_s),$$

we get

$$\begin{aligned} e^{-\lambda t} u(0, Z^x_t) - u(t,x) = &\int_0^t e^{-\lambda s} (-\partial_s u + a_{ij}\partial_i \partial_j u - \lambda u)(t-s, Z^x_s)\mathrm{d}s \\ &+ \int_0^t e^{-\lambda s} \nabla u(t-s, Z^x_s) \mathrm{d}W_s. \end{aligned}$$

Taking expectation of both sides, we obtain that

$$e^{-\lambda t} P^\sigma_t \varphi - u(t) = - \int_0^t e^{-\lambda s} P^\sigma_s f(t-s) \mathrm{d}s,$$

which is (3.40) by a change of variable and this completes the proof. □

Using the above Lemma, we have the following time Hölder regularity of ∇u.

Lemma 3.17. *Assume $\varphi \equiv 0$. Under condition (3.1) with some $p_0 \in (d, \infty)$, for any $\lambda \geqslant 0$, there is a constant $C = C(\Xi, p, \lambda)$ such that for all $t, s \in [0,T]$ and $f \in \widetilde{\mathbb{L}}^{p_0}(T)$ the solution u to (3.3) in the sense of Definition 2.3 satisfies*

$$\|\nabla u(t) - \nabla u(s)\|_\infty \leqslant C|t-s|^{\frac{1}{2} - \frac{d}{2p_0}} \|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}. \tag{3.41}$$

Remark 3.18. By Remark 2.6 and (3.41), we further have that there is a version of the solution such that $u \in C([0,T]; \mathbf{C}^1)$.

Proof. First, since $B, f \in \widetilde{\mathbb{L}}^{p_0}(T) \subset \widetilde{\mathbb{L}}^{p_0}_q(T)$, $\forall q$, we indeed have a unique solution u. Set

$$g(s) := B \cdot \nabla u(s) + f(s).$$

In view of (2.5),

$$|\!|\!| g |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)} \leqslant |\!|\!| b |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)} \|\nabla u\|_{\mathbb{L}^\infty_T} + |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)} \lesssim |\!|\!| f |\!|\!|_{\widetilde{\mathbb{L}}^{p_0}(T)}.$$

Then, for any $0 \leqslant s < t \leqslant T$, by (3.40), one sees that

$$\begin{aligned} \|\nabla u(t) - \nabla u(s)\|_\infty = &\left\| \int_s^t e^{-\lambda(t-r)} \nabla P^\sigma_{t-r} g(r) \mathrm{d}r \right\|_\infty \\ &+ \left\| \int_0^s \left(e^{-\lambda(t-r)} - e^{-\lambda(s-r)}\right) \nabla P^\sigma_{t-r} g(r) \mathrm{d}r \right\|_\infty \\ &+ \left\| \int_0^s e^{-\lambda(s-r)} \left(\nabla P^\sigma_{t-r} - \nabla P^\sigma_{s-r}\right) g(r) \mathrm{d}r \right\|_\infty \\ =: &\ \mathscr{I}_1 + \mathscr{I}_2 + \mathscr{I}_3. \end{aligned}$$

By (3.13), one sees that

$$\mathscr{I}_1 \lesssim \int_s^t (t-r)^{-1/2 - d/(2p_0)} |\!|\!| g(r) |\!|\!|_{p_0} \mathrm{d}r$$

$$\lesssim \int_s^t (t-r)^{-1/2-d/(2p_0)}\mathrm{d}r \|g\|_{\widetilde{\mathbb{L}}^{p_0}(T)}$$
$$\lesssim |t-s|^{\frac{1}{2}-\frac{d}{2p_0}} \|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}.$$

For $\mathscr{I}_2$, noting that $|e^{-x}-e^{-y}| \leqslant |x-y|$ for any $x, y > 0$, it follows from (3.13) that

$$\mathscr{I}_2 \lesssim |t-s| \int_0^s (t-r)^{-1/2-d/(2p_0)} \|g(r)\|_{p_0}\mathrm{d}r$$
$$\lesssim |t-s| \|g\|_{\widetilde{\mathbb{L}}^{p_0}(T)} \lesssim |t-s| \|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}.$$

For $\mathscr{I}_3$, by (3.16) with $q=\infty$, we have

$$\mathscr{I}_3 \lesssim \int_0^s ([(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}] \wedge 1)(s-r)^{-1/2-d/(2p_0)}\mathrm{d}r \|g\|_{\widetilde{\mathbb{L}}^{p_0}(T)}.$$

By a change of variable, we have

$$\mathscr{I}_3 \lesssim |t-s|^{\frac{1}{2}-\frac{d}{2p_0}} \int_0^\infty ([(s-r)^{-\frac{1}{2}}] \wedge 1)(s-r)^{-1/2-d/(2p_0)}\mathrm{d}r \|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}$$
$$\lesssim |t-s|^{1/2-d/(2p_0)} \|f\|_{\widetilde{\mathbb{L}}^{p_0}(T)}$$

and complete the proof. □

Moreover, we also have an estimate of time regularity for the solution to the following Cauchy problem.

Lemma 3.19. *Assume* (3.1). *Let* $\varphi \in C_b^\infty$ *and let* u *be the unique solution to the following Cauchy problem on* $[0,T]$ *in the sense of Definition 2.3*

$$\partial_t u = a_{ij}\partial_i\partial_j u + B \cdot \nabla u, \quad u_0 = \varphi. \tag{3.42}$$

Then there is a constant $C = C(\Xi)$ *such that for all* $0 \leqslant s < t \leqslant T$,

$$\|\nabla u(t) - \nabla u(s)\|_\infty \leqslant C(t-s)^{\frac{1}{2}-\frac{d}{2p_0}} s^{-1+\frac{d}{2p_0}} \|\varphi\|_\infty. \tag{3.43}$$

Proof. First, by (3.40), we have

$$u(t) = \int_0^t P^\sigma_{t-s}(B\cdot\nabla u)(s)\mathrm{d}s + P^\sigma_t \varphi, \tag{3.44}$$

which implies that

$$\|\nabla u(t)\|_\infty \lesssim \int_0^t (t-s)^{-\frac{1}{2}-\frac{d}{2p_0}} \|B\cdot\nabla u(s)\|_{p_0}\mathrm{d}s + t^{-\frac{1}{2}}\|\varphi\|_\infty$$
$$\lesssim \int_0^t (t-s)^{-\frac{1}{2}-\frac{d}{2p_0}} \|\nabla u(s)\|_\infty\mathrm{d}s + t^{-\frac{1}{2}}\|\varphi\|_\infty$$

because of (3.13). Hence, by Gronwall's inequality, we have

$$\|\nabla u(t)\|_\infty \lesssim t^{-\frac{1}{2}}\|\varphi\|_\infty. \tag{3.45}$$

By (3.44), (3.13) and (3.16), one sees that

$$\|\nabla u(t) - \nabla u(s)\|_\infty \leqslant \int_s^t \|\nabla P^\sigma_{t-r}(B\cdot\nabla u)(r)\|_\infty\mathrm{d}r + \int_0^s \|\nabla(P^\sigma_{t-r} - P^\sigma_{s-r})B\cdot\nabla u(r)\|_\infty\mathrm{d}r$$
$$+ \|\nabla(P^\sigma_t - P^\sigma_s)\varphi\|_\infty.$$
$$\lesssim \int_s^t (t-r)^{-\frac{1}{2}-\frac{d}{2p_0}} \|\nabla u(r)\|_\infty\mathrm{d}r$$
$$+ \int_0^s \left\{\left[(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\right] \wedge 1\right\}(s-r)^{-\frac{1}{2}-\frac{d}{2p_0}} \|\nabla u(r)\|_\infty\mathrm{d}r$$

$$
\begin{aligned}
&+\left\{\left[(t-s)^{\frac{1}{2}}s^{-\frac{1}{2}}\right]\wedge 1\right\}s^{-\frac{1}{2}}\|\varphi\|_\infty\\
:=&\mathscr{I}_1+\mathscr{I}_2+\mathscr{I}_3.
\end{aligned}
$$

Then, noting that $x\wedge 1\leqslant x^\theta$ for all $x\geqslant 0$ and $\theta\in[0,1]$, we have

$$
\mathscr{I}_3\lesssim\left\{\left[(t-s)^{\frac{1}{2}}s^{-\frac{1}{2}}\right]\wedge 1\right\}s^{-\frac{1}{2}}\|\varphi\|_\infty\lesssim(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-1+\frac{d}{2p_0}}\|\varphi\|_\infty.
$$

Based on (3.45) and a change of variable,

$$
\begin{aligned}
\mathscr{I}_1&\lesssim\|\varphi\|_\infty\int_s^t(t-r)^{-\frac{1}{2}-\frac{d}{2p_0}}r^{-\frac{1}{2}}\mathrm{d}r\lesssim(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-\frac{1}{2}}\|\varphi\|_\infty\\
&\lesssim(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-1+\frac{d}{2p_0}}\|\varphi\|_\infty
\end{aligned}
$$

since $p_0>d$. For $\mathscr{I}_2$, again by (3.45), we divide $[0,s]$ into $[0,s/2)$ and $[s/2,s]$ and have

$$
\begin{aligned}
\frac{\mathscr{I}_2}{\|\varphi\|_\infty}&\lesssim(\int_0^{s/2}+\int_{s/2}^s)\left\{\left[(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\right]\wedge 1\right\}(s-r)^{-\frac{1}{2}-\frac{d}{2p_0}}r^{-\frac{1}{2}}\mathrm{d}r\\
&\lesssim(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-1}\int_0^{s/2}r^{-\frac{1}{2}}\mathrm{d}r+s^{-\frac{1}{2}}\int_{s/2}^s\left\{\left[(t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\right]\wedge 1\right\}(s-r)^{-\frac{1}{2}-\frac{d}{2p_0}}\mathrm{d}r\\
&\lesssim(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-\frac{1}{2}}+s^{-\frac{1}{2}}\int_0^s\left\{\left[(t-s)^{\frac{1}{2}}r^{-\frac{1}{2}}\right]\wedge 1\right\}r^{-\frac{1}{2}-\frac{d}{2p_0}}\mathrm{d}r,
\end{aligned}
$$

where we used the fact that

$$
\begin{aligned}
&\left[\left((t-s)^{\frac{1}{2}}(s-r)^{-\frac{1}{2}}\right)\wedge 1\right](s-r)^{-\frac{1}{2}-\frac{d}{2p_0}}\\
&\leqslant(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}(s-r)^{-1}\leqslant 2(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-1}\quad\forall r\in[0,s/2).
\end{aligned}
$$

From a change of variable, we have

$$
\begin{aligned}
\mathscr{I}_2&\lesssim\left[(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-\frac{1}{2}}+s^{-\frac{1}{2}}(t-s)^{1-\frac{1}{2}-\frac{d}{2p_0}}\int_0^\infty(r^{-\frac{1}{2}}\wedge 1)r^{-\frac{1}{2}-\frac{d}{2p_0}}\mathrm{d}r\right]\|\varphi\|_\infty\\
&\lesssim(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-\frac{1}{2}}\|\varphi\|_\infty\lesssim(t-s)^{\frac{1}{2}-\frac{d}{2p_0}}s^{-1+\frac{d}{2p_0}}\|\varphi\|_\infty
\end{aligned}
$$

since $p_0\in(d,\infty)$ and complete the proof.

□

4. Convergence rate of the total variation

In this section, under $(\mathbf{H}_\sigma)$, $(\mathbf{H}_b^1)$ and $(\mathbf{H}_b^2)$, we will derive the convergence rate of $\|\mu^\varepsilon-\mu\|_{var}$. Recall that on the probability space $(\Omega,\mathscr{F},\mathbb{P},(\mathscr{F}_s)_{s\geqslant 0})$ we have a unique strong solution $(X^\varepsilon_\cdot,X_\cdot)$ to the following systems

$$
\mathrm{d}X^\varepsilon_t=b(t/\varepsilon,X^\varepsilon_t,\mu^\varepsilon_t)\mathrm{d}t+\sigma(X^\varepsilon_t)\mathrm{d}W_t,\quad X^\varepsilon_0=\xi,\tag{4.1}
$$

and

$$
\mathrm{d}X_t=\bar{b}(X_t,\mu_t)\mathrm{d}t+\sigma(X_t)\mathrm{d}W_t,\quad X_0=\xi,\tag{4.2}
$$

where μ^ε_t and μ_t are the distributions of X^ε_t and X_t respectively. For simplicity, in the sequel, let

$$
b_\varepsilon(t):=b(t/\varepsilon)\ (\varepsilon>0),\quad\text{and}\quad b_0:=\bar{b}.
$$

For any $x\in\mathbb{R}^d$ and $t\geqslant s\geqslant 0$, let $(Y^\varepsilon_{s,t}(x),Y_{s,t}(x))$ be the unique strong solution to the following SDEs

$$
\mathrm{d}Y^\varepsilon_{s,t}(x)=b_\varepsilon(t,Y^\varepsilon_{s,t}(x),\mu^\varepsilon_t)\mathrm{d}t+\sigma(Y^\varepsilon_{s,t}(x))\mathrm{d}W_t,\quad Y^\varepsilon_{s,s}(x)=x
$$

and

$$
\mathrm{d}Y_{s,t}(x)=\bar{b}(Y_{s,t}(x),\mu_t)\mathrm{d}t+\sigma(Y_{s,t}(x))\mathrm{d}W_t,\quad Y_{s,s}(x)=x.
$$

Set $Y_t^\varepsilon(x) := Y_{0,t}^\varepsilon(x)$ and $Y_t(x) := Y_{0,t}(x)$ for all $t \geqslant 0$ and $x \in \mathbb{R}^d$. Let $P^{x,\varepsilon}$ and P^x denote the distributions of $Y_\cdot^\varepsilon(x)$ and $Y_\cdot(x)$ in $C([0,T];\mathbb{R}^d)$ respectively. Based on the strong uniqueness of the above SDEs, we have

$$\int_{\mathbb{R}^d} P^{x,\varepsilon}\ \mathbb{P}\circ\xi^{-1}(\mathrm{d}x) = \mathbb{P}\circ(X_\cdot^\varepsilon)^{-1} \quad \text{and} \quad \int_{\mathbb{R}^d} P^x\ \mathbb{P}\circ\xi^{-1}(\mathrm{d}x) = \mathbb{P}\circ(X_\cdot)^{-1}. \tag{4.3}$$

Therefore, the estimates in Section 3.2 hold for X^ε and X, where the constants are independent of ε, since

$$\sup_{\varepsilon\geqslant 0}\sup_{\mu\in\mathcal{P}(\mathbb{R}^d)} |\!|\!| b_\varepsilon(\cdot,\mu)|\!|\!|_{\widetilde{\mathbb{L}}_T^{p_0}} < \infty.$$

Moreover, for any $t \in \mathbb{R}_+$ and $\varphi \in C_b^\infty$, consider the following Kolmogorov backward equation

$$\partial_s u^t + a_{ij}\partial_i\partial_j u^t + b_0(\cdot,\mu_s)\cdot\nabla u^t = 0, \tag{4.4}$$

with final condition

$$u^t(t) = \varphi.$$

By Proposition 2.7 and 3.11, there exists a unique solution u^t to (4.4), which is given by

$$u^t(s,x) = \mathbb{E}\varphi(Y_{s,t}(x)).$$

Define $\tilde{u}(s,x) = u^t(t-s,x)$. Then $\tilde{u}$ is the solution to

$$\partial_s\tilde{u} = a_{ij}\partial_i\partial_j\tilde{u} + b_0(\cdot,\mu_{t-s})\cdot\nabla\tilde{u}, \quad \tilde{u}_0 = \varphi.$$

By Lemma 3.19, we have

$$\|\nabla u^t(s)\|_\infty \lesssim (t-s)^{-\frac{1}{2}}\|\varphi\|_\infty \tag{4.5}$$

and

$$\|\nabla u^t(s_1) - \nabla u^t(s_2)\|_\infty \lesssim |s_1-s_2|^{\frac{1}{2}-\frac{d}{2p_0}}(s_1\wedge s_2)^{-1+\frac{d}{2p_0}}\|\varphi\|_\infty. \tag{4.6}$$

Then, for any $t\in[0,T]$, by applying the generalized Itô formula (3.4) to $u^t(s,Y_s^\varepsilon(x))$, one sees that

$$\mathbb{E}\varphi(Y_t^\varepsilon(x)) - u^t(0,x) = \mathbb{E}\int_0^t \Big(b_\varepsilon(s,Y_s^\varepsilon(x),\mu_s^\varepsilon) - b_0(Y_s^\varepsilon(x),\mu_s)\Big)\cdot\nabla u^t(s,Y_s^\varepsilon(x))\mathrm{d}s. \tag{4.7}$$

Noting that $u^t(0,x) = \mathbb{E}\varphi(Y_t(x))$, by (4.3), we have

$$\begin{aligned}|\mathbb{E}\varphi(X_t^\varepsilon) - \mathbb{E}\varphi(X_t)| &= \Big|\int_{\mathbb{R}^d}\big(\mathbb{E}\varphi(Y_t^\varepsilon(x)) - \mathbb{E}\varphi(Y_t(x))\big)\mathbb{P}\circ\xi^{-1}(\mathrm{d}x)\Big| \\ &\leqslant \sup_x\Big|\mathbb{E}\int_0^t\Big(b_\varepsilon(s,Y_s^\varepsilon(x),\mu_s^\varepsilon) - b_0(Y_s^\varepsilon(x),\mu_s)\Big)\cdot\nabla u^t(s,Y_s^\varepsilon(x))\mathrm{d}s\Big|.\end{aligned} \tag{4.8}$$

Here is the main result of this section:

Theorem 4.1. *Under the conditions* $(\mathbf{H}_\sigma)$ *and* $(\mathbf{H}_b^1)$*–*$(\mathbf{H}_b^2)$*, for any* $T>0$*, there is a constant* $C = C(\kappa_0,\kappa_1,d,T,p_0,\beta) > 0$ *such that for all* $\varepsilon>0$ *and* $t\in[0,T]$*,*

$$\|\mu_t^\varepsilon - \mu_t\|_{var} \leqslant C\inf_{h>0}\left(h^{\frac{1}{2}-\frac{d}{2p_0}} + \omega\left(\frac{h}{\varepsilon}\right)\right). \tag{4.9}$$

Proof. For simplicity, in the whole proof, we assume $\|\varphi\|_\infty = 1$ and drop the superscript t from u^t. First, let

$$\mathcal{B}^\varepsilon := \Big|\mathbb{E}\int_0^t\Big(b_\varepsilon(s,Y_s^\varepsilon(x),\mu_s^\varepsilon) - b_0(Y_s^\varepsilon(x),\mu_s)\Big)\cdot\nabla u(s,Y_s^\varepsilon(x))\mathrm{d}s\Big|$$

and

$$\mathcal{E}_h^\varepsilon := \Big|\mathbb{E}\int_0^t\Big(b_\varepsilon(s,Y_{\pi_h(s)}^\varepsilon(x),\mu_{\pi_h(s)}^\varepsilon) - b_0(Y_{\pi_h(s)}^\varepsilon(x),\mu_{\pi_h(s)}^\varepsilon)\Big)\cdot\nabla u(\pi_h(s),Y_{\pi_h(s)}^\varepsilon(x))\mathrm{d}s\Big|,$$

where $h\in(0,1)$. For any map $f:\mathbb{R}_+\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)\to\mathbb{R}^d$ and $h\in(0,1)$, define

$$U^\varepsilon_{1,h}(f):=\Big|\mathbb{E}\int_0^t\big((f\cdot\nabla u)(s,Y^\varepsilon_s(x),\mu^\varepsilon_s)-(f\cdot\nabla u)(s,Y^\varepsilon_{\pi_h(s)}(x),\mu^\varepsilon_s)\big)\mathrm{d}s\Big|$$
$$U^\varepsilon_{2,h}(f):=\Big|\mathbb{E}\int_0^t\Big(\Big[f(s,Y^\varepsilon_{\pi_h(s)}(x),\mu^\varepsilon_s)-f(s,Y^\varepsilon_{\pi_h(s)}(x),\mu^\varepsilon_{\pi_h(s)})\Big]\cdot\nabla u(s,Y^\varepsilon_{\pi_h(s)}(x))\Big)\mathrm{d}s\Big|$$
$$U^\varepsilon_{3,h}(f):=\Big|\mathbb{E}\int_0^t\Big(f(s,Y^\varepsilon_{\pi_h(s)}(x),\mu^\varepsilon_{\pi_h(s)})\cdot\Big[\nabla u(s,Y^\varepsilon_{\pi_h(s)}(x))-\nabla u(\pi_h(s),Y^\varepsilon_{\pi_h(s)}(x))\Big]\Big)\mathrm{d}s\Big|.$$

Then, we have

$$\begin{aligned}\mathcal{B}^\varepsilon\leqslant&\mathcal{E}^\varepsilon_h+\sum_{i=1}^3\Big[U^\varepsilon_{i,h}(b_\varepsilon)+U^\varepsilon_{i,h}(b_0)\Big]\\&+\Big|\mathbb{E}\int_0^t\Big(b_0(Y^\varepsilon_s(x),\mu^\varepsilon_s)-b_0(Y^\varepsilon_s(x),\mu_s)\Big)\cdot\nabla u(s,Y^\varepsilon_s(x))\mathrm{d}s\Big|.\end{aligned}$$

It follows from (4.5), (3.25) and $(\mathbf{H}^1_b)$ that

$$\begin{aligned}&\Big|\mathbb{E}\int_0^t\Big(b_0(Y^\varepsilon_s(x),\mu^\varepsilon_s)-b_0(Y^\varepsilon_s(x),\mu_s)\Big)\cdot\nabla u(s,Y^\varepsilon_s(x))\mathrm{d}s\Big|\\&\lesssim\int_0^t s^{-\frac{d}{2p_0}}(t-s)^{-\frac12}\|\mu^\varepsilon_s-\mu_s\|_{var}\mathrm{d}s,\end{aligned}$$

which implies that

$$\mathcal{B}^\varepsilon\lesssim\mathcal{E}^\varepsilon_h+\sum_{i=1}^3\Big[U^\varepsilon_{i,h}(b_\varepsilon)+U^\varepsilon_{i,h}(b_0)\Big]+\int_0^t s^{-\frac{d}{2p_0}}(t-s)^{-\frac12}\|\mu^\varepsilon_s-\mu_s\|_{var}\mathrm{d}s.\tag{4.10}$$

Now, we divide the rest of the proof into two steps. In Step 1, we estimate $U^\varepsilon_{i,h}(b_\varepsilon)+U^\varepsilon_{i,h}(b_0)$, $i=1,2,3$, one by one; In Step 2, we calculate $\mathcal{E}^\varepsilon_h$ under the assumption $(\mathbf{H}^2_b)$.

(Step 1) We only estimate $U^\varepsilon_{i,h}(b_\varepsilon)$, for $U^\varepsilon_{i,h}(b_0)$ we can proceed in the same way. First, we estimate $U^\varepsilon_{1,h}(b_\varepsilon)$. By (3.34) and (4.5), we have

$$\begin{aligned}U^\varepsilon_{1,h}(b_\varepsilon)&=\Big|\int_0^t\big(P^{Y^\varepsilon}_s(b_\varepsilon\cdot\nabla u)(s,\cdot,\mu^\varepsilon_s)(x)-P^{Y^\varepsilon}_{\pi_h(s)}(b_\varepsilon\cdot\nabla u)(s,\cdot,\mu^\varepsilon_s)(x)\big)\mathrm{d}s\Big|\\&\lesssim\int_0^t\Big[(h^\alpha(\pi_h(s))^{-\alpha})\wedge1\Big](\pi_h(s))^{-\frac{d}{2p_0}}|\!|\!|(b_\varepsilon\cdot\nabla u)(s,\cdot,\mu^\varepsilon_s)|\!|\!|_{p_0}\mathrm{d}s\\&\lesssim h^\alpha\int_0^t(\pi_h(s))^{-\alpha}(\pi_h(s))^{-\frac{d}{2p_0}}(t-s)^{-\frac12}\mathrm{d}s,\end{aligned}$$

where $\alpha=1/2-d/(2p_0)$. Noting that $\pi_h(s)=s$ for $s\leqslant h$, $\pi_h(s)\leqslant s$ for all $s\in[0,T]$ and $1-\alpha-d/(2p_0)-1/2=0$, one sees that

$$U^\varepsilon_{1,h}(b_\varepsilon)\lesssim h^{\frac12-\frac{d}{2p_0}}.$$

For $U^\varepsilon_{2,h}(b_\varepsilon)$, by (3.25), (4.5) and $(\mathbf{H}^1_b)$, we have

$$U^\varepsilon_{2,h}(b_\varepsilon)\lesssim\int_0^t(\pi_h(s))^{-\frac{d}{2p_0}}\|\mu^\varepsilon_s-\mu^\varepsilon_{\pi_h(s)}\|_{var}(t-s)^{-\frac12}\mathrm{d}s.$$

It follows from (3.35) with $\delta=\frac{d}{2p_0}$ that

$$U^\varepsilon_{2,h}(b_\varepsilon)\lesssim h^{\frac12-\frac{d}{2p_0}}\int_0^t(\pi_h(s))^{-\frac{d}{2p_0}}(t-\pi_h(s))^{-\frac12}\mathrm{d}s\lesssim h^{\frac12-\frac{d}{2p_0}},$$

since $p_0 > d$. Finally, in view of (3.25) and (4.6), because $p_0 < \infty$, we have

$$U^\varepsilon_{3,h}(b_\varepsilon) \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}} \int_0^t (\pi_h(s))^{-\frac{d}{2p_0}}(t-\pi_h(s))^{-1+\frac{d}{2p_0}}\mathrm{d}s \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}$$

and obtain that

$$\sum_{i=1}^3 \left(U^\varepsilon_{i,h}(b_\varepsilon) + U^\varepsilon_{i,h}(\bar{b})\right) \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}. \tag{4.11}$$

(Step 2) Let $M := [t/h]$. Without loss of generality we may assume that $M = t/h \in \mathbb{N}$ and note that

$$\begin{aligned}
\mathcal{E}^\varepsilon_h \leqslant & \Big|\mathbb{E}\int_0^h \Big(b_\varepsilon(s, Y^\varepsilon_s(x), \mu^\varepsilon_s) - b_0(Y^\varepsilon_s(x), \mu^\varepsilon_s)\Big)\cdot \nabla u(s, Y^\varepsilon_s(x))\mathrm{d}s\Big| \\
& + \Big|\sum_{k=1}^{M-1}\mathbb{E}\int_{kh}^{(k+1)h}\Big(b_\varepsilon(s, Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh}) - b_0(Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh})\Big)\cdot\nabla u(kh, Y^\varepsilon_{kh}(x))\mathrm{d}s\Big| \\
:= & \mathcal{E}_1 + \mathcal{E}_2.
\end{aligned}$$

From (3.25) and (4.5),

$$\mathcal{E}_1 \lesssim \int_0^h s^{-\frac{d}{2p_0}}(t-s)^{-\frac{1}{2}}\mathrm{d}s \lesssim h^{\frac{1}{2}-\frac{d}{2p_0}}.$$

By (4.5) and a change of variables, one sees that

$$\begin{aligned}
\mathcal{E}_2 &\lesssim \sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\Big|\mathbb{E}\int_{kh}^{(k+1)h}\Big(b_\varepsilon(s, Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh}) - b_0(Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh})\Big)\mathrm{d}s\Big| \\
&\lesssim \sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\Big|\varepsilon\mathbb{E}\int_{kh/\varepsilon}^{(k+1)h/\varepsilon}\Big(b(s, Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh}) - \bar{b}(Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh})\Big)\mathrm{d}s\Big|.
\end{aligned}$$

Based on the assumptions (1.6) and (3.25), we have

$$\begin{aligned}
\mathcal{E}_2 &\lesssim h\sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\omega\left(\frac{h}{\varepsilon}\right)\mathbb{E}H(Y^\varepsilon_{kh}(x), \mu^\varepsilon_{kh}) \\
&\lesssim h\sum_{k=1}^{M-1}(t-kh)^{-\frac{1}{2}}\omega\left(\frac{h}{\varepsilon}\right)(kh)^{-\frac{d}{2p_0}}\sup_\mu |\!|\!| H(\cdot,\mu)|\!|\!|_{p_0} \\
&\lesssim \omega\left(\frac{h}{\varepsilon}\right)\int_h^t (t-\pi_h(s))^{-\frac{1}{2}}(\pi_h(s))^{-\frac{d}{2p_0}}\mathrm{d}s \lesssim \omega\left(\frac{h}{\varepsilon}\right)
\end{aligned}$$

and obtain that

$$|\mathbb{E}\varphi(X^\varepsilon_t) - \mathbb{E}\varphi(X_t)| \lesssim \left(h^{\frac{1}{2}-\frac{d}{2p_0}} + \omega\left(\frac{h}{\varepsilon}\right)\right) + \int_0^t s^{-\frac{d}{2p_0}}(t-s)^{-\frac{1}{2}}\|\mu^\varepsilon_s - \mu_s\|_{var}\mathrm{d}s$$

because of (4.8), (4.10) and (4.11). Finally, taking the supremum over all $\varphi \in C^\infty_b$ with $\|\varphi\|_\infty = 1$ and by Gronwall's inequality of Volterra type (see [51, Example 2.4]), we complete the proof. □

5. Proof of Theoerm 1.2 and 1.3

In this section, we consider the process (X^ε, W, X) on the probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$ which satisfies the following system in $\mathbb{R}^d$:

$$X^\varepsilon_t = \xi + \int_0^t b(\frac{s}{\varepsilon}, X^\varepsilon_s, \mu^\varepsilon_s)\mathrm{d}s + \int_0^t \sigma(X^\varepsilon_s)\mathrm{d}W_s$$

and

$$X_t = \xi + \int_0^t \bar{b}(X_s, \mu_s)\mathrm{d}s + \int_0^t \sigma(X_s)\mathrm{d}W_s,$$

where W is a standard d-dimensional Brownian Motion, μ_t^ε and μ_t are the distributions of X_t^ε and X_t respectively and $(b, \bar{b}, \sigma)$ satisfies the conditions $(\mathbf{H}_b^1)$-$(\mathbf{H}_b^2)$ and $(\mathbf{H}_\sigma)$. We set

$$X^0 := X, \quad b_\varepsilon(t) := b(t/\varepsilon), \quad \text{and} \quad b_0 := \bar{b}.$$

Proof of Theoerm 1.2. Set

$$B(t,x) := b_0(x, \mu_t)$$

and consider the following backward parabolic PDE

$$\partial_t u + a_{ij}\partial_i\partial_j u - \lambda u + B\cdot\nabla u + B = 0,\ t\in[0,T], \quad u(T) = 0.$$

Since $\|B\|_{\widetilde{\mathbb{L}}^{p_0}(T)} \leqslant \sup_\mu \|\bar{b}(\cdot,\mu)\|_{p_0} < \infty$, by Lemma 2.5, for λ large enough there is a unique solution u in the sense of Definition 2.3 satisfying

$$\|\nabla u\|_{\mathbb{L}_T^\infty} \leqslant \frac{1}{2}$$

and for any $2/q + d/p_0 < 1$,

$$\|\nabla^2 u\|_{\widetilde{\mathbb{L}}_q^{p_0}(T)} \leqslant C,$$

which implies that for any $\lambda > 0$

$$\sup_{\varepsilon\geqslant 0} \mathbb{E}\exp\left(\lambda\int_0^T |\nabla^2 u(t, X_t^\varepsilon)|^2 \mathrm{d}t\right) < \infty, \tag{5.1}$$

where $X^0 := X$, because of (3.27). Moreover, by (3.41), for all $s,t\in[0,T]$,

$$\|\nabla u(t) - \nabla u(s)\|_\infty \lesssim |t-s|^{1/2 - d/(2p_0)}. \tag{5.2}$$

Define

$$\Phi_t(x) := x + u(t,x)$$

and

$$Y_t^\varepsilon := \Phi_t(X_t^\varepsilon), \quad Y_t := \Phi_t(X_t).$$

Then Φ_t is a C^1-diffeomorphism (see Remark 3.18) for any $t\in[0,T]$ with

$$\|\nabla\Phi\|_{\mathbb{L}_T^\infty} + \|\nabla\Phi^{-1}\|_{\mathbb{L}_T^\infty} \leqslant 4. \tag{5.3}$$

By the generalized Itô formula (3.4), we have

$$\mathrm{d}Y_t = \lambda u(t, X_t)\mathrm{d}t + (\sigma^*\nabla\Phi_t)(X_t)\mathrm{d}W_t$$

and

$$\mathrm{d}Y_t^\varepsilon = \lambda u(t, X_t^\varepsilon)\mathrm{d}t + \big(b_\varepsilon(t, X_t^\varepsilon, \mu_t^\varepsilon) - b_0(X_t^\varepsilon, \mu_t)\big)\cdot\nabla\Phi_t\big(X_t^\varepsilon\big)\mathrm{d}t + (\sigma^*\nabla\Phi_t(X_t^\varepsilon))\mathrm{d}W_t,$$

where σ^* is the transpose of σ. It follows from (5.3) that for any $t\in[0,T]$,

$$\begin{aligned}|X_t^\varepsilon - X_t|^2 \lesssim |Y_t^\varepsilon - Y_t|^2 \lesssim \lambda\|\nabla u\|_{\mathbb{L}_T^\infty}^2 &\int_0^t |X_s^\varepsilon - X_s|^2\mathrm{d}s \\ &+ \left[\int_0^t \big((\sigma^*\nabla\Phi_s)(X_s^\varepsilon) - (\sigma^*\nabla\Phi_s)(X_s)\big)\mathrm{d}W_s\right]^2 \\ &+ \Big|\int_0^t \big(b_\varepsilon(s, X_s^\varepsilon, \mu_s^\varepsilon) - b_0(X_s^\varepsilon, \mu_s)\big)\cdot\nabla\Phi_s\big(X_s^\varepsilon\big)\mathrm{d}s\Big|^2.\end{aligned}$$

Set

$$A_t^\varepsilon := \int_0^t \big(\mathcal{M}(\nabla^2 u)(s, X_s) + \mathcal{M}(\nabla^2 u)(s, X_s^\varepsilon) + \|\nabla u\|_{\mathbb{L}_T^\infty}\big)^2\,\mathrm{d}s$$

$$+\int_0^t (\mathcal{M}(\nabla\sigma)(X_s)+\mathcal{M}(\nabla\sigma)(X_s^\varepsilon)+\|\sigma\|_\infty)^2\,\mathrm{d}s$$

and

$$\eta_t^\varepsilon := \Big|\int_0^t \big(b_\varepsilon(s,X_s^\varepsilon,\mu_s^\varepsilon)-b_0(X_s^\varepsilon,\mu_s)\big)\cdot\nabla\Phi_s\big(X_s^\varepsilon\big)\mathrm{d}s\Big|^2.$$

Then, by (2.2), (5.1) and $(\mathbf{H}_\sigma)$, we have

$$\sup_\varepsilon \mathbb{E}\exp(A_T^\varepsilon)<\infty. \tag{5.4}$$

We note that by (2.1)

$$\begin{aligned}&\Big[\int_0^t \big((\sigma^*\nabla\Phi)(X_s^\varepsilon)-(\sigma^*\nabla\Phi)(X_s)\big)\mathrm{d}W_s\Big]^2\\ &\leqslant \int_0^t |X_s^\varepsilon-X_s|^2\mathrm{d}A_s^\varepsilon+M_t^\varepsilon,\end{aligned}$$

where M^ε is a martingale. Altogether, we have

$$|X_t^\varepsilon-X_t|^2\lesssim\int_0^t|X_s^\varepsilon-X_s|^2\mathrm{d}s+\int_0^t|X_s^\varepsilon-X_s|^2\mathrm{d}A_s^\varepsilon+M_t^\varepsilon+\eta_t^\varepsilon.$$

Hence, by (5.4) and the Stochastic Gronwall inequality (see Lemma 2.8 in [49] or [42]), one sees that for any $\ell\in(0,1)$,

$$\mathbb{E}\big(\sup_{t\in[0,T]}|X_t^\varepsilon-X_t|^{2\ell}\big)\lesssim\big(\mathbb{E}[\sup_{t\in[0,T]}\eta_t^\varepsilon]\big)^\ell.$$

Combining (5.2), (5.3) and (3.39), we have

$$\mathbb{E}[\sup_{t\in[0,T]}\eta_t^\varepsilon]\lesssim\mathbb{E}\int_0^T\Big|b_0(X_s^\varepsilon,\mu_s^\varepsilon)-b_0(X_s^\varepsilon,\mu_s)\Big|^2\mathrm{d}s+\inf_{h>0}\Big(h^{1-\delta}+h^{1-\frac{d}{p_0}}+\omega\Big(\frac{h}{\varepsilon}\Big)\Big). \tag{5.5}$$

Taking $\delta<d/p_0$ in (5.5), from (3.26) and (4.9), for any $2/q+d/p_0<1$, one sees that

$$\begin{aligned}\mathbb{E}[\sup_{t\in[0,T]}\eta_t^\varepsilon]&\lesssim\Big(\int_0^T\|\mu_s^\varepsilon-\mu_s\|_{var}^q\mathrm{d}s\Big)^{2/q}+\inf_{h>0}\Big(h^{1-\frac{d}{p_0}}+\Big(\omega\Big(\frac{h}{\varepsilon}\Big)\Big)^2\Big)\\ &\lesssim\inf_{h>0}\Big(h^{1-\frac{d}{p_0}}+\Big(\omega\Big(\frac{h}{\varepsilon}\Big)\Big)^2\Big)\end{aligned}$$

and this completes the proof. □

In the rest of this section, we assume that

$$b_\varepsilon(t,x,\mu)=b_\varepsilon(t,x),\quad b_0(x,\mu)=b_0(x)$$

and prove Theorem 1.3. The method is the same as the one of Theorem 1.2, except for using the elliptic equation to construct the Zvonkin's transformation instead of the parabolic.

Proof of Theorem 1.3. Consider the following elliptic PDE

$$a_{ij}\partial_i\partial_j u-\lambda u+b_0\cdot\nabla u+b_0=0. \tag{5.6}$$

Noting that $\|b_0\|_{\widetilde{\mathbb{L}}^{p_0}(T)}\leqslant\|b\|_{L^\infty(\mathbb{R}_+;\widetilde{L}^{p_0})}$ and by (2.9) for λ large enough, we have

$$\|\nabla u\|_\infty\leqslant\frac12$$

and

$$\|\nabla^2u\|_{p_0}\leqslant C\|b_0\|_{\widetilde{L}^{p_0}},\quad\forall p_0>d$$

It follows by (3.27) that for any $\lambda > 0$

$$\sup_{\varepsilon\geqslant 0}\mathbb{E}\exp\left(\lambda\int_0^T|\nabla^2 u(X_t^\varepsilon)|^2\mathrm{d}t\right)<\infty. \tag{5.7}$$

Define

$$\Phi(x):=x+u(x)$$

and

$$Y_t^\varepsilon:=\Phi(X_t^\varepsilon),\quad Y_t:=\Phi(X_t).$$

Then Φ is a C^1-diffeomorphism. Again by the generalized Itô formula (3.4), we have

$$\mathrm{d}Y_t=\lambda u(X_t)\mathrm{d}t+(\sigma^*\nabla\Phi)(X_t)\mathrm{d}W_t$$

and

$$\mathrm{d}Y_t^\varepsilon=\lambda u(X_t^\varepsilon)\mathrm{d}t+\big(b_\varepsilon(t,X_t^\varepsilon)-b_0(X_t^\varepsilon)\big)\cdot\nabla\Phi\big(X_t^\varepsilon\big)\mathrm{d}t+(\sigma^*\nabla\Phi)(X_t^\varepsilon)\mathrm{d}W_t.$$

Then, we have

$$|X_t^\varepsilon-X_t|^2\lesssim\int_0^t|X_s^\varepsilon-X_s|^2\mathrm{d}A_s^\varepsilon+M_t^\varepsilon+\eta_t^\varepsilon,$$

where $(M_t^\varepsilon)_{t\geqslant 0}$ is a martingale,

$$\begin{aligned}A_t^\varepsilon=&t+\int_0^t\big(\mathcal{M}(\nabla^2u)(X_s)+\mathcal{M}(\nabla^2u)(X_s^\varepsilon)+\|\sigma\|_\infty\big)^2\mathrm{d}s\\&+\int_0^t(\mathcal{M}(\nabla\sigma)(X_s)+\mathcal{M}(\nabla\sigma)(X_s^\varepsilon)+\|\nabla u\|_\infty)^2\mathrm{d}s\end{aligned}$$

and

$$\eta_t^\varepsilon=\Big|\int_0^t\big(b(s/\varepsilon,X_s^\varepsilon)-\bar b(X_s^\varepsilon)\big)\cdot\nabla\Phi\big(X_s^\varepsilon\big)\mathrm{d}s\Big|^2.$$

Then, in view of (2.2) and (5.7), we have

$$\sup_\varepsilon\mathbb{E}\exp(A_T^\varepsilon)<\infty,$$

which implies that for any $\ell\in(0,1)$,

$$\mathbb{E}\big(\sup_{t\in[0,T]}|X_t^\varepsilon-X_t|^{2\ell}\big)\lesssim\big(\mathbb{E}\sup_{t\in[0,T]}\eta_t^\varepsilon\big)^\ell$$

because of the Stochastic Gronwall inequality and we complete the proof by (3.39) with $\alpha=1$. $\square$

Acknowledgments. The first author would like to acknowledge the warm hospitality of Bielefeld University. And we would also like to thank Dr. Chengcheng Ling for many useful discussions.

Appendix A.

In this appendix, we prove the claim (1.15) in Example 1.7. To this end, we need the following lemma.

Lemma A.1. *Let*

$$h(t):=\int_{\mathbb{R}}B(\sin(\xi t))\nu(\mathrm{d}\xi).$$

and

$$\bar h:=\Big(\frac{1}{2\pi}\int_0^{2\pi}B(\sin(\xi\tau))\mathrm{d}\tau\Big)\nu(\mathbb{R}\setminus\{0\})+B(0)\nu(\{0\}),$$

where $B : [-1,1]$ is measurable and ν is a finite measure on $\mathbb{R}$. Assume that there is a constant $C_B > 0$ such that

$$|B(u)| \leqslant C_B, \quad \forall u \in [-1,1].$$

Then, for any $t, T \in \mathbb{R}_+$,

$$\Big|\frac{1}{T}\int_t^{t+T}\big(h(s)-\bar{h}\big)\mathrm{d}s\Big| \leqslant \frac{4\pi C_B}{T}\int_{\mathbb{R}\setminus\{0\}}\frac{\nu(\mathrm{d}\xi)}{|\xi|}.$$

Proof of Lemma A.1. If $\int_{\mathbb{R}\setminus\{0\}}\frac{\nu(\mathrm{d}\xi)}{|\xi|} = \infty$, this is trivial. So, we assume that $\int_{\mathbb{R}\setminus\{0\}}\frac{\nu(\mathrm{d}\xi)}{|\xi|} < \infty$. First, one sees that

$$\begin{aligned}
\mathscr{I} &:= \left|\frac{1}{T}\int_t^{t+T}\big(h(s)-\bar{h}\big)\mathrm{d}s\right| \\
&= \left|\frac{1}{T}\int_t^{t+T}\int_{\mathbb{R}\setminus\{0\}}B(\sin(\xi s))\nu(\mathrm{d}\xi)\mathrm{d}s - \frac{1}{2\pi}\int_t^{t+2\pi}B(\sin(\tau))\nu(\mathbb{R}\setminus\{0\})\mathrm{d}\tau\right| \\
&= \left|\int_{\mathbb{R}\setminus\{0\}}\Big[\frac{1}{T}\int_t^{t+T}B(\sin(\xi s))\mathrm{d}s - \frac{1}{2\pi}\int_t^{t+2\pi}B(\sin(\tau))\mathrm{d}\tau\Big]\nu(\mathrm{d}\xi)\right|,
\end{aligned}$$

by Fubini's theorem. From a change of variable, we have

$$\mathscr{I} = \left|\int_{\mathbb{R}\setminus\{0\}}\Big[\frac{1}{T|\xi|}\int_{t\xi}^{(t+T)\xi}B(\sin s)\mathrm{d}s - \frac{1}{2\pi}\int_t^{t+2\pi}B(\sin\tau)\mathrm{d}\tau\Big]\nu(\mathrm{d}\xi)\right|,$$

where $\int_a^b := -\int_b^a$ if $a > b$. Set

$$G := \int_0^{2\pi}B(\sin s)\mathrm{d}s = \int_t^{t+2\pi}B(\sin s)\mathrm{d}s, \quad \forall t \in \mathbb{R}.$$

Then, noting that $s \to \sin s$ has a period 2π, we have

$$\begin{aligned}
\int_{t\xi}^{(t+T)\xi}B(\sin s)\mathrm{d}s &= \Big[\frac{T|\xi|}{2\pi}\Big]G + \int_{\mathrm{sgn}(\xi)[\frac{T|\xi|}{2\pi}]2\pi+t\xi}^{T\xi+t\xi}B(\sin s)\mathrm{d}s \\
&:= \Big[\frac{T|\xi|}{2\pi}\Big]G + H_t(\xi)
\end{aligned}$$

where $\mathrm{sgn}(\xi) := \xi/|\xi|$, which implies that

$$\begin{aligned}
\mathscr{I} &= \left|\int_{\mathbb{R}\setminus\{0\}}\Big[\frac{1}{T|\xi|}\Big(\Big[\frac{T|\xi|}{2\pi}\Big]G + H_t(\xi)\Big) - \frac{1}{2\pi}G\Big]\nu(\mathrm{d}\xi)\right| \\
&\leqslant \int_{\mathbb{R}\setminus\{0\}}\Big|\frac{1}{T|\xi|}\Big[\frac{T|\xi|}{2\pi}\Big] - \frac{1}{2\pi}\Big|\nu(\mathrm{d}\xi)G + \int_{\mathbb{R}\setminus\{0\}}\frac{1}{T|\xi|}H_t(\xi)\nu(\mathrm{d}\xi).
\end{aligned}$$

We note that

$$\Big|\frac{1}{T|\xi|}\Big[\frac{T|\xi|}{2\pi}\Big] - \frac{1}{2\pi}\Big| = \frac{1}{T|\xi|}\Big|\Big[\frac{T|\xi|}{2\pi}\Big] - \frac{T|\xi|}{2\pi}\Big| \leqslant \frac{1}{T|\xi|}$$

and

$$G \vee H_t(\xi) \leqslant \int_0^{2\pi}|B(\sin s)|\mathrm{d}s \leqslant 2\pi C_B.$$

Therefore, we have

$$\mathscr{I} \leqslant \frac{4\pi C_B}{T}\int_{\mathbb{R}\setminus\{0\}}\frac{\nu(\mathrm{d}\xi)}{|\xi|}$$

and complete the proof.

□

Now we can give the

Proof of (1.15). Since $(\mathbf{H}_b^1)$ holds for b obviously, it suffice to show that $(\mathbf{H}_b^2)$ holds. We note that in Example 1.7

$$|F(u,x)| \leqslant |F(u,0)| + |F(u,0) - F(t,x)| \leqslant L_F + L_F|x|$$

because of (1.14), which implies that

$$|F\big(u, \int_{\mathbb{R}^d} \phi(x,y)\mu(\mathrm{d}y)\big)| \leqslant L_F(1 + \int_{\mathbb{R}^d} |\phi(x,y)|\mu(\mathrm{d}y)).$$

Hence, by Lemma A.1, we see that

$$\left|\frac{1}{T}\int_t^{t+T} \big(b(s,x,\mu) - \bar{b}(x,\mu)\big)\mathrm{d}s\right| \leqslant \frac{4\pi L_F}{T}\int_{\mathbb{R}\setminus\{0\}} \frac{\nu(\mathrm{d}\xi)}{|\xi|} H(x,\mu),$$

where $H(x,\mu) = 1 + \int_{\mathbb{R}^d} |\phi(x,y)|\mu(\mathrm{d}y)$. It is easy to see that

$$\sup_\mu |\!|\!| H(\cdot,\mu) |\!|\!|_{p_0} \leqslant \|1\|_\infty + \int_{\mathbb{R}^d} |\!|\!| \phi(\cdot,y) |\!|\!|_{p_0} \mu(\mathrm{d}y) \leqslant 1 + \sup_y |\!|\!| \phi(\cdot,y) |\!|\!|_{p_0}.$$

This completes the proof. □

References

[1] Bakhtin V. and Kifer Y., Diffusion approximation for slow motion in fully coupled averaging. *Probab. Theory Related Fields* **129** (2004), 157–181.

[2] Barbu V. and Röckner M., Probabilistic representation for solutions to nonlinear Fokker-Planck equations. *SIAM J. Math. Anal.* **50** (2018), 4246–4260.

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

[4] Barbu V. and Röckner M., Nonlinear Fokker-Planck equations with fractional Laplacian and McKean-Vlasov SDEs with Lévy-Noise. *arXiv:2210.05612* (2022).

[5] Barbu V. and Röckner M., Uniqueness for nonlinear Fokker-Planck equations and for McKean-Vlasov SDEs: The degenerate case. *arXiv:2203.00122* (2022).

[6] Bogoliubov N. N. and Mitropolsky Y. A., *Asymptotic Methods in the Theory of Non-linear Oscillations.* Translated from the second revised Russian edition International Monographs on Advanced Mathematics and Physics Hindustan Publishing Corp., Delhi, Gordon and Breach Science Publishers, Inc., New York 1961 x+537 pp.

[7] Carmona, R. and Delarue, F., Probabilistic analysis of mean-field games. *SIAM J. Control Optim.* **51** (2013), 2705–2734.

[8] Carmona R. and Delarue F., *Probabilistic theory of mean field games with applications. II. Mean field games with common noise and master equations. Probability Theory and Stochastic Modeling*, **84**. Springer, Cham, 2018. xxiv+697 pp.

[9] Carrillo J. A., Gvalani R. S., Pavliotis G. A. and Schlichting A., Long-time behavior and phase transitions for the McKean-Vlasov equation on the tours. *Arch. Ration. Mech. Anal.* **235** (2020), 635–690.

[10] Cerrai S., A Khasminskii type averaging principle for stochastic reaction-diffusion equations. *Ann. Appl. Probab.* **19** (2009), 899–948.

[11] Cerrai S. and Freidlin M., Averaging principle for a class of stochastic reaction-diffusion equations. *Probab. Theory Related Fields* **144** (2009), 137–177.

[12] Cerrai S. and Lunardi A., Averaging principle for nonautonomous slow-fast systems of stochastic reaction-diffusion equations: the almost periodic case. *SIAM J. Math. Anal.* **49** (2017), 2843–2884.

[13] Chen Z.–Q., Hu E., Xie L. and Zhang X., Heat kernels for non-symmetric diffusion operators with jumps. *J. Differential Equations* **263** (2017), 6576–6634.

[14] Cheng M. and Liu Z., The second Bogolyubov theorem and global averaging principle for SPDEs with monotone coefficients. To appear in *SIAM J. Math. Anal.* Preprint version available at *https://arxiv.org/abs/2109.00371.*

[15] Crippa G. and De Lellis D., Estimates and regularity results for the DiPerna-Lions flow. *J. Reine Angew. Math.* **616** (2008), 15–46.

[16] Dareiotis K. and Gerencsér M., On the regularisation of the noise for the Euler-Maruyama scheme with irregular drift. *Electron. J. Probab.* **25** (2020) 1-18.

[17] Duan J. and Wang W., *Effective Dynamics of Stochastic Partial Differential Equations.* Elsevier Insights. Elsevier, Amsterdam, 2014, xii+270 pp.

[18] Freidlin M. I. and Wentzell A. D., *Random Perturbations of Dynamical Systems.* Translated from the 1979 Russian original by Joseph Szücs. Third edition. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 260. Springer, Heidelberg, 2012, xxviii+458 pp.

[19] Hairer M. and Li X.–M., Averaging dynamics driven by fractional Brownian motion. *Ann. Probab.* **48** (2020), 1826–1860.

[20] Hammersley W., Siska D. and Szpruch L., McKean-Vlasov SDEs under measure dependent Lyapunov conditions. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** (2021), 1032–1057.

[21] Han Y., Solving McKean-Vlasov SDEs via relative entropy. *arXiv:2204.05709* (2022).

[22] Hao Z., Röckner M. and Zhang X., Euler scheme for density dependent stochastic differential equations. *J. Differential Equations* **274** (2021), 996–1014.

[23] Hao Z., Röckner M. and Zhang X., Strong convergence of propagation of chaos for McKean-Vlasov SDEs with singular interactions. *arXiv:2204.07952* (2022).

[24] Hong W., Li S. and Liu W., Strong convergence rates in averaging principle for slow-fast McKean-Vlasov SPDEs. *J. Differential Equations* **316** (2022), 94–135.

[25] Khasminshii R. Z., On the principle of averaging the Itô's stochastic differential equations. (Russian) *Kybernetika (Prague)* **4** (1968), 260–279.

[26] Kifer Y., Some recent advances in averaging. *Modern dynamical systems and applications, 385-403, Cambridge Univ. Press, Cambridge*, 2004.

[27] Krylov N. V. and Bogolyubov N. N., *Introduction to Non-Linear Mechanics.* Annals of Mathematics Studies, No. 11 Princeton University Press, Princeton, N. J., 1943. iii+105 pp.

[28] Krylov N. V. and Röckner M., Strong solutions of stochastic equations with singular time dependent drift. *Probab. Theory Relat. Fields* **131** (2005), 154–196.

[29] Lacker D., On a strong form of propagation of chaos for McKean-Vlasov equations, *Electron. Commun. Probab.* **23** (2018), 1–11.

[30] Lacker D., Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *arXiv:2105.02983* (2021).

[31] Lê K. and Ling C., Taming singular stochastic differential equations: A numerical method. *arXiv:2110.01343* (2021).

[32] Maslowski B., Seidler J. and Vrkoč I., An averaging principle for stochastic evolution equations. II. *Math. Bohem.* **116** (1991), 191–224.

[33] McKean H. P., A class of Markov processes associated with nonlinear parabolic equations. *Proc Nat. Acad Sci USA* **56** (1966), 1907–1911.

[34] Mishura Y. S. and Veretennikov A. Yu., Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations. *Theory Probab. Math. Statist.* **103** (2020), 59–101.

[35] Pei B., Inahama Y. and Xu Y., Averaging principle for fast-slow system driven by mixed fractional Brownian rough path. *J. Differential Equations* **301** (2021), 202–235.

[36] Röckner M., Sun X. and Xie L., Strong and weak convergence in the averaging principle for SDEs with Hölder coefficients. *arXiv:1907.09256* (2019).

[37] Röckner M., Sun X. and Xie Y., Strong convergence order for slow-fast McKean-Vlasov stochastic differential equations. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** (2021), 547–576.

[38] Röckner M., Xie L. and Yang l., Asymptotic behavior of multiscale stochastic partial differential equations. *arXiv:2010.14897* (2020).

[39] Röckner M. and Zhang X., Well-posedness of distribution dependent SDEs with singular drifts. *Bernoulli* **27** (2021), 1131–1158.

[40] Sanders J. A. and Verhulst F., *Averaging Methods in Nonlinear Dynamical Systems.* Applied Mathematical Sciences, 59. Springer-Verlag, New York, 1985. x+247 pp.

[41] Scheutzow M., A stochastic Gronwall lemma. *Infinite Dimens. Anal. Quantum Probab. Rel. Top.* **16** (2013), 1350019, 4pp.

[42] Shen G., Xiang J. and Wu J., Averaging principle for distribution dependent stochastic differential equations driven by fractional Brownian motion and standard Brownian motion. *J. Differential Equations* **321** (2022), 381–414.

[43] Sznitman A.–S., Topics in propagation of chaos. In *École d'Été de Prob. de Saint-Flour XIX*-1989. *Lect. Notes in Math.* **1464** 165-251. Berlin: Springer-Verlag 1991.

[44] Veretennikov A. Yu., On large deviations in the averaging principle for SDEs with a "full dependence". *Ann. Probab.* **27** (1999), 284–296.

[45] Wang F.–Y., Distribution dependent SDEs for Landau type equations. *Stochastic Process. Appl.* **128** (2018), 595–621.

[46] Wang W. and Roberts A. J., Average and deviation for slow-fast stochastic partial differential equations. *J. Differential Equations* **253** (2012), 1265–1286.

[47] Xia P., Xie L., Zhang X. and Zhao G., $L^q(L^p)$-theory of stochastic differential equations. *Stochastic Processes Appl.* **130** (2020), 5188–5211.
[48] Zhang, X. and Zhao, G., Singular Brownian diffusion processes. *Commun. Math. Stat.* **6** (2018), 533–581.
[49] Zhao G., On distribution depend SDEs with singular drifts. *arXiv:2003.04829v3* (2020).
[50] Zhang X., Stochastic Volterra equations in Banach spaces and stochastic partial differential equation. *J. Funct. Anal.* **258** (2010), 1361–1425.
[51] Zhang X., Weak solutions of McKean-Vlasov SDEs with supercritical drifts. To appear in *Commun. Math. Stat.* Preprint version available at *https://arxiv.org/abs/2010.15330*.
[52] Zvonkin A. K., A transformation of the phase space of a diffusion process that removes the drift. (Russian) *Mat. Sb. (N.S.)* **93(135)** (1974), 129-149.

Mengyu Cheng: Fakultät für Mathematik, Universität Bielefeld, 33615 Bielefeld, Germany; School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, P. R. China
Email: mengyucheng@mail.dlut.edu.cn; mengyu.cheng@hotmail.com

Zimo Hao: Fakultät für Mathematik, Universität Bielefeld, 33615, Bielefeld, Germany; School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P. R. China
Email: zimohao@whu.edu.cn; zhao@math.uni-bielefeld.de

Michael Röckner: Fakultät für Mathematik, Universität Bielefeld, 33615, Bielefeld, Germany; Academy of Mathematics and System Sciences, Chinese Academy of Science, Beijing 100190, P. R. China
Email: roeckner@math.uni-bielefeld.de