Stochastic averaging principle for two-time-scale SDEs with distribution dependent coefficients driven by fractional Brownian motion

Guangjun Shen, Jiayuan Yin and Jiang-Lun Wu*

October 1, 2022

Abstract

In this paper, we derive an averaging principle for a fast-slow system of stochastic differential equations (SDEs) involving distribution dependent coefficients driven by both fractional Brownian motion (fBm) and standard Brownian motion (Bm). We first establish the existence and uniqueness of solutions of the fast-slow system and the corresponding averaging equation. Then, we show that the slow component strongly converges to the solution of the associated averaged equation.

Keywords: averaging principle; fast-slow systems; fractional Brownian motion; standard Brownian motion.

2020 Mathematics Subject Classification: 60H10; 60G22; 34C29; 35Q83

1 Introduction

In the seminal papers [14, 15], Kac proposed the "propagation of chaos" of mean field particle systems in order to study nonlinear PDEs in Vlasov's kinetic theory. This motivated McKean [20] to study nonlinear Fokker-Planck equations by utilising stochastic differential equations with distribution dependent drift coefficients. In general, nonlinear Fokker-Planck equations can be characterised by distribution dependent stochastic differential equations, which are also named as McKean-Vlasov SDEs or mean field SDEs. A distinct feature of such systems is the appearance of probability laws in the coefficients of the resulting equations (for more comprehensive overview, the reader is referred to Wang [31], Huang and Wang [11], Mehri and Stannat [21], Huang, Ren and Wang [13] and the references therein).

In this paper, we are concerned with the averaging principle for fast-slow systems of distribution dependent stochastic differential equations (DDSDEs, for short) of the form

$$\begin{cases} dX_t^\epsilon = b(t, X_t^\epsilon, \mathscr{L}_{X_t^\epsilon}, Y_t^\epsilon)dt + \sigma(t, \mathscr{L}_{X_t^\epsilon})dB_t^H, \ X_0^\epsilon = x \in \mathbb{R}^n, \\ dY_t^\epsilon = \frac{1}{\epsilon} f(t, X_t^\epsilon, \mathscr{L}_{X_t^\epsilon}, Y_t^\epsilon)dt + \frac{1}{\sqrt{\epsilon}} g(t, X_t^\epsilon, \mathscr{L}_{X_t^\epsilon}, Y_t^\epsilon)dW_t, \ Y_0^\epsilon = y \in \mathbb{R}^m, \end{cases} \tag{1.1}$$

*Corresponding author. This research is supported by the National Natural Science Foundation of China (12071003)

where $\mathscr{L}_{X_t^\epsilon}$ stands for the law of X_t^ϵ. The parameter ϵ represents the ratio between the natural time scales of $X_t^\epsilon \in \mathbb{R}^n$ and $Y_t^\epsilon \in \mathbb{R}^m$. We are concerned with situations where $\epsilon \ll 1$, i.e., with a separation of scales, in such a case the vector X_t^ϵ is called the "slow component" of the systems, and the vector Y_t^ϵ is called the "fast component" of the systems. The driving process B_t^H and W_t are independent d_1 dimensional fractional Brownian motions with Hurst parameter $H \in (\frac{1}{2}, 1)$ and d_2 dimensional Wiener processes, respectively. The coefficients

$$\begin{aligned}
b &: [0,T] \times \mathbb{R}^n \times \mathscr{P}_\theta(\mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^n, \\
\sigma &: [0,T] \times \mathscr{P}_\theta(\mathbb{R}^n) \to \mathbb{R}^{n\times d_1}, \\
f &: [0,T] \times \mathbb{R}^n \times \mathscr{P}_\theta(\mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^m, \\
g &: [0,T] \times \mathbb{R}^n \times \mathscr{P}_\theta(\mathbb{R}^n) \times \mathbb{R}^m \to \mathbb{R}^{m\times d_2},
\end{aligned}$$

with

$$\mathscr{P}_\theta(\mathbb{R}^n) := \left\{ \mu \in \mathbb{P}(\mathbb{R}^n) : \mu(|\cdot|^\theta) := \int_{\mathbb{R}^n} |x|^\theta \mu(dx) < \infty \right\}, \quad \theta \in [2, \infty),$$

where $\mathbb{P}$ is the set of probability measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. The space $\mathscr{P}_\theta(\mathbb{R}^n)$ is a Polish space under the L^θ-Wasserstein distance

$$\mathbb{W}_\theta(\mu_1, \mu_2) := \inf_{\pi \in \mathscr{C}(\mu_1,\mu_2)} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} |x-y|^\theta \pi(dx, dy) \right)^{\frac{1}{\theta}}, \quad \mu_1, \mu_2 \in \mathscr{P}_\theta(\mathbb{R}^n),$$

where $\mathscr{C}(\mu_1, \mu_2)$ is the set of probability measures on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals μ_1 and μ_2, respectively.

We recall that fractional Brownian motion with Hurst parameter $H \in (0, 1)$ is a centred Gaussian process $B^H = \{B_t^H, t \geq 0\}$ with covariance function

$$R_H(s,t) := \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H}), \ s, t \geq 0.$$

For $H = \frac{1}{2}$, fBm B^H is the standard Brownian motion. For $H \neq \frac{1}{2}$, B^H is neither a semimartingale nor a Markov process. As a consequence, classical techniques of stochastic analysis are not applicable. In particular, an fBm with a Hurst parameter $H \in (\frac{1}{2}, 1)$ possesses a property of long range memory, which roughly implies that the decay of stochastic dependence with respect to the past is only subexponentially slow, what makes this kind of noise a realistic choice for problems with long memory in the applied sciences including hydrology, telecommunication, turbulence, image processing, and finance ([2], [22], [23], [10]), and this is why this kind of noise is being used now very often.

The averaging principle, initiated by Khasminskii [16], is a very efficient and important tool in study of stochastic differential equations for modeling problems arising in many practical research areas. It in fact provides a powerful tool for simplifying dynamical systems, and obtains approximate solutions to differential equations. The averaging principle enables us to study complex equations with related averaging equations, which paves a convenient and easy way to study many important properties. To date, the stochastic averaging principle has been developed for many more general types of stochastic differential equations (see, for example, [8], [4], [24],[37],[38], [19], [28], [29],[30], [17], [9] just to mention a few).

A natural generalisation of the averaging principle, which will be carried out in this paper, can be illustrated as follows. Assume that for every fixed x the rapid variables induce a unique invariant measure $\nu^{t,x,\mu}$. Then, as $\epsilon \to 0$, X_t^ϵ converges on every finite interval $[0,T]$ to the solution $\bar{X}$ which is the solution of the following averaged equation,

$$\begin{cases} d\bar{X}_t = \bar{b}(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t})dt + \sigma(t, \mathscr{L}_{\bar{X}_t})dB_t^H, \\ \bar{X}_0 = x, \end{cases} \tag{1.2}$$

where

$$\bar{b}(t, x, \mu) = \int_{\mathbb{R}^m} b(t, x, \mu, z)\nu^{t,x,\mu}(dz),$$

and $\nu^{t,x,\mu}$ is the unique invariant measure for the transition semigroup of the solution of the following frozen equation,

$$\begin{cases} dY_s = f(t, x, \mu, Y_s)ds + g(t, x, \mu, Y_s)d\tilde{W}_s, \\ Y_0 = y, \end{cases} \tag{1.3}$$

where $\tilde{W}_t$ is a d_2-dimensional Brownian motion on another given complete probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ and $\{\tilde{\mathscr{F}}_t, t \geq 0\}$ is the natural filtration generated by $\tilde{W}_t$. It is worth noting that, for any initial data $y \in \mathbb{R}^m$, Equation (1.3) has a unique strong solution $\{Y_s^{t,x,\mu,y}\}_{s\geq 0}$, which is a homogeneous Markov process, so its transition semigroup has a unique invariant measure $\nu^{t,x,\mu}$ under appropriate conditions. Hence, the definition of the averaged coefficient $\bar{b}$ is meaningful. From mathematical point of view, it is possible to model systems with different time-scales and then operate a rigorous dimensionality reduction, approximating the behavior of the slow component X_t^ϵ with $\bar{X}_t$ and controlling the error of such approximation. For this reason multi-scale stochastic systems are widely used in many areas of physics, chemistry, biology, financial mathematics and many other applications areas (see, for example, [6], [32]).

In the distribution independent case, there have been many fundamental studies addressing the averaging principle for two-time scale stochastic systems driven Brownian motion, Lévy process and fractional Brownian motion. Xu, Liu and Miao [34] proved the L^2 convergence for two-time-scales with special non-Lipschitz which extends the existing results from Lipschitz to non-Lipschitz case. Liu et al. [18] used the techniques of time discretisation and truncation to study the averaging principle for stochastic differential equations with slow and fast time-scales, where the drift coefficients satisfied local Lipschitz conditions with respect to the slow and fast variables. Givon [7] considered two-time-scale system of jump-diffusion stochastic differential equations and studied the convergence rate of the slow components to the effective dynamics. Xu and Liu [35] proved a stochastic averaging principle for two time-scale jump-diffusion SDEs under the non-Lipschitz coefficients. Pei et al. [25] studied the averaging principle for a fast-slow system of rough differential equations driven by mixed fractional Brownian rough path. The fast component is driven by Brownian motion, while the slow component is driven by fractional Brownian motion. Pei et al. [26] considered averaging principle for fast-slow systems involving both fractional Brownian motion and standard Brownian motion. More recently, there has been an increasing interest to study the stochastic averaging principle for two time-scale distribution dependent stochastic differential equations. Under some proper assumptions on the coefficients, Röckner, Sun and Xie [27] proved that the slow component strongly converges to the solution of the corresponding averaged equation with convergence order $\frac{1}{3}$ using the approach of time discretisation. Furthermore, under stronger regularity conditions on the coefficients, they used the technique of Poisson equation to improve the order to $\frac{1}{2}$. Xu et al. [36] considered strong averaging principle for two-time-scale stochastic McKean-Vlasov equations. Using the variational approach and classical Khasminskii time discretisation, Hong, Li and Liu [12] studied the asymptotic behavior for a class of McKean-Vlasov stochastic partial differential equations with slow and fast time-scales. The main results can be applied to demonstrate the averaging principle for various McKean-Vlasov nonlinear SPDEs.

Although there exist many investigations in the literature devoted to studying stochastic averaging principle for slow and fast time-scales stochastic McKean-Vlasov equations driven by Brownian motion, or by Lévy processes, and so on, as we know, there is not any consideration of averaging principle for slow and fast time-scales stochastic McKean-Vlasov equations driven by fractional Brownian motion. Moreover, due to their distribution dependent nature, they are potentially useful and important for modelling complex systems in diverse areas of applications. Comparing to the classical two-time-scale stochastic McKean-Vlasov equations driven by Brownian motion and Lévy processes, the two-time-scale distribution dependent SDEs driven by fractional Brownian motion are much more complex, therefore, a stochastic averaging principle for such SDEs is naturally interesting and would also be very useful. This motivates us to carry out the present paper, aiming to establish a stochastic averaging principle for the DDSDEs where the fast component is driven by Brownian motion and the slow component is driven by fractional Brownian motion.

Throughout this paper, the letter C will denote a positive constant, with or without subscript, its value may change in different occasions. We will write the dependence of the constant on parameters explicitly if it is essential.

The rest of the paper is organised as follows. In Section 2, we prove the existence and uniqueness of solutions to the two-time-scale SDEs with distribution dependent coefficients driven by fractional Brownian motion. In Section 3, we establish an approximation theorem as an averaging principle for the solutions of the concerned DDSDEs.

2 Existence and uniqueness theorem

In this section, by utilising the Carathéodory approximation technique, we will establish the existence and uniqueness theorem for solutions of fast-slow systems of distribution dependent stochastic differential equations (1.1) driven by fBm and standard Brownian motion under the following Assumption 2.1. It is worthwhile to mention that throughout this section the parameter $\epsilon > 0$ is arbitrarily fixed. We would like to point out that this would not affect the derivation of our averaging principle in the next section.

Assumption 2.1 *There exists a non-decreasing function $K(t), K(0) = 1$ such that for any $t, t_i \in [0,T], p > 0, x_i \in \mathbb{R}^n, y_i \in \mathbb{R}^m, \mu_i \in \mathscr{P}_\theta(\mathbb{R}^n), \nu_i \in \mathscr{P}_\theta(\mathbb{R}^m), i = 1,2.$*

$$\begin{aligned} &|b(t_1,x_1,\mu_1,y_1) - b(t_2,x_2,\mu_2,y_2)|^p \\ &\qquad \le K(|t_1-t_2|^p)[\kappa(|x_1-x_2|^p + |y_1-y_2|^p + \mathbb{W}_\theta(\mu_1,\mu_2)^p)], \end{aligned} \tag{2.1}$$

$$\|\sigma(t,\mu_1) - \sigma(t,\mu_2)\|^p \le K(t^p)\kappa(\mathbb{W}_\theta(\mu_1,\mu_2)^p), \tag{2.2}$$

$$\begin{aligned} &|f(t_1,x_1,\mu_1,y_1) - f(t_2,x_2,\mu_2,y_2)|^p \\ &\qquad \le K(|t_1-t_2|^p)[\kappa(|x_1-x_2|^p + |y_1-y_2|^p + \mathbb{W}_\theta(\mu_1,\mu_2)^p)], \end{aligned} \tag{2.3}$$

$$\begin{aligned}\|g(t_1,x_1,\mu_1,y_1)-g(t_2,x_2,\mu_2,y_2)\|^p \\ \leq K(|t_1-t_2|^p)[\kappa(|x_1-x_2|^p+|y_1-y_2|^p+\mathbb{W}_\theta(\mu_1,\mu_2)^p)],\end{aligned} \tag{2.4}$$

and

$$|b(t,0,\delta_0,0)|^p+\|\sigma(t,\delta_0)\|^p+|f(t,0,\delta_0,0)|^p+\|g(t,0,\delta_0,0)\|^p \leq K(t^p), \tag{2.5}$$

where $\kappa:\mathbb{R}^+\to\mathbb{R}^+$ *is continuous and non-decreasing concave function with* $\kappa(0)=0$, $\kappa(v)>0$, *for every* $v>0$ *such that* $\int_{0+}\frac{1}{\kappa(v)}dv=+\infty$.

Example 2.2 *We can give a few concrete examples of the function* $\kappa(\cdot)$*. Let* $L>0$*, and let* $\delta\in(0,1)$ *be sufficiently small. Define*
$\kappa_1(u)=Lu, u\geq 0.$

$$\kappa_2(u)=\begin{cases} u\log(u^{-1}), & 0\leq u\leq\delta; \\ \delta\log(\delta^{-1})+\kappa'_2(\delta-)(u-\delta), & u>\delta. \end{cases}$$

$$\kappa_3(u)=\begin{cases} u\log(u^{-1})\log\log(u^{-1}), & 0\leq u\leq\delta; \\ \delta\log(\delta^{-1})\log\log(\delta^{-1})+\kappa'_3(\delta-)(u-\delta), & u>\delta, \end{cases}$$

where κ' *denotes the derivative of the function* κ*. They are all concave nondecreasing functions satisfying* $\int_{0+}\frac{du}{\kappa_i(u)}=\infty, i=1,2,3$*. Furthermore, we observed that the Lipschitz condition is a special case of our proposed condition.*

For any $p\geq 1$, let $\mathcal{S}^p([0,T];\mathbb{R}^n)$ be the space of $\mathbb{R}^n$-valued, continuous $(\mathscr{F}_t)_{t\in[0,T]}$-adapted process ψ on [0,T] satisfying

$$\|\psi\|_{\mathcal{S}^p}:=\left(\mathbb{E}\sup_{t\in[0,T]}|\psi_t|^p\right)^{\frac{1}{p}}<\infty.$$

Now, we define the Carathéodory approximation as follows. For any integer $n\geq 1$, define $X_t^{\epsilon,n}=x, Y_t^{\epsilon,n}=y$ for $-1\leq t\leq 0$ and

$$X_t^{\epsilon,n}=x+\int_0^t b(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})ds+\int_0^t\sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})dB^H_s, \tag{2.6}$$

and

$$Y_t^{\epsilon,n}=y+\frac{1}{\epsilon}\int_0^t f(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})ds+\frac{1}{\sqrt{\epsilon}}\int_0^t g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})dW_s. \tag{2.7}$$

for $t\in[0,T]$. It is noted that comparing with Picard's successive approximation technique, the advantage of using Carathéodory approximation technique is that we do not need to compute $X_t^{\epsilon,1},\cdots,X_t^{\epsilon,n-1}$ (res. $Y_t^{\epsilon,1},\cdots,Y_t^{\epsilon,n-1}$) to compute $X_t^{\epsilon,n}$ (res. $Y_t^{\epsilon,n}$). In fact, we can compute $X_t^{\epsilon,n}$ (res. $Y_t^{\epsilon,n}$) directly over intervals of length $\frac{1}{n}$. Our results are new even when the coefficients appeared in Assumption 2.1 satisfy Lipschitz condition. Observe that $\sigma(s,\mathscr{L}_{X_s^\epsilon})$ is a deterministic function, then $\int_0^t\sigma(s,\mathscr{L}_{X_s^\epsilon})dB_s^H$ can be regarded as a Wiener integral with respect to fractional Brownian motion. We need to prepare two lemmas in order to establish the main result in this section.

Lemma 2.3 *Suppose that Assumption 2.1 holds with* $p\geq\theta$ *and* $p>\frac{1}{H}$*. Then*

$$\mathbb{E}\left(\sup_{t\in[0,T]}|X_t^{\epsilon,n}|^p\right)\leq C_{p,\epsilon,T,H,x,y},$$

and

$$\mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}|^p\bigg) \le C_{p,\epsilon,T,H,x,y}.$$

Proof. For any $n \ge 1$. By the elementary inequality

$$|x_1+x_2+x_3|^p \le 3^{p-1}(|x_1|^p+|x_2|^p+|x_3|^p),$$

and Hölder inequality, (2.6)–(2.7), we have

$$\begin{aligned}
\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}|^p\bigg) &\le 3^{p-1}\mathbb{E}|x|^p + 3^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t b(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})ds\bigg|^p\bigg)\\
&\quad + 3^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t \sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})dB_s^H\bigg|^p\bigg)\\
&=: 3^{p-1}\mathbb{E}|x|^p + 3^{p-1}I_{11} + 3^{p-1}I_{12},
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n|^p}\bigg) &\le 3^{p-1}\mathbb{E}|y|^p + (\frac{3}{\epsilon})^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t f(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})ds\bigg|^p\bigg)\\
&\quad + (\frac{3}{\sqrt{\epsilon}})^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})dW_s\bigg|^p\bigg)\\
&=: 3^{p-1}\mathbb{E}|y|^p + (\frac{3}{\epsilon})^{p-1}I_{13} + (\frac{3}{\sqrt{\epsilon}})^{p-1}I_{14}.
\end{aligned}$$

For the term I_{11}, owing to (2.1) in the Assumption 2.1, Hölder inequality, and the fact $\mathbb{W}_\theta(\mathscr{L}_{X_1},\mathscr{L}_{X_2})^p \le \mathbb{E}|X_1-X_2|^p$, we can obtain

$$\begin{aligned}
I_{11} &= \mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t b(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}}) - b(s,0,\delta_0,0) + b(s,0,\delta_0,0)ds\bigg|^p\bigg)\\
&\le C_{p,T}\mathbb{E}\bigg(\sup_{t\in[0,T]}\int_0^t |b(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}}) - b(s,0,\delta_0,0) + b(s,0,\delta_0,0)|^p ds\bigg)\\
&\le C_{p,T}\mathbb{E}\bigg(\sup_{t\in[0,T]}\int_0^t (\kappa(|X^{\epsilon,n}_{s-\frac{1}{n}}|^p + |Y^{\epsilon,n}_{s-\frac{1}{n}}|^p + \mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},\delta_0)^p) + K(T^p))ds\bigg)\\
&\le C_{p,T}\int_0^T (1+\kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}|^p + \mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}|^p))ds\\
&\le C_{p,T}\int_0^T (1+\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}|^p + \mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}|^p)ds,
\end{aligned}$$

where the last inequality is due to $\kappa(\cdot)$ is concave and increasing, there must exist a positive number a such that

$$\kappa(u) \le a(1+u). \tag{2.8}$$

For the term I_{12}, it comes from Alòs and Nualart [1], Fan et al. [5] and (2.2), we have

$$\begin{aligned}
I_{12} &= \mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t \sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})dB_s^H\bigg|^p\bigg)\\
&\le C_{p,T,H}\int_0^T \|\sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})\|^p ds\\
&\le C_{p,T,H}\int_0^T \|\sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})-\sigma(s,\delta_0)+\sigma(s,\delta_0)\|^p ds\\
&\le C_{p,T,H}K(T^p)\int_0^T (\kappa(\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}|^p)+1)ds\\
&\le C_{p,T,H}\int_0^T (1+\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}|^p)ds.
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}|^p\bigg) &\le C\mathbb{E}|x|^p + C_{p,T,H}\int_0^T \bigg(1+\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}|^p\bigg)ds\\
&\le C\mathbb{E}|x|^p + C_{p,T,H}\int_0^T \bigg(1+\mathbb{E}\bigg(\sup_{r\in[0,s]}|X_r^{\epsilon,n}|^p\bigg)+\mathbb{E}\bigg(\sup_{r\in[0,s]}|Y_r^{\epsilon,n}|^p\bigg)\bigg)ds.
\end{aligned} \tag{2.9}$$

For the term I_{13}, it follows from Hölder inequality and (2.3) in Assumption 2.1, we have

$$\begin{aligned}
I_{13} &= \mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t [f(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-f(s,0,\delta_0,0)+f(s,0,\delta_0,0)]ds\bigg|^p\bigg)\\
&\le C_{p,T}\mathbb{E}\bigg(\sup_{t\in[0,T]}\int_0^t |[f(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-f(s,0,\delta_0,0)+f(s,0,\delta_0,0)]|^p ds\bigg)\\
&\le C_{p,T}\mathbb{E}\bigg(\sup_{t\in[0,T]}\int_0^t [\kappa(|X^{\epsilon,n}_{s-\frac{1}{n}}|^p+|Y^{\epsilon,n}_{s-\frac{1}{n}}|^p+\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},\delta_0)^p)+K(T^p)]ds\bigg)\\
&\le C_{p,T}\int_0^T \bigg(1+\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}|^p\bigg)ds.
\end{aligned}$$

For the term I_{14}, it follows from the Burkholder-Davis-Gundy inequality, Hölder inequality and (2.4) in Assumption 2.1, we have

$$\begin{aligned}
I_{14} &= \mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})dW_s\bigg|^p\bigg)\\
&\le C_{p,T}\bigg|\mathbb{E}\int_0^T \|g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})\|^2 ds\bigg|^{\frac{p}{2}}\\
&\le C_{p,T}\bigg|\mathbb{E}\int_0^T \|g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-g(s,0,\delta_0,0)+g(s,0,\delta_0,0)\|^2 ds\bigg|^{\frac{p}{2}}\\
&\le C_{p,T}\mathbb{E}\int_0^T (\kappa(|X^{\epsilon,n}_{s-\frac{1}{n}}|^p+|Y^{\epsilon,n}_{s-\frac{1}{n}}|^p+\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},\delta_0)^p)+K(T^p))ds\\
&\le C_{p,T}\int_0^T (1+\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}|^p)ds.
\end{aligned}$$

Thus, we can get

$$\begin{aligned}\mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}|^p\bigg) &\le C\mathbb{E}|y|^p + C_{p,\epsilon,T}\int_0^T (1+\mathbb{E}|X_{s-\frac{1}{n}}^{\epsilon,n}|^p + \mathbb{E}|Y_{s-\frac{1}{n}}^{\epsilon,n}|^p)ds \\ &\le C\mathbb{E}|y|^p + C_{p,\epsilon,T}\int_0^T \bigg(1+\mathbb{E}\bigg(\sup_{r\in[0,s]}|X_r^{\epsilon,n}|^p\bigg) + \mathbb{E}\bigg(\sup_{r\in[0,s]}|Y_r^{\epsilon,n}|^p\bigg)\bigg)ds.\end{aligned} \tag{2.10}$$

Combining (2.9) and (2.10), we have

$$\begin{aligned}&\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}|^p\bigg) + \mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}|^p\bigg) \\ &\le C_{p,T,x,y} + C_{p,\epsilon,T}\int_0^T \bigg(\mathbb{E}\bigg(\sup_{r\in[0,s]}|X_r^{\epsilon,n}|^p\bigg) + \mathbb{E}\bigg(\sup_{r\in[0,s]}|Y_r^{\epsilon,n}|^p\bigg)\bigg)ds.\end{aligned}$$

It follows from the Gronwall's inequality that

$$\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}|^p\bigg) + \mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}|^p\bigg) \le C_{p,\epsilon,T,H,x,y}. \tag{2.11}$$

Therefore, this shows the boundedness of $X_t^{\epsilon,n}$ and $Y_t^{\epsilon,n}$. ■

Lemma 2.4 *Suppose that Assumption 2.1 holds with $p \ge \theta$ and $p > \frac{1}{H}$. Then*

$$\mathbb{E}(|X_t^{\epsilon,n} - X_s^{\epsilon,n}|^p) \le C_{p,\epsilon,T,H,x,y}.[(t-s)^p + (t-s)^{pH}], \tag{2.12}$$

and

$$\mathbb{E}(|Y_t^{\epsilon,n} - Y_s^{\epsilon,n}|^p) \le C_{p,\epsilon,T,H,x,y}.[(t-s)^p + (t-s)^{\frac{p}{2}}]. \tag{2.13}$$

Proof. By (2.6) and (2.7), we have

$$X_t^{\epsilon,n} - X_s^{\epsilon,n} = \int_s^t b(r, X_{r-\frac{1}{n}}^{\epsilon,n}, \mathscr{L}_{X_{r-\frac{1}{n}}^{\epsilon,n}}, Y_{r-\frac{1}{n}}^{\epsilon,n})dr + \int_s^t \sigma(r, \mathscr{L}_{X_{r-\frac{1}{n}}^{\epsilon,n}})dB_r^H,$$

and

$$\begin{aligned}&Y_t^{\epsilon,n} - Y_s^{\epsilon,n} \\ &= \frac{1}{\epsilon}\int_s^t f(r, X_{r-\frac{1}{n}}^{\epsilon,n}, \mathscr{L}_{X_{r-\frac{1}{n}}^{\epsilon,n}}, Y_{r-\frac{1}{n}}^{\epsilon,n})dr + \frac{1}{\sqrt{\epsilon}}\int_s^t g(r, X_{r-\frac{1}{n}}^{\epsilon,n}, \mathscr{L}_{X_{r-\frac{1}{n}}^{\epsilon,n}}, Y_{r-\frac{1}{n}}^{\epsilon,n})dW_r.\end{aligned}$$

By the elementary inequality $|x_1+x_2|^p \leq 2^{p-1}(|x_1|^p+|x_2|^p)$ and Hölder inequality, we have

$$\begin{aligned}
&\mathbb{E}|X_t^{\epsilon,n}-X_s^{\epsilon,n}|^p\\
&\leq 2^{p-1}\mathbb{E}\bigg|\int_s^t b(r,X^{\epsilon,n}_{r-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}},Y^{\epsilon,n}_{r-\frac{1}{n}})dr\bigg|^p+2^{p-1}\mathbb{E}\bigg|\int_s^t \sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})dB^H_r\bigg|^p\\
&\leq 2^{p-1}(t-s)^{p-1}\mathbb{E}\int_s^t \big|b(r,X^{\epsilon,n}_{r-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}},Y^{\epsilon,n}_{r-\frac{1}{n}})\big|^p dr+2^{p-1}\mathbb{E}\bigg|\int_s^t \sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})dB^H_r\bigg|^p\\
&\leq 2^{p-1}(t-s)^{p-1}\mathbb{E}\int_s^t |b(r,X^{\epsilon,n}_{r-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}},Y^{\epsilon,n}_{r-\frac{1}{n}})-b(r,0,\delta_0,0)+b(r,0,\delta_0,0)|^p dr\\
&\quad+2^{p-1}\mathbb{E}\bigg|\int_s^t \sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})dB^H_r\bigg|^p\\
&\leq 2^{p-1}(t-s)^{p-1}\mathbb{E}\int_s^t (\kappa(|X^{\epsilon,n}_{r-\frac{1}{n}}|^p+|Y^{\epsilon,n}_{r-\frac{1}{n}}|^p+\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}},\delta_0)^p)+K(T^p))dr\\
&\quad+2^{p-1}\mathbb{E}\bigg|\int_s^t \sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})dB^H_r\bigg|^p\\
&\leq C_p(t-s)^{p-1}\int_s^t (1+\mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{r-\frac{1}{n}}|^p)dr+2^{p-1}\mathbb{E}\bigg|\int_s^t \sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})dB^H_r\bigg|^p.
\end{aligned}$$

As for the term $\mathbb{E}\big|\int_s^t \sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})dB^H_r\big|^p$, using the methods from Alòs and Nualart [1] (see also Fan et al. [5], Shen et al. [29]), we have

$$\begin{aligned}
\mathbb{E}\bigg|\int_s^t \sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})dB^H_r\bigg|^p &\leq C_{p,H}(t-s)^{pH-1}\int_s^t \|\sigma(r,\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}})\|^p dr\\
&\leq C_{p,H}(t-s)^{pH-1}\int_s^t K(T^p)(\kappa(\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}},\delta_0)^p)+1)dr\\
&\leq C_{p,H}(t-s)^{pH-1}\int_s^t (1+\mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p)dr.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}|Y_t^{\epsilon,n}-Y_s^{\epsilon,n}|^p\\
&\leq (\frac{2}{\epsilon})^{p-1}\mathbb{E}\bigg|\int_s^t f(r,X^{\epsilon,n}_{r-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}},Y^{\epsilon,n}_{r-\frac{1}{n}})dr\bigg|^p\\
&\quad+(\frac{2}{\sqrt{\epsilon}})^{p-1}\mathbb{E}\bigg|\int_s^t g(r,X^{\epsilon,n}_{r-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}},Y^{\epsilon,n}_{r-\frac{1}{n}})dW_r\bigg|^p\\
&=:(\frac{2}{\epsilon})^{p-1}J_{21}+(\frac{2}{\sqrt{\epsilon}})^{p-1}J_{22}.
\end{aligned}$$

For the term J_{21}, by the Hölder inequality, we have

$$\begin{aligned} J_{21} &= \mathbb{E}\bigg|\int_s^t f(s, X^{\epsilon,n}_{r-\frac{1}{n}}, \mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}}, Y^{\epsilon,n}_{r-\frac{1}{n}})dr\bigg|^p \\ &\leq (t-s)^{p-1}\mathbb{E}\int_s^t |f(s, X^{\epsilon,n}_{r-\frac{1}{n}}, \mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}}, Y^{\epsilon,n}_{r-\frac{1}{n}})|^p dr \\ &\leq (t-s)^{p-1}\mathbb{E}\int_s^t |[f(s, X^{\epsilon,n}_{r-\frac{1}{n}}, \mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}}, Y^{\epsilon,n}_{r-\frac{1}{n}}) - f(s,0,\delta_0,0) + f(s,0,\delta_0,0)]|^p dr \\ &\leq (t-s)^{p-1}\mathbb{E}\int_s^t (\kappa(|X^{\epsilon,n}_{r-\frac{1}{n}}|^p + |Y^{\epsilon,n}_{r-\frac{1}{n}}|^p + \mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}}, \delta_0)^p) + K(T^p))dr \\ &\leq C_{p,T}(t-s)^{p-1}\int_s^t (\mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p + \mathbb{E}|Y^{\epsilon,n}_{r-\frac{1}{n}}|^p + 1)dr. \end{aligned}$$

For the term J_{22}, by the Burkholder-Davis-Gundy inequality, we have

$$\begin{aligned} J_{22} &\leq C_p\mathbb{E}\bigg[\int_s^t \|g(s, X^{\epsilon,n}_{r-\frac{1}{n}}, \mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}}, Y^{\epsilon,n}_{r-\frac{1}{n}})\|^2 dr\bigg]^{\frac{p}{2}} \\ &\leq C_p(t-s)^{\frac{p}{2}-1}\mathbb{E}\int_s^t \|g(s, X^{\epsilon,n}_{r-\frac{1}{n}}, \mathscr{L}_{X^{\epsilon,n}_{r-\frac{1}{n}}}, Y^{\epsilon,n}_{r-\frac{1}{n}})\|^p dr \\ &\leq C_{p,T}(t-s)^{\frac{p}{2}-1}\int_s^t (\mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p + \mathbb{E}|Y^{\epsilon,n}_{r-\frac{1}{n}}|^p + 1)dr. \end{aligned}$$

Above all, by the Lemma 2.3 we can get that

$$\begin{aligned} &\mathbb{E}|X^{\epsilon,n}_t - X^{\epsilon,n}_s|^p \\ &\leq C_p(t-s)^{p-1}\int_s^t (1 + \mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p + \mathbb{E}|Y^{\epsilon,n}_{r-\frac{1}{n}}|^p)dr + C_{p,H}(t-s)^{pH-1}\int_s^t (1 + \mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p)dr \\ &\leq C_{p,\epsilon,T,H,x,y}.[(t-s)^p + (t-s)^{pH}], \end{aligned}$$

and

$$\begin{aligned} &\mathbb{E}|Y^{\epsilon,n}_t - Y^{\epsilon,n}_s|^p \\ &\leq C_{p,\epsilon,T}(t-s)^{p-1}\int_s^t (\mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p + \mathbb{E}|Y^{\epsilon,n}_{r-\frac{1}{n}}|^p + 1)dr \\ &\quad + C_{p,\epsilon,T}(t-s)^{\frac{p}{2}-1}\int_s^t (\mathbb{E}|X^{\epsilon,n}_{r-\frac{1}{n}}|^p + \mathbb{E}|Y^{\epsilon,n}_{r-\frac{1}{n}}|^p + 1)dr \\ &\leq C_{p,\epsilon,T,H,x,y}.[(t-s)^p + (t-s)^{\frac{p}{2}}]. \end{aligned}$$

This completes the proof. ∎

Theorem 2.5 *Suppose that Assumption 2.1 holds. For any $\epsilon > 0$, $p \geq \theta$ and $p > \frac{1}{H}$, there exists a unique solution $(X^\epsilon_t, Y^\epsilon_t), t \geq 0$ to system (1.1) with initial value $x \in \mathbb{R}^n, y \in \mathbb{R}^m$ and for all $T > 0, (X^\epsilon, Y^\epsilon) \in \mathcal{S}^p([0,T];\mathbb{R}^n) \times \mathcal{S}^p([0,T];\mathbb{R}^m), \mathbb{P} - a.s$ and*

$$\begin{cases} X^\epsilon_t = x + \int_0^t b(s, X^\epsilon_s, \mathscr{L}_{X^\epsilon_s}, Y^\epsilon_s)ds + \int_0^t \sigma(s, \mathscr{L}_{X^\epsilon_s})dB^H_s, \\ Y^\epsilon_t = y + \frac{1}{\epsilon}\int_0^t f(s, X^\epsilon_s, \mathscr{L}_{X^\epsilon_s}, Y^\epsilon_s)ds + \frac{1}{\sqrt{\epsilon}}\int_0^t g(s, X^\epsilon_s, \mathscr{L}_{X^\epsilon_s}, Y^\epsilon_s)dW_s. \end{cases} \tag{2.14}$$

Proof. We split the proof in two steps.
Step one: Existence. We firstly prove that $(X_t^{\epsilon,n})_{n\geq 1}$, $(Y_t^{\epsilon,n})_{n\geq 1}$ are Cauchy sequences in $\mathcal{S}^p([0,T])$.

Note that for $n > m \geq 1$, it is routine to obtain

$$
\begin{aligned}
&\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}-X_t^{\epsilon,m}|^p\bigg)\\
&\leq 2^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t[(b(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-b(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})]ds\bigg|^p\bigg)\\
&\quad+2^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t[\sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})-\sigma(s,\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}})]dB^H_s\bigg|^p\bigg)\\
&:=2^{p-1}II_{11}+2^{p-1}II_{12},
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}-Y_t^{\epsilon,m}|^p\bigg)\\
&\leq(\frac{2}{\epsilon})^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t[f(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-f(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})]ds\bigg|^p\bigg)\\
&\quad+(\frac{2}{\sqrt{\epsilon}})^{p-1}\mathbb{E}\bigg(\sup_{t\in[0,T]}\bigg|\int_0^t[g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-g(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})]dW_s\bigg|^p\bigg)\\
&:=(\frac{2}{\epsilon})^{p-1}II_{21}+(\frac{2}{\sqrt{\epsilon}})^{p-1}II_{22}.
\end{aligned}
$$

For the term II_{11}. By the Assumption 2.1, Hölder inequality and Lemma 2.4, we have

$$
\begin{aligned}
II_{11}&\leq C_{T,p}\mathbb{E}\bigg(\sup_{t\in[0,T]}\int_0^t\bigg|[(b(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-b(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})]\bigg|^p ds\bigg)\\
&\leq C_{T,p}\mathbb{E}\int_0^T[|(b(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-b(s,X^{\epsilon,m}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},Y^{\epsilon,m}_{s-\frac{1}{n}})|^p\\
&\qquad+|b(s,X^{\epsilon,m}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},Y^{\epsilon,m}_{s-\frac{1}{n}})-b(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})|^p]ds\\
&\leq C_{T,p}\mathbb{E}\int_0^T[\kappa(|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}})^p)\\
&\qquad+\kappa(|X^{\epsilon,m}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{m}}|^p+|Y^{\epsilon,m}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{m}}|^p)+\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}})^p)]ds\\
&\leq C_{T,p}\int_0^T[\kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p)\\
&\qquad+\kappa(2\mathbb{E}|X^{\epsilon,m}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{m}}|^p+\mathbb{E}|Y^{\epsilon,m}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{m}}|^p)]ds\\
&\leq C_{T,p}\int_0^T\kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p)ds\\
&\qquad+C_{T,p}\int_0^T\kappa\bigg(C_{T,H,p,\epsilon,|x|,|y|}[(\frac{1}{m}-\frac{1}{n})^p+(\frac{1}{m}-\frac{1}{n})^{pH}+(\frac{1}{m}-\frac{1}{n})^{\frac{p}{2}}]\bigg)ds.
\end{aligned}
$$

For the term II_{12}, by Assumption 2.1, Hölder inequality and Lemma 2.4, we have

$$\begin{aligned}
II_{12} &\leq C_{T,p,H}\mathbb{E}\int_0^T \|[\sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})-\sigma(s,\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}})]\|^p ds\\
&\leq C_{T,p,H}\mathbb{E}\int_0^T [\|\sigma(s,\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}})-\sigma(s,\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}})\|^p+\|\sigma(s,\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}})-\sigma(s,\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}})\|^p]ds\\
&\leq C_{T,p,H}\mathbb{E}\int_0^T K(T)[\kappa(\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}})^p)+\kappa(\mathbb{W}_\theta(\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}})^p)]ds\\
&\leq C_{T,p,H}\int_0^T \kappa(\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p)ds\\
&\quad + C_{T,p,H}\int_0^T \kappa\bigg(C_{T,H,p,\epsilon,|x|,|y|}[(\frac{1}{m}-\frac{1}{n})^p+(\frac{1}{m}-\frac{1}{n})^{pH}]\bigg)ds.
\end{aligned}$$

Using the same way of the proof of II_{11}, we have

$$\begin{aligned}
II_{21} &\leq C_{T,p}\mathbb{E}\int_0^T |[f(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-f(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})]|^p ds\\
&\leq C_{T,p}\mathbb{E}\int_0^T [|f(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-f(s,X^{\epsilon,m}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},Y^{\epsilon,m}_{s-\frac{1}{n}})|^p\\
&\qquad +|f(s,X^{\epsilon,m}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},Y^{\epsilon,m}_{s-\frac{1}{n}})-f(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})|^p]ds\\
&\leq C_{T,p}\int_0^T \kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p)\\
&\qquad +\kappa(2\mathbb{E}|X^{\epsilon,m}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{m}}|^p+\mathbb{E}|Y^{\epsilon,m}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{m}}|^p)ds\\
&\leq C_{T,p}\int_0^T \kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p)\\
&\quad + C_{T,p}\int_0^T \kappa\bigg(C_{T,H,p,\epsilon,|x|,|y|}[(\frac{1}{m}-\frac{1}{n})^p+(\frac{1}{m}-\frac{1}{n})^{pH}+(\frac{1}{m}-\frac{1}{n})^{\frac{p}{2}}]\bigg)ds.
\end{aligned}$$

For the term II_{22}, by the Assumption 2.1, Burkholder-Davis-Gundy inequality and Lemma 2.4, we have

$$\begin{aligned}
II_{22} &\leq C_p\mathbb{E}\bigg(\int_0^T \|g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-g(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})\|^2 ds\bigg)^{\frac{p}{2}}\\
&\leq C_{T,p}\mathbb{E}\int_0^T \|g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-g(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})\|^p ds\\
&\leq C_{T,p}\mathbb{E}\int_0^T [\|g(s,X^{\epsilon,n}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,n}_{s-\frac{1}{n}}},Y^{\epsilon,n}_{s-\frac{1}{n}})-g(s,X^{\epsilon,m}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},Y^{\epsilon,m}_{s-\frac{1}{n}})\|^p\\
&\qquad +\|g(s,X^{\epsilon,m}_{s-\frac{1}{n}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{n}}},Y^{\epsilon,m}_{s-\frac{1}{n}})-g(s,X^{\epsilon,m}_{s-\frac{1}{m}},\mathscr{L}_{X^{\epsilon,m}_{s-\frac{1}{m}}},Y^{\epsilon,m}_{s-\frac{1}{m}})\|^p]ds\\
&\leq C_{T,p}\int_0^T \kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p)ds\\
&\quad + C_{T,p}\int_0^T \kappa\bigg(C_{T,H,p,\epsilon,|x|,|y|}[(\frac{1}{m}-\frac{1}{n})^p+(\frac{1}{m}-\frac{1}{n})^{pH}+(\frac{1}{m}-\frac{1}{n})^{\frac{p}{2}}]\bigg)ds.
\end{aligned}$$

Above all, we can conclude

$$
\begin{aligned}
&\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}-X_t^{\epsilon,m}|^p\bigg)\\
&\le C_{T,p,H}\int_0^T\kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p)ds\\
&\quad+C_{T,p}\int_0^T\kappa\bigg(C_{T,H,p,\epsilon,|x|,|y|}[(\frac{1}{m}-\frac{1}{n})^p+(\frac{1}{m}-\frac{1}{n})^{pH}+(\frac{1}{m}-\frac{1}{n})^{\frac{p}{2}}]\bigg)ds,
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}-Y_t^{\epsilon,m}|^p\bigg)\\
&\le C_{T,p,\epsilon}\int_0^T\kappa(2\mathbb{E}|X^{\epsilon,n}_{s-\frac{1}{n}}-X^{\epsilon,m}_{s-\frac{1}{n}}|^p+\mathbb{E}|Y^{\epsilon,n}_{s-\frac{1}{n}}-Y^{\epsilon,m}_{s-\frac{1}{n}}|^p)ds\\
&\quad+C_{T,p,\epsilon}\int_0^T\kappa\bigg(C_{T,H,p,\epsilon,|x|,|y|}[(\frac{1}{m}-\frac{1}{n})^p+(\frac{1}{m}-\frac{1}{n})^{pH}+(\frac{1}{m}-\frac{1}{n})^{\frac{p}{2}}]\bigg)ds.
\end{aligned}
$$

Using the fact that $\kappa(0)=0$ and Fatou's Lemma, we obtain for every $\delta>0$,

$$
\begin{aligned}
Z(t)&\le C_{T,p,\epsilon,H}\int_0^T\kappa(2Z(s))ds\\
&\le \delta+C_{T,p,\epsilon,H}\int_0^T\kappa(2Z(s))ds.
\end{aligned}
$$

where

$$
Z(T)=\limsup_{n,m\to\infty}\bigg[\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}-X_t^{\epsilon,m}|^p\bigg)+\mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}-Y_t^{\epsilon,m}|^p\bigg)\bigg].
$$

Hence, the Bihari inequality yields

$$
Z(T)\le\frac{1}{2}G^{-1}\bigg[G(2\delta)+C_{T,p,\epsilon,H}\bigg].
$$

where $G(2\delta)+C_{T,p,H}\in Dom(G^{-1})$, G^{-1} is the inverse function of $G(\cdot)$ and

$$
G(v)=\int_1^v\frac{ds}{\kappa(s)},\quad v>0.
$$

By Assumption 2.1, one sees that $\lim_{\delta\downarrow 0}G(\delta)=-\infty$ and $Dom(G^{-1})=(-\infty,G(\infty))$. Letting $\delta\to 0$ gives

$$
\mathbb{E}\bigg(\sup_{t\in[0,T]}|X_t^{\epsilon,n}-X_t^{\epsilon,m}|^p\bigg)+\mathbb{E}\bigg(\sup_{t\in[0,T]}|Y_t^{\epsilon,n}-Y_t^{\epsilon,m}|^p\bigg)=0.
$$

Consequently, $(X_t^{\epsilon,n})_{n\ge1}$, $(Y_t^{\epsilon,n})_{n\ge1}$ are Cauchy sequences in $\mathcal{S}^p([0,T])$ with $p\ge\theta$ and $p>\frac{1}{H}$, and then the limit, denoted by $X_t^\epsilon, Y_t^\epsilon$ is a solution to (1.1).

Step two: Uniqueness. Let $(X_t^\epsilon, Y_t^\epsilon)$, $(\tilde{X}_t^\epsilon, \tilde{Y}_t^\epsilon)$ be two solutions for (1.1) on the same probability space with the same initial value, note that

$$X_t^\epsilon - \tilde{X}_t^\epsilon = \int_0^t [b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - b(s, \tilde{X}_s^\epsilon, \mathscr{L}_{\tilde{X}_s^\epsilon}, \tilde{Y}_s^\epsilon)]ds + \int_0^t [\sigma(s, \mathscr{L}_{X_s^\epsilon}) - \sigma(s, \mathscr{L}_{\tilde{X}_s^\epsilon})]dB_s^H,$$

and

$$\begin{aligned} Y_t^\epsilon - \tilde{Y}_t^\epsilon =& (\frac{1}{\epsilon}) \int_0^t [f(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - f(s, \tilde{X}_s^\epsilon, \mathscr{L}_{\tilde{X}_s^\epsilon}, \tilde{Y}_s^\epsilon)]ds \\ &+ (\frac{1}{\sqrt{\epsilon}}) \int_0^t [g(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - g(s, \tilde{X}_s^\epsilon, \mathscr{L}_{\tilde{X}_s^\epsilon}, \tilde{Y}_s^\epsilon)]dW_s. \end{aligned}$$

By the same ways of the proof of II_{11}, II_{12}, we have

$$\begin{aligned} \mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\epsilon - \tilde{X}_t^\epsilon|^p\right) \leq& 2^{p-1}\mathbb{E}\left(\sup_{t\in[0,T]} \left| \int_0^t [b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - b(s, \tilde{X}_s^\epsilon, \mathscr{L}_{\tilde{X}_s^\epsilon}, \tilde{Y}_s^\epsilon)]ds \right|^p\right) \\ &+ 2^{p-1}\mathbb{E}\left(\sup_{t\in[0,T]} \left| \int_0^t [\sigma(s, \mathscr{L}_{X_s^\epsilon}) - \sigma(s, \mathscr{L}_{\tilde{X}_s^\epsilon})]dB_s^H \right|^p\right) \\ \leq& C_{T,p,H} \int_0^T \kappa(2\mathbb{E}|X_s^\epsilon - \tilde{X}_s^\epsilon|^p + \mathbb{E}|Y_s^\epsilon - \tilde{Y}_s^\epsilon|^p)ds. \end{aligned}$$

By the same ways of the proof of II_{21}, II_{22}, we have

$$\begin{aligned} \mathbb{E}\left(\sup_{t\in[0,T]} |Y_t^\epsilon - \tilde{Y}_t^\epsilon|^p\right) \leq& \mathbb{E}\Bigg[(\frac{1}{\epsilon}) \int_0^T |f(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - f(s, \tilde{X}_s^\epsilon, \mathscr{L}_{\tilde{X}_s^\epsilon}, \tilde{Y}_s^\epsilon)|ds \\ &+ (\frac{1}{\sqrt{\epsilon}}) \int_0^T \|g(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - g(s, \tilde{X}_s^\epsilon, \mathscr{L}_{\tilde{X}_s^\epsilon}, \tilde{Y}_s^\epsilon)\|dW_s\Bigg]^p \\ \leq& C_{T,p,\epsilon,H} \int_0^T \kappa(2\mathbb{E}|X_s^\epsilon - \tilde{X}_s^\epsilon|^p + \mathbb{E}|Y_s^\epsilon - \tilde{Y}_s^\epsilon|^p)ds. \end{aligned}$$

Hence, we have

$$\begin{aligned} &\mathbb{E}\left(\sup_{t\in[0,T]} |X_t^\epsilon - \tilde{X}_t^\epsilon|^p\right) + \mathbb{E}\left(\sup_{t\in[0,T]} |Y_t^\epsilon - \tilde{Y}_t^\epsilon|^p\right) \\ &\leq C_{T,p,\epsilon,H} \int_0^T \kappa(2\mathbb{E}|X_s^\epsilon - \tilde{X}_s^\epsilon|^p + 2\mathbb{E}|Y_s^\epsilon - \tilde{Y}_s^\epsilon|^p)ds. \end{aligned}$$

Then, the Bihari inequality implies that $(X_t^\epsilon, Y_t^\epsilon) = (\tilde{X}_t^\epsilon, \tilde{Y}_t^\epsilon)$, $t \in [0, T]$, $\mathbb{P} - a.s.$ This completes the proof. ■

3 Averaging principle

In this section, assume $\theta = 2$, we will establish the convergence, under non-Lipschitz coefficients, of X_t^ϵ, the slow component in (1.1), to $\bar{X}_t$, the solution of the averaged equations (3.3) in the sense of convergence in mean square. We need the following conditions on the coefficients f, g.

Assumption 3.1 *There exist constants $\beta_i > 0, i = 1, 2$, such that the following hold*

$$2\langle y_1 - y_2, f(t_1, x_1, \mu_1, y_1) - f(t_2, x_2, \mu_2, y_2)\rangle + \|g(t_1, x_1, \mu_1, y_1) - g(t_2, x_2, \mu_2, y_2)\|^2 \\ \leq -\beta_1|y_1 - y_2|^2 + K(|t_1 - t_2|^2)\kappa(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2), \tag{3.1}$$

and

$$2\langle y, f(t, x, \mu, y)\rangle + \|g(t, x, \mu, y)\|^2 \leq -\beta_2|y|^2 + C(1 + |x|^2 + \mu(|\cdot|^2)). \tag{3.2}$$

Next, we state our main result.

Theorem 3.2 *Suppose that Assumptions 2.1 and 3.1 hold, for any $T > 0$, $t \in [0, T]$, we have*

$$\lim_{\epsilon\to 0} \mathbb{E}|X^\epsilon - \bar{X}|^2 = 0,$$

where $\bar{X}$ is the solition of the following averaged equation,

$$\begin{cases} d\bar{X}_t = \bar{b}(t, \bar{X}_t, \mathscr{L}_{\bar{X}_t})dt + \sigma(t, \mathscr{L}_{\bar{X}_t})dB_t^H, \\ \bar{X}_0 = x. \end{cases} \tag{3.3}$$

here

$$\bar{b}(t, x, \mu) = \int_{\mathbb{R}^m} b(t, x, \mu, z)\nu^{t,x,\mu}(dz),$$

and $\nu^{t,x,\mu}$ is the unique invariant measure for the transition semigroup of the following frozen equation,

$$\begin{cases} dY_s = f(t, x, \mu, Y_s)ds + g(t, x, \mu, Y_s)d\tilde{W}_s, \\ Y_0 = y, \end{cases} \tag{3.4}$$

where $\tilde{W}_t$ is a d_2-dimensional Brownian motion on another complete probability space $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$ and $\{\tilde{\mathscr{F}}_t, t \geq 0\}$ is natural filtration generated by $\tilde{W}_t$. It is easy to prove for any initial data $y \in \mathbb{R}^m$ that equation (3.4) has a unique strong solution $\{Y_s^{t,x,\mu,y}\}_{s\geq 0}$, which is a homogeneous Markov process. Moreover, $\sup_{s\geq 0} \tilde{\mathbb{E}}|Y_s^{t,x,\mu,y}|^2 \leq C_T[1 + |x|^2 + |y|^2 + \mu(|\cdot|^2)]$, for any $t \in [0, T]$, where $\tilde{\mathbb{E}}$ is the expectation on $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}})$.

Noting that for any bounded measurable function $\varphi : \mathbb{R}^m \to \mathbb{R}$, we have

$$P_s^{t,x,\mu}\varphi(y) := \tilde{\mathbb{E}}\varphi(Y_s^{t,x,\mu,y}), \quad y \in \mathbb{R}^m, s \geq 0,$$

where $\{P_s^{t,x,\mu}\}_{s\geq 0}$ be the transition semigroup of $Y_s^{t,x,\mu,y}$.

Lemma 3.3 *Suppose that Assumptions 2.1 and 3.1 hold. Then for any $T > 0$, $t_1, t_2 \in [0, T]$, we have*

$$\tilde{\mathbb{E}}[|Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2}|^2] \le e^{-\beta_1 s}|y_1 - y_2|^2 + C_T(K(|t_1 - t_2|^2)[\kappa(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2)].$$

Proof.

$$\begin{aligned} Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2} =& y_1 - y_2 + \int_0^s [f(t_1, x_1, \mu_1, Y_r^{t_1,x_1,\mu_1,y_1}) - f(t_2, x_2, \mu_2, Y_r^{t_2,x_2,\mu_2,y_2})]dr \\ &+ \int_0^s [g(t_1, x_1, \mu_1, Y_r^{t_1,x_1,\mu_1,y_1}) - g(t_2, x_2, \mu_2, Y_r^{t_2,x_2,\mu_2,y_2})]d\tilde{W}_r. \end{aligned}$$

By Itô's formula, we have

$$\begin{aligned} &\tilde{\mathbb{E}}|Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2}|^2 \\ &= \tilde{\mathbb{E}} \int_0^s [2\langle f(t_1, x_1, \mu_1, Y_r^{t_1,x_1,\mu_1,y_1}) - f(t_2, x_2, \mu_2, Y_r^{t_2,x_2,\mu_2,y_2}), Y_r^{t_1,x_1,\mu_1,y_1} - Y_r^{t_2,x_2,\mu_2,y_2}\rangle \\ &\qquad + ||g(t_1, x_1, \mu_1, Y_r^{t_1,x_1,\mu_1,y_1}) - g(t_2, x_2, \mu_2, Y_r^{t_2,x_2,\mu_2,y_2})||^2]dr. \end{aligned}$$

By Assumptions 3.1, we have

$$\begin{aligned} &\frac{d}{ds}\tilde{\mathbb{E}}|Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2}|^2 \\ &\quad \le -\beta_1(\tilde{\mathbb{E}}|Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2}|^2) + C_T K(|t_1 - t_2|^2)[\kappa(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2)]. \end{aligned}$$

Hence, by Gronwall's inequality([7].pp.584), we obtain

$$\tilde{\mathbb{E}}|Y_s^{t_1,x_1,\mu_1,y_1} - Y_s^{t_2,x_2,\mu_2,y_2}|^2 \le e^{-\beta_1 s}|y_1 - y_2|^2 + C_T K(|t_1 - t_2|^2)[\kappa(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2)].$$

This completes the proof. ∎

Remark 3.1 *Under Assumptions 2.1 and 3.1, the averaged equation (3.3) has a unique solution $\{\bar{X}_t, t \ge 0\}$.*

In fact, for any $x_1, x_2 \in \mathbb{R}^n, t_1, t_2 \in [0, T]$ and any initial value $y \in \mathbb{R}^m$, by Assumption 2.1, Lemmas 3.3 and

$$\begin{aligned} &|\bar{b}(t_1, x_1, \mu_1) - \bar{b}(t_2, x_2, \mu_2)|^2 \\ &= \left| \int_{\mathbb{R}^m} b(t_1, x_1, \mu_1, y)\nu^{t_1,x_1,\mu_1}(dy) - \int_{\mathbb{R}^m} b(t_2, x_2, \mu_2, y)\nu^{t_2,x_2,\mu_2}(dy) \right|^2 \\ &= \lim_{s\to\infty} |\tilde{\mathbb{E}}b(t_1, x_1, \mu_1, Y_s^{t_1,x_1,\mu_1,y}) - \tilde{\mathbb{E}}b(t_2, x_2, \mu_2, Y_s^{t_2,x_2,\mu_2,y})|^2 \\ &\le \lim_{s\to\infty} \tilde{\mathbb{E}}|b(t_1, x_1, \mu_1, Y_s^{t_1,x_1,\mu_1,y}) - b(t_2, x_2, \mu_2, Y_s^{t_2,x_2,\mu_2,y})|^2 \\ &\le K(|t_1 - t_2|^2)[\kappa(|x_1 - x_2|^2 + \tilde{\mathbb{E}}|Y_s^{t_1,x_1,\mu_1,y} - Y_s^{t_2,x_2,\mu_2,y}| + \mathbb{W}_2(\mu_1, \mu_2)^2)] \\ &\le K(|t_1 - t_2|^2)[\kappa(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2 + C_T\kappa(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2))] \\ &\le K(|t_1 - t_2|^2)[\kappa_1(|x_1 - x_2|^2 + \mathbb{W}_2(\mu_1, \mu_2)^2)]. \end{aligned} \tag{3.5}$$

Next, notice that $\kappa : \mathbb{R}^+ \to \mathbb{R}^+$ is a continuous and non-decreasing concave function fulfilling $\kappa(0) = 0, \kappa(v) > 0$, and $\int_{0+} \frac{1}{\kappa(v)}dv = +\infty$ for each $v > 0$, we then let $\kappa_1(v) := \kappa(v + C_T\kappa(v))$, then one can see that $\kappa_1 : \mathbb{R}^+ \to \mathbb{R}^+$ is also a continuous and non-decreasing concave function satisfying $\kappa_1(0) = 0, \kappa_1(v) > 0$ and $\int_{0+} \frac{1}{\kappa_1(v)}dv = +\infty$. Consequently, this indicates that $\bar{b}$ satisfies (2.1). As we have already showed that there is a unique solution to (1.1), by the same argument, we conclude that there is a unique solution to Equation (3.3).

3.1 Some estimates for the solution $(X_t^\epsilon, Y_t^\epsilon)$

Firstly, we prove some uniform bounds for the solution $(X_t^\epsilon, Y_t^\epsilon)$

Lemma 3.4 *Suppose that Assumptions 2.1 and 3.1 hold. Then, for any $T > 0$, $t \in [0, T]$, we have*

$$\sup_{\epsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}|X_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_2},$$

and

$$\sup_{\epsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}|Y_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_2},$$

where the positive constant $C_{T,H,|x|,|y|.\beta_2}$ with subscripts dependents on $T, H, |x|, |y|, \beta_2$.

Proof. For any $t \in [0, T]$, it follows from Hölder inequality and Assumption 2.1, we get

$$\begin{aligned}
\mathbb{E}|X_t^\epsilon|^2 &= \mathbb{E}\bigg|x + \int_0^t b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)ds + \int_0^t \sigma(s, \mathscr{L}_{X_s^\epsilon})dB_s^H\bigg|^2 \\
&\le 3\mathbb{E}|x|^2 + 3\mathbb{E}\bigg|\int_0^t b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)ds\bigg|^2 + 3\mathbb{E}\bigg|\int_0^t \sigma(s, \mathscr{L}_{X_s^\epsilon})dB_s^H\bigg|^2 \\
&\le 3\mathbb{E}|x|^2 + T\mathbb{E}\int_0^t |b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)|^2 ds + 3\mathbb{E}\bigg|\int_0^t \sigma(s, \mathscr{L}_{X_s^\epsilon})dB_s^H\bigg|^2 \\
&\le 3\mathbb{E}|x|^2 + T\mathbb{E}\int_0^t |b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - b(s, 0, \delta_0, 0) + b(s, 0, \delta_0, 0)|^2 ds \\
&\quad + 3\mathbb{E}\bigg|\int_0^t \sigma(s, \mathscr{L}_{X_s^\epsilon})dB_s^H\bigg|^2 \\
&\le 3\mathbb{E}|x|^2 + C_T\mathbb{E}\int_0^t (\kappa(|X_s^\epsilon|^2 + |Y_s^\epsilon|^2 + \mathbb{W}_2(\mathscr{L}_{X_s^\epsilon}, \delta_0)^2) + K(T^2))ds \\
&\quad + 3\mathbb{E}\bigg|\int_0^t \sigma(s, \mathscr{L}_{X_s^\epsilon})dB_s^H\bigg|^2 \\
&\le 3\mathbb{E}|x|^2 + C_T\int_0^t (1 + 2\mathbb{E}|X_s^\epsilon|^2 + \mathbb{E}|Y_s^\epsilon|^2)ds + C_{T,H}\int_0^t (1 + \mathbb{E}|X_s^\epsilon|^2)ds \\
&\le C_{T,|x|} + C_{T,H}\int_0^t (\mathbb{E}|X_s^\epsilon|^2 + \mathbb{E}|Y_s^\epsilon|^2)ds.
\end{aligned} \tag{3.6}$$

As for the Y_t^ϵ, using Itô's formula and Assumption 3.1, we have

$$\begin{aligned}
\frac{d}{dt}\mathbb{E}|Y_t^\epsilon|^2 &= \frac{2}{\epsilon}\mathbb{E}\langle f(t, X_t^\epsilon, \mathscr{L}_{X_t^\epsilon}, Y_t^\epsilon), Y_t^\epsilon\rangle + \frac{1}{\epsilon}\mathbb{E}||g(t, X_t^\epsilon, \mathscr{L}_{X_t^\epsilon}, Y_t^\epsilon)||^2 \\
&\le -\frac{\beta_2}{\epsilon}\mathbb{E}|Y_t^\epsilon|^2 + \frac{C}{\epsilon}(\mathbb{E}|X_t^\epsilon|^2 + 1).
\end{aligned} \tag{3.7}$$

Then by the comparison theorem, we have

$$\begin{aligned}
\sup_{t\in[0,T]} \mathbb{E}|Y_t^\epsilon|^2 &\le |y|^2 e^{-\frac{\beta_2 t}{\epsilon}} + \frac{C}{\epsilon}\int_0^t e^{-\frac{\beta_2(t-s)}{\epsilon}}\bigg(\sup_{s\in[0,t]} \mathbb{E}|X_s^\epsilon|^2 + 1\bigg)ds \\
&\le |y|^2 + C_{T,\beta_2}\bigg(\sup_{s\in[0,t]} \mathbb{E}|X_s^\epsilon|^2 + 1\bigg).
\end{aligned} \tag{3.8}$$

Combing (3.6) with (3.8), we have

$$\sup_{t\in[0,T]}\mathbb{E}|X_t^\epsilon|^2 \le C_{T,|x|,|y|} + C_{T,H,\beta_2}\int_0^t \sup_{s\in[0,t]}\mathbb{E}|X_s^\epsilon|^2 ds.$$

So, Gronwall's inequality yields

$$\sup_{t\in[0,T]}\mathbb{E}|X_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_2}. \tag{3.9}$$

Hence, we have

$$\sup_{t\in[0,T]}\mathbb{E}|X_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_2}, \quad \sup_{t\in[0,T]}\mathbb{E}|Y_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_2}.$$

This completes the proof. ■

Similar to the proof of Lemma 3.4, for $0 \le t \le t+h \le T$, we have

$$\begin{aligned}
&\mathbb{E}|X_{t+h}^\epsilon - X_t^\epsilon|^2 \\
&= \mathbb{E}\left|\left[\int_t^{t+h} b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)ds + \int_t^{t+h}\sigma(s,\mathscr{L}_{X_s^\epsilon})dB_s^H\right]\right|^2 \\
&\le C\mathbb{E}\left[\left|\int_t^{t+h} b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon)ds\right|^2 + \left|\int_t^{t+h}\sigma(s,\mathscr{L}_{X_s^\epsilon})dB_s^H\right|^2\right] \\
&\le C_T h\mathbb{E}\int_t^{t+h}(\kappa(2\mathbb{E}|X_s^\epsilon|^2 + \mathbb{E}|Y_s^\epsilon|^2) + K(T^2))ds + C_T h^{2H-1}\mathbb{E}\int_t^{t+h}\kappa(\mathbb{E}|X_s^\epsilon|^2)ds \\
&\le C_T h\mathbb{E}\int_t^{t+h}(2\mathbb{E}|X_s^\epsilon|^2 + \mathbb{E}|Y_s^\epsilon|^2 + 1)ds + C_T h^{2H-1}\mathbb{E}\int_t^{t+h}(1+\mathbb{E}|X_s^\epsilon|^2)ds \\
&\le C_{T,H,|x|,|y|,\beta_2}(h^2 \vee h^{2H}).
\end{aligned}$$

Hence, we have the following result.

Lemma 3.5 *Suppose that Assumptions 2.1 and 3.1 hold. Then, for $0 \le t \le t+h \le T$, we have*

$$\mathbb{E}|X_{t+h}^\epsilon - X_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_2}(h^2 \vee h^{2H}).$$

3.2 Some estimates for the auxiliary process $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon)$

The aim of this paper is to estimate the difference between X_t^ϵ and $\bar{X}_t$. To this end we introduce an auxiliary process $(\hat{X}_t^\epsilon, \hat{Y}_t^\epsilon)$ and divide [0,T] into intervals of size $\delta < 1$ (δ is a fixed number depending on ϵ). We will use the discretisation techniques from Khasminskii in [16] (we can see, for example, Röckner, Sun and Xie [27], Pei, Inahama and Xu [26]) to construct a process $\hat{Y}_t^\epsilon$ with initial value $\hat{Y}_0^\epsilon = Y_0^\epsilon = y$ such that for $t \in [k\delta, \min((k+1)\delta, T)]$,

$$\hat{Y}_t^\epsilon = \hat{Y}_{k\delta}^\epsilon + \frac{1}{\epsilon}\int_{k\delta}^t f(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon)ds + \frac{1}{\sqrt{\epsilon}}\int_{k\delta}^t g(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon)dW_s.$$

This can be rewritten as

$$\hat{Y}_t^\epsilon = y + \frac{1}{\epsilon}\int_0^t f(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)ds + \frac{1}{\sqrt{\epsilon}}\int_0^t g(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)dW_s,$$

where $s(\delta) := [\frac{s}{\delta}]\delta$ is the nearest breakpoint preceding s, and $[\frac{s}{\delta}]$ is the integer part of $\frac{s}{\delta}$. Similar, we can define the process $\hat{X}_t^\epsilon$ with initial value $\hat{X}_t^\epsilon = x$ by

$$\hat{X}_t^\epsilon = x + \int_0^t b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)ds + \int_0^t \sigma(s, \mathscr{L}_{X_s^\epsilon})dB_s^H.$$

Remark 3.2 *By the construction of $\hat{Y}_t^\epsilon$ and use similar argument as in the proof of Lemma 3.4, it is easy to obtain*

$$\sup_{\epsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}|\hat{Y}_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_2}.$$

Lemma 3.6 *Suppose that Assumptions 2.1 and 3.1 hold. Then, for any $T > 0$, $t \in [0, T]$, we have*

$$\sup_{\epsilon\in(0,1)} \sup_{t\in[0,T]} \mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_1,\beta_2}\kappa(\delta^{2H}).$$

Proof. For any $t \in [0, T]$, we have

$$\begin{aligned} Y_t^\epsilon - \hat{Y}_t^\epsilon =& \frac{1}{\epsilon}\int_0^t [f(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - f(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)]ds \\ &+ \frac{1}{\sqrt{\epsilon}}\int_0^t [g(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - g(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)]dW_s. \end{aligned}$$

Using Itô's formula, we have

$$\begin{aligned} &\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 \\ &\le \frac{2}{\epsilon}\mathbb{E}\int_0^t \langle f(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - f(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon), |Y_t^\epsilon - \hat{Y}_t^\epsilon|\rangle ds \\ &\quad + \frac{1}{\epsilon}\mathbb{E}\int_0^t ||g(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - g(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)||^2 ds. \end{aligned}$$

By Assumption 3.1, we have

$$\begin{aligned} \frac{d}{dt}\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 &\le -\frac{\beta_1}{\epsilon}\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 + \frac{C_T}{\epsilon}\kappa(2\mathbb{E}|X_t^\epsilon - X_{t(\delta)}^\epsilon|^2) \\ &\le -\frac{\beta_1}{\epsilon}\mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 + \frac{C_{T,H,|x|,|y|,\beta_2}}{\epsilon}\kappa(2\delta^{2H}). \end{aligned}$$

Hence, by comparison theorem and Lemma 3.5, we can get

$$\begin{aligned} \mathbb{E}|Y_t^\epsilon - \hat{Y}_t^\epsilon|^2 &\le \frac{C_{T,H,|x|,|y|,\beta_2}}{\epsilon}\int_0^t e^{-\frac{\beta_1(t-s)}{\epsilon}}\kappa(\delta^{2H})ds \\ &\le C_{T,H,|x|,|y|,\beta_1,\beta_2}\kappa(\delta^{2H}). \end{aligned}$$

This completes the proof. ■

Lemma 3.7 *Suppose that Assumptions 2.1 and 3.1 hold. Then, for any $T > 0$, $t \in [0, T]$, we have*

$$\sup_{t\in[0,T]} \mathbb{E}|X_t^\epsilon - \hat{X}_t^\epsilon|^2 \le C_{T,H,|x|,|y|,\beta_1,\beta_2}\kappa(\delta^{2H} + \kappa(\delta^{2H})).$$

Proof. For any $t \in [0,T]$,

$$X_t^\epsilon - \hat{X}_t^\epsilon = \int_0^t b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) ds.$$

By Lemma 3.5 and 3.6, we obtain

$$\begin{aligned}
\sup_{t\in[0,T]} \mathbb{E}|X_t^\epsilon - \hat{X}_t^\epsilon|^2 &\le \mathbb{E}\Big[\int_0^T |b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)| ds\Big]^2 \\
&\le C_T \mathbb{E}\int_0^T |b(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}, Y_s^\epsilon) - b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon)|^2 ds \\
&\le C_T \mathbb{E}\int_0^T K(\delta^2)[\kappa(|X_s^\epsilon - X_{s(\delta)}^\epsilon|^2 + \mathbb{W}_2(\mathscr{L}_{X_s^\epsilon}, \mathscr{L}_{X_{s(\delta)}^\epsilon})^2 + |Y_s^\epsilon - \hat{Y}_s^\epsilon|^2)] ds \\
&\le C_{T,H,|x|,|y|,\beta_1,\beta_2} \kappa(\delta^{2H} + \kappa(\delta^{2H})).
\end{aligned}$$

This completes the proof. ■

3.3 The estimate for $|\hat{X}_t^\epsilon - \bar{X}_t^\epsilon|$

Lemma 3.8 *Suppose that Assumptions 2.1 and 3.1 hold. Then for any $T > 0$, $t \in [0,T]$, we have*

$$\sup_{t\in[0,T]} \mathbb{E}\Big|\int_0^t (b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon})) ds\Big|^2 \le C_{T,H,|x|,|y|}(\frac{\epsilon}{\delta} + \delta).$$

Proof. By elementary inequality, estimate (3.5) and Lemma 3.4, we have

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\Big|\int_0^t (b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon})) ds\Big|^2 \\
&\le C \sup_{t\in[0,T]} \mathbb{E}\Big|\sum_{k=0}^{[t/\delta]-1} \int_{k\delta}^{(k+1)\delta} b(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}) ds\Big|^2 \\
&\quad + C \sup_{t\in[0,T]} \mathbb{E}\Big|\int_{t(\delta)}^t b(t(\delta), X_{t(\delta)}^\epsilon, \mathscr{L}_{X_{t(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(t(\delta), X_{t(\delta)}^\epsilon, \mathscr{L}_{X_{t(\delta)}^\epsilon}) ds\Big|^2 \\
&\le C \sup_{t\in[0,T]} \mathbb{E}\Big([\frac{t}{\delta}] \sum_{k=0}^{[t/\delta]-1} \Big|\int_{k\delta}^{(k+1)\delta} b(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}) ds\Big|^2\Big) \\
&\quad + C_T \delta \sup_{t\in[0,T]} \mathbb{E}\int_{t(\delta)}^t (1 + |X_{t(\delta)}^\epsilon|^2 + |\hat{Y}_s^\epsilon|^2 + \mathbb{E}|X_{t(\delta)}^\epsilon|^2) ds \\
&\le \frac{C_T}{\delta^2} \max_{0\le k\le[\frac{T}{\delta}]-1} \mathbb{E}\Big|\int_{k\delta}^{(k+1)\delta} b(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(k\delta, X_{k\delta}^\epsilon, \mathscr{L}_{X_{k\delta}^\epsilon}) ds\Big|^2 \\
&\quad + C_{T,H,|x|,|y|,\beta_2} \delta^2 \\
&\le C_T \frac{\epsilon^2}{\delta^2} \max_{0\le k\le[\frac{T}{\delta}]-1} \int_0^{\frac{\delta}{\epsilon}} \int_\zeta^{\frac{\delta}{\epsilon}} \phi_k(s,\zeta) ds d\zeta + C_{T,H,|x|,|y|,\beta_2} \delta^2,
\end{aligned}$$

where $0 \le \zeta \le s \le \frac{\delta}{\epsilon}$, and

$$\begin{aligned}\phi_k(s,\zeta) = \mathbb{E}[|\langle b(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, \hat{Y}^\epsilon_{s\epsilon+k\delta}) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}), \\ b(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, \hat{Y}^\epsilon_{\zeta\epsilon+k\delta}) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}})\rangle|].\end{aligned} \tag{3.10}$$

Now, we consider the process $\tilde{Y}^{\epsilon,s,x,\mu,y}_t, t \ge s$, defined as follows

$$\tilde{Y}^{\epsilon,s,x,\mu,y}_t = y + \frac{1}{\epsilon}\int_s^t f(s,x,\mu,\tilde{Y}^{\epsilon,s,x,\mu,y}_r)dr + \frac{1}{\sqrt{\epsilon}}\int_s^t g(s,x,\mu,\tilde{Y}^{\epsilon,s,x,\mu,y}_r)dW_r.$$

By the construction of $\hat{Y}^\epsilon_t$, for any $k \in \mathbb{N}$, we have

$$\hat{Y}^\epsilon_t = \tilde{Y}_t^{\epsilon,k\delta,X^\epsilon_{k\delta},\mathscr{L}_{X^\epsilon_{k\delta}},\hat{Y}^\epsilon_{k\delta}}.$$

Moreover, by the time shift transformation, for any fixed k and $s \in [0, \frac{\delta}{\epsilon}]$, we have

$$\begin{aligned}\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_{s\epsilon+k\delta} &= y + \frac{1}{\epsilon}\int_{k\delta}^{k\delta+s\epsilon} f(k\delta,x,\mu,\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_r)dr + \frac{1}{\sqrt{\epsilon}}\int_{k\delta}^{k\delta+s\epsilon} g(k\delta,x,\mu,\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_r)dW_r \\ &= y + \frac{1}{\epsilon}\int_0^{s\epsilon} f(k\delta,x,\mu,\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_{r+k\delta})dr + \frac{1}{\sqrt{\epsilon}}\int_0^{s\epsilon} g(k\delta,x,\mu,\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_{r+k\delta})dW^*_r \\ &= y + \int_0^s f(k\delta,x,\mu,\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_{r\epsilon+k\delta})dr + \int_0^s g(k\delta,x,\mu,\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_{r\epsilon+k\delta})d\hat{W}^*_r,\end{aligned}$$

where $W^*_t := W_{t+k\delta} - W_{k\delta}$ is the shift version of W_t, $\hat{W}^*_r = \frac{1}{\sqrt{\epsilon}}W^*_{r\epsilon}$. Recall that

$$Y^{k\delta,x,\mu,y}_s = y + \int_0^s f(k\delta,x,\mu,Y^{k\delta,x,\mu,y}_r)dr + \int_0^s g(k\delta,x,\mu,Y^{k\delta,x,\mu,y}_r)d\tilde{W}_r.$$

Hence, we have

$$\tilde{Y}^{\epsilon,k\delta,x,\mu,y}_{s\epsilon+k\delta} \sim Y^{k\delta,x,\mu,y}_s, \tag{3.11}$$

where $\sim$ denotes coincidence in the sense of distribution.

$$\begin{aligned}\phi_k(s,\zeta) &= \mathbb{E}\{\mathbb{E}[\langle b(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, \hat{Y}^\epsilon_{s\epsilon+k\delta}) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}), \\ &\qquad b(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, \hat{Y}^\epsilon_{\zeta\epsilon+k\delta}) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}})\rangle|\mathscr{F}_{k\delta}]\} \\ &= \mathbb{E}\{\mathbb{E}[\langle b(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, \hat{Y}^\epsilon_{s\epsilon+k\delta}) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}), \\ &\qquad b(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, \hat{Y}^\epsilon_{\zeta\epsilon+k\delta}) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}})\rangle]\}. \\ &= \mathbb{E}\Bigg[\tilde{\mathbb{E}}\Bigg\langle b\Bigg(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, Y_s^{k\delta,X^\epsilon_{k\delta},\mathscr{L}_{X^\epsilon_{k\delta}},\hat{Y}^\epsilon_{k\delta}}\Bigg) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}), \\ &\qquad b\Bigg(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, Y_\zeta^{k\delta,X^\epsilon_{k\delta},\mathscr{L}_{X^\epsilon_{k\delta}},\hat{Y}^\epsilon_{k\delta}}\Bigg) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}})\Bigg\rangle\Bigg] \\ &= \mathbb{E}\Bigg[\tilde{\mathbb{E}}\Bigg\langle \tilde{\mathbb{E}}\Bigg[b\Bigg(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, Y_s^{k\delta,X^\epsilon_{k\delta},\mathscr{L}_{X^\epsilon_{k\delta}},\hat{Y}^\epsilon_{k\delta}}\Bigg)\Bigg|\tilde{\mathscr{F}}_\zeta\Bigg] - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}), \\ &\qquad b\Bigg(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}}, Y_\zeta^{k\delta,X^\epsilon_{k\delta},\mathscr{L}_{X^\epsilon_{k\delta}},\hat{Y}^\epsilon_{k\delta}}\Bigg) - \bar{b}(k\delta, X^\epsilon_{k\delta}, \mathscr{L}_{X^\epsilon_{k\delta}})\Bigg\rangle\Bigg].\end{aligned} \tag{3.12}$$

Therefore, by Cauchy-Schwarz's inequality

$$
\begin{aligned}
\phi_k(s,\zeta) \le{}& \mathbb{E}\bigg[\tilde{\mathbb{E}}\bigg|b\bigg(k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, Y^{k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, \hat{Y}^{\epsilon}_{k\delta}}_{s-\zeta}\bigg) - \bar{b}(k\delta, X^{\epsilon}_{k\delta}\mathscr{L}_{X^{\epsilon}_{k\delta}})\bigg|^2\bigg]^{\frac{1}{2}} \\
&+ \mathbb{E}\bigg[\tilde{\mathbb{E}}\bigg|b\bigg(k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, Y^{k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, \hat{Y}^{\epsilon}_{k\delta}}_{\zeta}\bigg) - \bar{b}(k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}})\bigg|^2\bigg]^{\frac{1}{2}},
\end{aligned}
$$

which, (2.1) and (3.5) imply that

$$
\begin{aligned}
&\tilde{\mathbb{E}}\bigg|b\bigg(k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, Y^{k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, \hat{Y}^{\epsilon}_{k\delta}}_{\zeta}\bigg) - \bar{b}(k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}})\bigg|^2 \\
&\le C\tilde{\mathbb{E}}\bigg(1 + |X^{\epsilon}_{k\delta}|^2 + \bigg|Y^{k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, \hat{Y}^{\epsilon}_{k\delta}}_{\zeta}\bigg|^2 + \mathbb{E}|X^{\epsilon}_{k\delta}|^2\bigg).
\end{aligned}
$$

Thus, exist $\alpha > 0$, we have

$$
\begin{aligned}
\phi_k(s,\zeta) &\le C_T\mathbb{E}\bigg\{\tilde{\mathbb{E}}\bigg[1 + |X^{\epsilon}_{k\delta}|^2 + \bigg|Y^{k\delta, X^{\epsilon}_{k\delta}, \mathscr{L}_{X^{\epsilon}_{k\delta}}, \hat{Y}^{\epsilon}_{k\delta}}_{\zeta}\bigg|^2 + \mathscr{L}_{X^{\epsilon}_{k\delta}}(|\cdot|^2)\bigg]e^{\frac{-\alpha}{2}(s-\zeta)}\bigg\} \\
&\le C_T\mathbb{E}\bigg(1 + |X^{\epsilon}_{k\delta}|^2 + |\hat{Y}^{\epsilon}_{k\delta}|^2 + \mathbb{E}|X^{\epsilon}_{k\delta}|^2\bigg)e^{\frac{-\alpha}{2}(s-\zeta)} \\
&\le C_{T,H,|x|,|y|,\beta_2}e^{\frac{-\alpha}{2}(s-\zeta)}.
\end{aligned}
\tag{3.13}
$$

Therefore, by equation (3.13) and choosing $\delta = \delta(\epsilon)$ such that $\frac{\delta}{\epsilon}$ is sufficiently large, we have

$$
\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\bigg|\int_0^t (b(s(\delta), X^{\epsilon}_{s(\delta)}, \mathscr{L}_{X^{\epsilon}_{s(\delta)}}, \hat{Y}^{\epsilon}_s) - \bar{b}(s(\delta), X^{\epsilon}_{s(\delta)}, \mathscr{L}_{X^{\epsilon}_{s(\delta)}}))ds\bigg|^2 \\
&\le C_T\frac{\epsilon^2}{\delta^2}\max_{0\le k\le[\frac{T}{\delta}]-1}\int_0^{\frac{\delta}{\epsilon}}\int_{\zeta}^{\frac{\delta}{\epsilon}}\phi_k(s,\zeta)dsd\zeta + C_{T,H,|x|,|y|,\beta_2}\delta^2 \\
&\le C_{T,H,|x|,|y|,\beta_2}\frac{\epsilon^2}{\delta^2}\max_{0\le k\le[\frac{T}{\delta}]-1}\int_0^{\frac{\delta}{\epsilon}}\int_{\zeta}^{\frac{\delta}{\epsilon}} e^{\frac{-\alpha}{2}(s-\zeta)}dsd\zeta + C_{T,H,|x|,|y|,\beta_2}\delta^2 \\
&\le C_{T,H,|x|,|y|,\beta_2}\frac{\epsilon^2}{\delta^2}\bigg(\frac{2}{\alpha}\frac{\delta}{\epsilon} - \frac{4}{\alpha^2} + \frac{4}{\alpha^2}e^{\frac{-\alpha\delta}{2\epsilon}}\bigg) + C_{T,H,|x|,|y|,\beta_2}\delta^2 \\
&\le C_{T,H,|x|,|y|,\beta_2}(\frac{\epsilon}{\delta} + \delta).
\end{aligned}
$$

This completes the proof. ∎

Lemma 3.9 *Suppose that Assumptions 2.1 and 3.1 hold. Then for any $T > 0$, $t \in [0,T]$, we have*

$$
\lim_{\epsilon\to 0}\sup_{t\in[0,T]} \mathbb{E}|\hat{X}^{\epsilon}_t - \bar{X}_t|^2 = 0.
$$

Proof. Recall that

$$\begin{aligned}\hat{X}_t^\epsilon - \bar{X}_t = &\int_0^t [b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(s, \bar{X}_s, \mathscr{L}_{\bar{X}_s})]ds \\ &+ \int_0^t [\sigma(s, \mathscr{L}_{X_s^\epsilon}) - \sigma(s, \mathscr{L}_{\bar{X}_s})]dB_s^H \\ = &\int_0^t [b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon})]ds \\ &+ \int_0^t [\bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}) - \bar{b}(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon})]ds \\ &+ \int_0^t [\bar{b}(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}) - \bar{b}(s, \bar{X}_s, \mathscr{L}_{\bar{X}_s})]ds \\ &+ \int_0^t [\sigma(s, \mathscr{L}_{X_s^\epsilon}) - \sigma(s, \mathscr{L}_{\bar{X}_s})]dB_s^H.\end{aligned}$$

For any $t \in [0, T]$, by the simple inequality

$$|x_1 + x_2 + \cdots + x_k|^p \leq k^{p-1}(|x_1|^p + |x_2|^p + \cdots + |x_k|^p),$$

and Hölder inequality, we have

$$\begin{aligned}\sup_{t\in[0,T]} \mathbb{E}|\hat{X}_t^\epsilon - \bar{X}_t|^2 \leq &C \sup_{t\in[0,T]} \mathbb{E}\left| \int_0^t [b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon})]ds \right|^2 \\ &+ C \sup_{t\in[0,T]} \mathbb{E}\left| \int_0^t [\bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}) - \bar{b}(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon})]ds \right|^2 \\ &+ C \sup_{t\in[0,T]} \mathbb{E}\left| \int_0^t [\bar{b}(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}) - \bar{b}(s, \bar{X}_s, \mathscr{L}_{\bar{X}_s})]ds \right|^2 \\ &+ C \sup_{t\in[0,T]} \mathbb{E}\left| \int_0^t [\sigma(s, \mathscr{L}_{X_s^\epsilon}) - \sigma(s, \mathscr{L}_{\bar{X}_s})]dB_s^H \right|^2 \\ \leq &C \sup_{t\in[0,T]} \mathbb{E}\left| \int_0^t [b(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}, \hat{Y}_s^\epsilon) - \bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon})]ds \right|^2 \\ &+ C_T \sup_{t\in[0,T]} \mathbb{E} \int_0^t |\bar{b}(s(\delta), X_{s(\delta)}^\epsilon, \mathscr{L}_{X_{s(\delta)}^\epsilon}) - \bar{b}(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon})|^2 ds \\ &+ C_T \sup_{t\in[0,T]} \mathbb{E} \int_0^t |\bar{b}(s, X_s^\epsilon, \mathscr{L}_{X_s^\epsilon}) - \bar{b}(s, \bar{X}_s, \mathscr{L}_{\bar{X}_s})|^2 ds \\ &+ C \sup_{t\in[0,T]} \mathbb{E}\left| \int_0^t [\sigma(s, \mathscr{L}_{X_s^\epsilon}) - \sigma(s, \mathscr{L}_{\bar{X}_s})]dB_s^H \right|^2 \\ =: &\sup_{t\in[0,T]} \sum_{i=1}^4 J_i(t).\end{aligned}$$

For the term $J_1(t)$, by Lemma 3.8 we have

$$J_1(t) \leq C_{T,H,|x|,|y|,\beta_2}(\frac{\epsilon}{\delta} + \delta).$$

For the terms $J_2(t)$ and $J_3(t)$, by the equation (3.5), Lemma 3.5 and Lemma 3.7, we have

$$J_2(t) \leq C_T \int_0^t K(\delta^2)\kappa(2\mathbb{E}|X^\epsilon_{s(\delta)} - X^\epsilon_s|^2)ds \leq C_{T,H,|x|,|y|,\beta_1,\beta_2}\kappa(\delta^{2H}),$$

and

$$\begin{aligned} J_3(t) &\leq C_T \int_0^t \kappa(2\mathbb{E}|X^\epsilon_s - \hat{X}^\epsilon_s|^2) + \kappa(2\mathbb{E}|\hat{X}^\epsilon_s - \bar{X}_s|^2)ds \\ &\leq C_{T,H,|x|,|y|,\beta_1,\beta_2}\kappa(\delta^{2H} + \kappa(\delta^{2H})) + C_T \int_0^t \kappa(\mathbb{E}|\hat{X}^\epsilon_s - \bar{X}_s|^2)ds. \end{aligned}$$

For the term $J_4(t)$, thanks to the result from the proof of ([5], Theorem 3.1) and $J_3(t)$, we have

$$\begin{aligned} J_4(t) &\leq C_{T,H}\mathbb{E}\int_0^t ||\sigma(s, \mathscr{L}_{X^\epsilon_s}) - \sigma(s, \mathscr{L}_{\bar{X}_s})||^2 ds \\ &\leq C_{T,H}K(T^2)\int_0^t \kappa(\mathbb{E}|X^\epsilon_s - \hat{X}^\epsilon_s|^2) + \kappa(\mathbb{E}|\hat{X}^\epsilon_s - \bar{X}_s|^2)ds \\ &\leq C_{T,H,|x|,|y|,\beta_1,\beta_2}\kappa(\delta^{2H} + \kappa(\delta^{2H})) + C_{T,H}\int_0^t \kappa(\mathbb{E}|\hat{X}^\epsilon_s - \bar{X}_s|^2)ds. \end{aligned}$$

Above all, we have

$$\begin{aligned} &\sup_{t\in[0,T]} \mathbb{E}|\hat{X}^\epsilon_t - \bar{X}_t|^2 \\ &\leq C_{T,H,|x|,|y|,\beta_1,\beta_2}\left[\frac{\epsilon}{\delta} + \delta + \kappa(\delta^{2H}) + \kappa(\delta^{2H} + \kappa(\delta^{2H}))\right] + C_{T,H}\int_0^T \kappa\left(\sup_{r\in[0,s]} \mathbb{E}|\hat{X}^\epsilon_r - \bar{X}_r|^2\right)ds. \end{aligned}$$

With the aid of taking $\delta = \sqrt{\epsilon}$ and letting $\epsilon \to 0$, yields

$$\begin{aligned} &\lim_{\epsilon\to 0}\sup_{t\in[0,T]} \mathbb{E}|\hat{X}^\epsilon_t - \bar{X}_t|^2 \\ &\leq C_{T,H}\int_0^T \kappa\left(\lim_{\epsilon\to 0}\sup_{r\in[0,s]} \mathbb{E}|\hat{X}^\epsilon_r - \bar{X}_r|^2\right)ds. \end{aligned}$$

Hence, by the result of [[3],Lemma 3.7], we have

$$\lim_{\epsilon\to 0}\sup_{t\in[0,T]} \mathbb{E}|\hat{X}^\epsilon_t - \bar{X}_t|^2 = 0.$$

This completes the proof. ■

3.4 The estimate for $|X^\epsilon_t - \bar{X}_t|$

Suppose Assumption 2.1 and 3.1 hold. Taking $\delta = \sqrt{\epsilon}$, by Lemma 3.7 and 3.9, any $T > 0$, initial values $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, we have

$$\lim_{\epsilon\to 0}\left[\sup_{t\in[0,T]} \mathbb{E}|X^\epsilon_t - \bar{X}_t|^2\right] \leq \lim_{\epsilon\to 0}\left[\sup_{t\in[0,T]} \mathbb{E}(|X^\epsilon_t - \hat{X}^\epsilon_t|^2 + |\hat{X}^\epsilon_t - \bar{X}_t|^2)\right] = 0.$$

This completes the proof of Theorem 3.2.

Acknowledgements. This research was financially supported by the Natural Science Foundation of China (No. 12071003). The authors would like to thank the referee and Prof Tusheng Zhang (Associate Editor) for their insightful comments and stimulation which have led us to improve the presentation of the paper.

References

[1] E. Alòs and D. Nualart, Stochastic integration with respect to the fractional Brownian motion. *Stochastics* **75** (2003), 129-152.

[2] F. Biagini, Y. Hu, B. Øksendal and T. Zhang, *Stochastic Calculus for Fractional Brownian Motion and Applications*. Springer-Verlag, London (2008).

[3] G. Cao and K. He, On a type of stochastic differential equations driven by countably many brownian motions, *J. Funct. Anal.* **203** (2003), 262-285.

[4] Z. Dong, X. Sun, H. Xiao and J. Zhai, Averaging principle for one dimensional stochastic Burgers equation. *J. Differential Equations.* **265** (2018), 4749-4797.

[5] X. Fan, X. Huang, Y. Suo and C. Yuan, Distribution dependent SDEs driven by fractional Brownian motions. *Stochastic Process. Appl.* **151** (2022) 23–67.

[6] J. P. Fouque, G. Papanicolaou, R. Sircar and K. Solna, *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives*. Cambridge University Press (2011).

[7] D. Givon, Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems. *Multiscale Model. Simul.* **6** (2007), 577-594.

[8] Z. Guo, G. Lv and J. Wei. Averaging principle for stochastic differential equations under a weak condition. *Chaos.* **30** (2020), 123139 .

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

[10] Y. Hu, *Analysis on Gaussian Spaces*. World Scientific Publishing Co. Pte. Ltd., Hackensack, JZ (2017).

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

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

[13] X. Huang, P. Ren and F-Y Wang, Distribution dependent stochastic differential equations. *Front. Math. China.* **16** (2021), 257-301.

[14] M. Kac, Foudations of kinetic theory. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability 1954-1955. vol. III. pages 171-197. University of California Press. Berkeley and Los Angeles (1956).

[15] M. Kac, *Probability and Related Topics in the Physical Sciences.* Interscience Publishers, New York (1959).

[16] R. Khasminskii, On the principle of averaging the Itô stochastic differential equations. *Kybernetika* **4** (1968), 260-279.

[17] Y. Li, X. Mao, Q. Song, F. Wu and G. Yin, Strong convergence of Euler-Maruyama schemes for McKean-Vlasov stochastic differential equations under local Lipschitz conditions of state variables. *IMA J. Numer. Anal.* (2022) https://doi.org/10.1093/imanum/drab107

[18] W. Liu, M. Röckner, X. Sun and Y. Xie, Averaging principle for slow-fast stochastic differential equations with time dependent locally Lipschitz coefficients. *J. Differential Equations* **268** (2020), 2910-2948.

[19] D. Luo, Q. Zhu and Z. Luo, An averaging principle for stochastic fractional differential equations with time-delays. *Appl. Math. Lett.***105** (2020), 106290.

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

[21] S. Mehri and W. Stannat, Weak solutions to Vlasov-McKean equations under Lyapunov-type conditions. *Stoch. Dyn.* **19** (2019) 1950042.

[22] Y. Mishura, *Stochastic Calculus for Fractional Brownian Motion and Related Processes.* Lect. Notes in Math. **1929** (2008).

[23] D. Nualart, *Malliavin Calculus and Related Topics.* 2nd edition, Springer, New York (2006).

[24] B. Pei, Y. Xu and J.-L. Wu, Stochastic averaging for stochastic differential equations driven by fractional Brownian motion and standard Brownian motion. *Appl. Math. Lett.* **100** (2020), 106006.

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

[26] B. Pei, Y. Inahama and Y. Xu, Averaging principles for mixed fast-slow systems driven by fractional Brownian. *arXiv:2001.06945v4*

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

[28] G. Shen, J. Song and J.-L. Wu, Stochastic averaging principle for distribution dependent stochastic differential equations. *Appl. Math. Lett.* **125** (2022), 107761.

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

[30] G. Shen, J.-L. Wu, R. Xiao, and X. Yin, An averaging principle for neutral stochastic fractional order differential equations with variable delays driven by Lévy noise. *Stoch. Dyn.* **22** (2022) 2250009

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

[32] F. Wu, T. Tian, J. B. Rawlings and G. Yin, Approximate method for stochastic chemical kinetics with two-time scales by chemical Langevin equations. *J. Chem. Phys.* **144** (2016), 174112.

[33] F. Xi and C. Zhu, Jump type stochastic differential equations with non-lipschitz coefficients: non confluence, Feller and strong Feller properties, and exponential ergodicity, *J. Differ. Equ.* **266** (2019), 4668-4711.

[34] J. Xu, J. Liu and Y. Miao, Strong averaging principle for two-time-scale SDEs with non-Lipschitz coefficients. *J. Math. Anal. Appl.* **468** (2018), 116-140.

[35] J. Xu, J. Liu, Stochastic averaging principle for two time-scale jump-diffusion SDEs under the non-Lipschitz coefficients. Stochastics (2020) DOI:10.1080/17442508.2020.1784897

[36] J. Xu, J. Liu, J. Liu and Y. Miao, Strong averaging principle for two-time-scale stochastic McKean-Vlasov equations. *Appl. Math. Optim.* **84 (Suppl 1)** (2021), 837-867.

[37] Y. Xu, J. Duan and W. Xu, An averaging principle for stochastic dynamical systems with Lévy noise. *Phys. D* **240** (2011), 1395-1401.

[38] Y. Xu, B. Pei and J.-L. Wu, Stochastic averaging principle for differential equations with non-Lipschitz coefficients driven by fractional Brownian motion. *Stoch. Dyn.***17** (2017), 1750013.

Guangjun Shen
Department of Mathematics, Anhui Normal University, Wuhu 241002, China
gjshen@163.com

Jiayuan Yin
Department of Mathematics, Anhui Normal University, Wuhu 241002, China
yinjiayuan18@163.com

Jiang-Lun Wu
Department of Mathematics, Computational Foundry, Swansea University
Swansea, SA1 8EN, UK
j.l.wu@swansea.ac.uk