Two-time-scale stochastic differential delay equations driven by multiplicative fractional Brownian noise: averaging principle

Min HAN[a], Yong XU[a], Bin PEI[a,b], Jiang-Lun WU[c]

[a] *School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an, 710072, China*
[b] *Research & Development Institute of Northwestern Polytechnical University in Shenzhen, Shenzhen, 518057, China*
[c] *Department of Mathematics, Swansea University, Swansea, SA1 8EN, UK*

Abstract

The main goal of this article is to study an averaging principle for a class of two-time-scale stochastic differential delay equations in which the slow-varying process includes a multiplicative fractional Brownian noise with Hurst parameter $H \in (\frac{1}{2}, 1)$ and the fast-varying process is a rapidly-changing diffusion. We would like to emphasize that the approach proposed in this paper is based on the fact that a stochastic integral with respect to fractional Brownian motion with Hurst parameter in $(\frac{1}{2}, 1)$ can be defined as a generalized Stieltjes integral. In particular, to prove a limit theorem for the averaging principle, we will introduce a sequence of stopping times to control the size of multiplicative fractional Brownian noise. Then, inspired by the Khasminskii's approach, an averaging principle is developed in the sense of convergence in the p-th moment uniformly in time.

Keywords. Averaging principle, two-time-scale, stochastic differential delay equations, multiplicative fractional Brownian noise

Mathematics subject classification. 60G22, 60H10, 60F25,

1. Introduction

Let $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\geq 0}, \mathbb{P})$ be a stochastic basis satisfying the usual conditions. Given $H \in (0,1)$, a continuous centered Gaussian process $(B^H(t))_{t\geq 0}$ with the covariance function

$$\mathbb{E}[B^H(t)B^H(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H}), \qquad t, s \in \mathbb{R}_+,$$

is called one-dimensional fractional Brownian motion (FBM) and H is the corresponding Hurst parameter. Since FBM characterized by the stationarity of its increments and a medium- or long-memory property, so it is in sharp contrast with martingales and Markov processes [20, 22]. FBM also exhibits power scaling and path regularity properties with Hurst parameter H. It has become a popular choice for applications where classical processes cannot model the property of long memory [2, 6]. Due to the long-memory property of FBM when $H \in (\frac{1}{2}, 1)$, thus, in this paper, we restrict ourselves to consider values of the Hurst parameter bigger than $\frac{1}{2}$.

It is well known that owing to different rates of interactions of subsystems and components, singularly perturbed systems which have a wide range of applications in science and engineering usually exhibit multi-scale behavior. Although there has been vast literature on the study for singularly perturbed systems [7, 9], the multi-scale property makes the underlying systems highly

Email addresses: mhan2019@hotmail.com (Min HAN), hsxu3@nwpu.edu.cn (Yong XU), binpei@nwpu.edu.cn (Bin PEI), j.l.wu@swansea.ac.uk (Jiang-Lun WU)

complex, thus difficult to analyze. The averaging principle pioneered by Khasminskii [16] for a class of diffusions provides an effective way to reduce the complexity of the systems in which both fast and slow components co-exist reflected by a time-scale separation parameter $\varepsilon \in (0,1)$. The idea of averaging principle is that there esixts a limit system given by an average of the slow component with respect to the invariant measure of the fast component and it can approximate the slow component in a suitable sense whenever $\varepsilon \downarrow 0$. The work on stochastic averaging principles proposed by Khasminskii [16] inspired much of the subsequent development; see [12, 17, 18, 25, 35, 36, 37, 39, 40] for stochastic differential equations (SDEs) and [3, 4, 10, 27, 28, 29, 30, 34] for stochastic partial differential equations. In particular, Hairer and Li [14] considered slow-fast systems where the slow system is driven by FBM and proved the convergence to the averaged solution took place in probability. Very recently, Pei, Inahama and Xu answered affirmatively that an averaging principle still holds for fast-slow mixed SDEs driven by both Brownian motion (BM) and FBM $H \in (\frac{1}{2}, 1)$ in the mean square sense [25] and $H \in (\frac{1}{3}, \frac{1}{2}]$ in the mean sense [26]. The aforementioned references are all concerned with systems without memory. Nevertheless, in response to the great needs of dynamical systems with memory (delay), there is also extensive literature on stochastic differential delay equations (SDDEs); see for example, [8, 19] and [21]. Bao, Song, Yin and Yuan [1] studied ergodicity and strong limit results for an averaging principle for a class of two-time-scale functional SDEs. Later, Hu and Yuan [15] extended results in [1] to neutral functional SDEs with two-time-scales. Using weak convergence method, Wu and Yin [36] developed an averaging principle for functional diffusions with two-time scales in which the slow-varying process includes path-dependent functionals and the fast-varying process is a rapidly-changing diffusion. Nevertheless, except some developments for functional diffusions such as [1, 15, 36], the investigation on two-time-scale SDDEs with non-martingale-type noises is even more scarce to the best of our knowledge.

In contrast to the rapid progress in two-time-scale delay systems and non-martingale-type noises, the study on averaging principles for SDDEs driven by multiplicative fractional Brownian noise is still in its infancy. In addition, the underlying random noise in financial mathematics, which consists of two parts: one part, describing the economical background for a stock price (a long memory which is a property of FBM), and the other part, coming from the randomness inherent for the stock market (a Brownian noise), is much more natural. Because, BM is lack of memory, and FBM with $H \in (\frac{1}{2}, 1)$ is too smooth, a model driven by both processes is free of such drawbacks. For examples, a mixed Black and Scholes model was firstly proposed by Schoenmakers and Kloeden [31] to discuss the problem of arbitrage. Cheridito [5] studied the martingale properties of the linear combination of BM and FBM independently.

With the motivation above, this work aims to establish an averaging principle for fast-slow mixed SDDEs. In this paper, we shall bring delays, Brownian noise, multiplicative fractional Brownian noise and two-time-scale system together, and prove a limit theorem for the averaging principle for SDDEs driven by multiplicative fractional Brownian noise with Hurst parameter $H \in (\frac{1}{2}, 1)$ and Brownian noise. Since multiplicative fractional Brownian noise and Brownian noise coexisting, we see that the techniques in the present paper are much more complicated and different from those of [1, 15], our main tools consist of precise estimates in Besov-type spaces (see (2.5) below) and fractional calculus approach, i.e. generalized Stieltjes integral method, following the methodology presented in Nualart and Rascanu [23]. Moreover, the technique adopted in [25, Lemma 4.2], which is a key ingredient in obtaining averaging principle, does not work for the case SDDEs and one of the outstanding issues is the infinite-dimensional phase space of the segment processes (see X_t, Y_t bellow), which makes the goal of estimating the displacement of the segment process a very difficult task (see Lemma 3.8 bellow). To overcome these difficulties, new approaches have to be developed. A key of our approach is the use of the newly developed fractional calculus approach. Our main idea of the proof for the limit theorem is based on considering a suitable sequence of stopping times to control the size of the multiplicative fractional Brownian noise. Then, inspired by the Khasminskii's approach, a limit theorem of the averaging principle is proved in the sense of convergence in the p-th moment uniformly in time. Let us point out again that the novelty of our paper is the segment processes and delay dealing with two-time-scale equations driven by both multiplicative fractional Brownian noise and Brownian noise and some previous works are generalized and improved partially, e.g.[1, 25, 39, 40].

This paper is organized as follows. Section 2 presents some necessary notations and assumptions. A limit theorem of the averaging principle for two-time-scale mixed SDDEs driven by multiplicative fractional Brownian noise subject to an additional fast-varying diffusion process is then proved in Section 3. Finally, an appendix is provided at the end of the paper as technical complements.

2. Preliminaries

In this section, we will recall some basic facts on the generalised Stieltje integral that will play a main role in our paper, see e.g. [23, 13, 22, 41] for more details. Throughout this paper, unless otherwise specified, we use the following notation. Let $\mathbb{R}^n$ be an n-dimensional Euclidean space with norm $|\cdot|$. Let $T>0$. Fix the parameter α, such that $0<\alpha<\frac{1}{2}$, denote by $W^{\alpha,1}(0,T;\mathbb{R}^n)$ the space of measurable functions $f:[0,T]\to\mathbb{R}^n$ such that

$$\|f\|_{\alpha,1}:=\int_0^T\frac{|f(s)|}{s^\alpha}ds+\int_0^T\int_0^s\frac{|f(s)-f(\zeta)|}{(s-\zeta)^{\alpha+1}}d\zeta ds<\infty.$$

Following Zähle [41], for $f\in W^{\alpha,1}(0,T;\mathbb{R}^n), 0\le s<t\le T$, we can define a generalized Stieltje integral

$$\int_0^T f(r)dg(r)=(-1)^\alpha\int_0^T D^\alpha_{0+}f(r)D^{1-\alpha}_{T-}g_{T-}(r)dr, \tag{2.1}$$

$$\int_s^t f(r)dg(r)=\int_0^T f(r)\mathbf{1}_{(s,t)}dg(r), \tag{2.2}$$

where, in general, for $0\le a<c\le T, g_{c-}(r):=g(r)-g(c)$, and for $a<t<c$ the Weyl derivatives are given respectively by

$$\begin{aligned} D^\alpha_{a+}f(t) &= \frac{1}{\Gamma(1-\alpha)}\Big(\frac{f(t)}{(t-a)^\alpha}+\alpha\int_a^t\frac{f(t)-f(\zeta)}{(t-\zeta)^{\alpha+1}}d\zeta\Big),\\ D^{1-\alpha}_{c-}g_{c-}(t) &= \frac{(-1)^{1-\alpha}}{\Gamma(\alpha)}\Big(\frac{g(t)-g(c)}{(c-t)^{1-\alpha}}+(1-\alpha)\int_t^c\frac{g(t)-g(\zeta)}{(\zeta-t)^{2-\alpha}}d\zeta\Big),\end{aligned}$$

where Γ denotes the Gamma function. It can be proved that the integral (2.1) exists and that the following crucial inequality holds

$$\Big|\int_0^T f(t)dg(t)\Big|\le\frac{\|g\|_{\alpha,0,T}}{\Gamma(1-\alpha)\Gamma(\alpha)}\|f\|_{\alpha,1},$$

where

$$\|g\|_{\alpha,0,T}:=\sup_{0\le s<t\le T}\Big(\frac{|g(t)-g(s)|}{(t-s)^{1-\alpha}}+\int_s^t\frac{|g(\zeta)-g(s)|}{(\zeta-s)^{2-\alpha}}d\zeta\Big)<\infty.$$

For the sake of shortness, we denote $\Lambda_{\alpha,g}:=\frac{\|g\|_{\alpha,0,T}}{\Gamma(1-\alpha)\Gamma(\alpha)}$.

From now on, given the m-dimensional FBM denoted by $(B^H_t)_{t\ge0}$ with $H\in(\frac{1}{2},1)$, we take a parameter $\alpha\in(1-H,\frac{1}{2})$ which will be fixed througout this paper. For $f\in W^{\alpha,1}(0,T;\mathbb{R}^n)$ the integral

$$\int_0^T f(s)dB^H_s$$

will be understood in the sense of definition (2.1) pathwise, which makes sense due to $\Lambda_{\alpha,B^H} < \infty$ a.s. (cf. [23]), that is

$$\Big|\int_0^T f(t)dB_t^H\Big| \le \Lambda_{\alpha,B^H}\|f\|_{\alpha,1}. \tag{2.3}$$

Furthermore, by the classical Fernique's theorem, for any $0 < \vartheta < 2$, we have

$$\mathbb{E}\big[e^{(\Lambda_{\alpha,B^H})^{\vartheta}}\big] < \infty. \tag{2.4}$$

We follow the approach [13, 23] to introduce some necessary spaces and norms. Let $\tau > 0$, $(s,t) \subset [-\tau, T]$. We will denote by $W_0^{\alpha,\infty}(s,t;\mathbb{R}^n)$ the space of measurable functions $f : [s,t] \to \mathbb{R}^n$ such that

$$\|f\|_{\alpha,\infty(s,t)} := \sup_{r\in[s,t]}\Big(|f(r)| + \int_s^r \frac{|f(r)-f(u)|}{(r-u)^{\alpha+1}}du\Big) < \infty, \tag{2.5}$$

For shortness, denote $\|f(r)\|_{\alpha(s)} := |f(r)| + \int_s^r \frac{|f(r)-f(u)|}{(r-u)^{\alpha+1}}du$. We also need to introduce a new norm in the space $W_0^{\alpha,\infty}(s,t;\mathbb{R}^n)$, that is, for any $\lambda \ge 1$

$$\|f\|^p_{\alpha,\lambda(s,t)} \quad := \quad \sup_{r\in[s,t]} e^{-\lambda r}\bigg(|f(r)| + \int_s^r \frac{|f(r)-f(u)|}{(r-u)^{\alpha+1}}du\bigg)^p, \qquad p \ge 1.$$

We will use the notation $\|f\|^p_{\alpha,\infty(\tau)} := \|f\|^p_{\alpha,\infty(-\tau,T)}$, $\|f\|^p_{\alpha,\lambda(\tau)} := \|f\|^p_{\alpha,\lambda(-\tau,T)}$ and $\|f(r)\|_{\alpha(\tau)} := \|f(r)\|_{\alpha(-\tau)}$. Note that when $\tau = 0$, we shall omit (τ) in the name of the corresponding norm.

Now, we recall an auxiliary technical lemma from [11].

Lemma 2.1. *For any non-negative a and b such that $a + b < 1$, and for any $\lambda \ge 1$, there exists a positive constant C such that*

$$\int_0^t e^{-\lambda(t-r)}(t-r)^{-a}r^{-b}dr \le C\lambda^{a+b-1}.$$

In addition, for $b \le 0$ and $0 \le a < 1$, and for any $\lambda \ge 1$, we have

$$\int_0^t e^{-\lambda(t-r)}(t-r)^{-a}r^{-b}dr \le \Gamma(1-a)t^{-b}\lambda^{a-1}.$$

Later on, we will also need the following estimate which follows from [23, Proposition 4.1 and Proposition 4.3].

Lemma 2.2. *For measurable functions $f : [0,T] \to \mathbb{R}^n$, there exists a constant $C > 0$ such that*

$$\Big\|\int_0^t f(r)dr\Big\|_\alpha \le C\int_0^t |f(r)|(t-r)^{-\alpha}dr,$$

and

$$\Big\|\int_0^t f(r)dB_r^H\Big\|_\alpha \le C\Lambda_{\alpha,B^H}\int_0^t ((t-r)^{-2\alpha} + r^{-\alpha})\Big(|f(r)| + \int_0^r \frac{|f(r)-f(q)|}{(r-q)^{1+\alpha}}dq\Big)dr.$$

Througout this paper, C and C_* denote positive constants that may depend on the parameters α, T and the initial values and vary from line to line. C_* is used to emphasize that the constant depends on the corresponding parameter $*$ which is one or more than one parameter.

3. Systems of Fast-Slow SDDEs

Let $\mathbb{R}^n \otimes \mathbb{R}^m$ denote the collection of all $n \times m$ matrices with real entries. For an $A \in \mathbb{R}^n \otimes \mathbb{R}^m$, $\|A\|$ stands for its Frobenius matrix norm. For a fixed $\tau > 0$, let $\mathscr{L} := C([-\tau, 0]; \mathbb{R}^n)$ denote the family of all continuous functions from $[-\tau, 0] \longrightarrow \mathbb{R}^n$, endowed with the uniform norm $\|\cdot\|_\infty$. For $h(\cdot) \in C([-\tau, \infty); \mathbb{R}^n)$ and $t \geq 0$, define the segment $h_t \in \mathscr{L}$ by $h_t(\theta) := h(t+\theta), \theta \in [-\tau, 0]$. Let $B^H = \{B_t^H, t \in [0,T]\}$ and $W = \{W_t, t \in [0,T]\}$ be independent m-dimensional FBM adapted to $\{\mathscr{F}_t\}$ and m-dimensional $\{\mathscr{F}_t\}$-Bm, respectively.

We are concerned with the following mixed SDDEs driven by multiplicative fractional Brownian noise with Hurst parameter $H \in (\frac{1}{2}, 1)$ and Brownian noise:

$$dX^\varepsilon(t) = b_1(X_t^\varepsilon, Y_t^\varepsilon)\, dt + \sigma_1(X^\varepsilon(t-\tau))dB_t^H, \qquad t > 0, \qquad X_0^\varepsilon = \xi \in \mathscr{L}, \tag{3.1}$$

$$dY^\varepsilon(t) = \frac{1}{\varepsilon} b_2(X_t^\varepsilon, Y^\varepsilon(t), Y^\varepsilon(t-\tau))\, dt + \frac{1}{\sqrt{\varepsilon}} \sigma_2(X_t^\varepsilon, Y^\varepsilon(t), Y^\varepsilon(t-\tau))\, dW_t, \tag{3.2}$$

with the initial value $Y_0^\varepsilon = \eta \in \mathscr{L}$, where the parameter $0 < \varepsilon \ll 1$ represents the ratio between the natural time scale of the X^ε and Y^ε variables and $b_1 : \mathscr{L} \times \mathscr{L} \to \mathbb{R}^n, \sigma_1 : \mathbb{R}^n \to \mathbb{R}^n \otimes \mathbb{R}^m, b_2 : \mathscr{L} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma_2 : \mathscr{L} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \otimes \mathbb{R}^m$ are Gâteaux differentiable. The integral $\int \cdot\, dW$ should be interpreted as an Itô stochastic integral and the integral $\int \cdot\, dB^H$ as a generalised Stieltjes integral.

We denote by $\nabla^{(i)}$ the gradient operators for the i-th component. Throughout this article, for any $\chi, \psi \in \mathscr{L}$ and $x_1, x_2, y_1, y_2 \in \mathbb{R}^n$, we assume that

- (H1) $\nabla b_1 = (\nabla^{(1)} b_1, \nabla^{(2)} b_1)$ is bounded, and there exists $L_1 > 0$ such that
$$|b_1(\chi, \psi)| \leq L_1(1 + \|\chi\|_\infty).$$
- (H2) The function σ_1 is C^1 such that its Frechet derivative w.r.t x is bounded and globally Lipschitz continuous, i.e. there exist $L_2, L_3 > 0$ such that
$$|D^1\sigma_1(x)| \leq L_2 \qquad \text{and} \qquad |D^1\sigma_1(x) - D^1\sigma_1(y)| \leq L_3|x - y|.$$
- (H3) $\nabla b_2 = (\nabla^{(1)} b_2, \nabla^{(2)} b_2, \nabla^{(3)} b_2)$ and $\nabla \sigma_2 = (\nabla^{(1)} \sigma_2, \nabla^{(2)} \sigma_2, \nabla^{(3)} \sigma_2)$ are bounded.
- (H4) There exist $\lambda_1 > \lambda_2 > 0$, independent of χ, such that
$$\begin{aligned} 2\langle x_1 - x_2, b_2(\chi, x_1, y_1) - b_2(\chi, x_2, y_2)\rangle + \|\sigma_2(\chi, x_1, y_1) - \sigma_2(\chi, x_2, y_2)\| \\ \leq -\lambda_1|x_1 - x_2|^2 + \lambda_2|y_1 - y_2|^2. \end{aligned}$$
- (H5) For the intial value $X_0^\varepsilon = \xi \in \mathscr{L}$, there exists a $\lambda_3 > 0$ such that
$$|\xi(t) - \xi(s)| \leq \lambda_3|t - s|, \qquad s, t \in [-\tau, 0].$$

According to [33, Theorem 4.1] and [19, Theorem 2.2, pp.150], the existence and uniqueness of the solutions of (3.1) are guaranteed by the conditions (H1)-(H3) and (H5).

Lemma 3.1. *Suppose that* (H1)-(H3) *and* (H5) *hold. Then,* (3.1) *has a unique strong solution* $(X^\varepsilon(t), Y^\varepsilon(t))_{t \geq -\tau}$, *i.e.,*

$$\begin{aligned}
X^\varepsilon(t) &= \xi(0) + \int_0^t b_1(X_s^\varepsilon, Y_s^\varepsilon)\, ds + \int_0^t \sigma_1(X^\varepsilon(s-\tau))dB_s^H, \quad t > 0, \\
Y^\varepsilon(t) &= \eta(0) + \frac{1}{\varepsilon}\int_0^t b_2(X_s^\varepsilon, Y^\varepsilon(s), Y^\varepsilon(s-\tau))\, ds \\
&\quad + \frac{1}{\sqrt{\varepsilon}}\int_0^t \sigma_2(X_s^\varepsilon, Y^\varepsilon(s), Y^\varepsilon(s-\tau))\, dW_s, \quad t > 0, \\
X_0^\varepsilon &= \xi \in \mathscr{L}, Y_0^\varepsilon = \eta \in \mathscr{L}.
\end{aligned}$$

3.1. Ergodicity of the Frozen Equation with Memory

Consider an SDE with memory associated with the fast motion and frozen slow component in the following form

$$dY(t) = b_2(\chi, Y(t), Y(t-\tau))dt + \sigma_2(\chi, Y(t), Y(t-\tau))dW_t, \quad t > 0, \tag{3.3}$$

with the initial value $Y_0 = \eta \in \mathscr{L}$.

Under (H3), (3.3) has a unique strong solution $(Y(t))_{t\geq-\tau}$ (see, e.g. [19, Theorem 2.2, pp. 150]). To highlight the initial value $\eta \in \mathscr{L}$ and frozen segment $\chi \in \mathscr{L}$, we write the corresponding solution process $(Y^{\chi,\eta}(t))_{t\geq-\tau}$ and the segment process $(Y_t^{\chi,\eta})_{t\geq0}$ instead of $(Y(t))_{t\geq-\tau}$ and $(Y_t)_{t\geq0}$, respectively.

In fact, the unique invariant measure with respect to the frozen equation (3.3) has been obtained in [1]. So, we recall the ergodicity result here.

Lemma 3.2. *Under* (H3)-(H4)*,* $Y_t^{\chi,\eta}$ *has a unique invariant measure* μ^χ*, and there exist constants* $C, \rho > 0$ *such that*

$$\left|\mathbb{E}[b_1(\chi, Y_t^{\chi,\eta})] - \bar{b}_1(\chi)\right| \leq Ce^{-\rho t}(1 + \|\chi\|_\infty + \|\eta\|_\infty), \quad t \geq 0, \quad \eta \in \mathscr{L},$$

where

$$\bar{b}_1(\chi) = \int_{\mathscr{L}} b_1(\chi, \varphi)\mu^\chi(d\varphi), \quad \chi \in \mathscr{L}, \tag{3.4}$$

and μ^χ *is a unique invariant measure with respect to the frozen equation (3.3).*

Let $\tilde{\mathscr{F}}_t$ be the σ-field generated by $\{Y_r^{\chi,\eta}, r \leq t\}$ and for $0 \leq \zeta \leq s \leq T$, set

$$\mathcal{J}(s, \zeta, \chi, \eta) = \mathbb{E}[\langle b_1(\chi, Y_s^{\chi,\eta}) - \bar{b}_1(\chi), b_1(\chi, Y_\zeta^{\chi,\eta}) - \bar{b}_1(\chi)\rangle]. \tag{3.5}$$

Then, the following lemma holds.

Lemma 3.3. *For* $0 \leq \zeta \leq s \leq T$*, there exist constants* $C, \rho > 0$ *which are independent of* s, ζ *such that*

$$\mathcal{J}(s, \zeta, \chi, \eta) \leq C(1 + \|\chi\|_\infty^2 + \|\eta\|_\infty^2)e^{-\frac{\rho}{2}(s-\zeta)}. \tag{3.6}$$

Proof: By (3.5), invoking the Markov property of $Y_t^{\chi,\eta}$, one has

$$\begin{aligned} \mathcal{J}(s, \zeta, \chi, \eta) &= \mathbb{E}[\langle b_1(\chi, Y_\zeta^{\chi,\eta}) - \bar{b}_1(\chi), \mathbb{E}[(b_1(\chi, Y_s^{\chi,\eta}) - \bar{b}_1(\chi))|\tilde{\mathscr{F}}_\zeta]\rangle] \\ &\leq \mathbb{E}\big[\langle b_1(\chi, Y_\zeta^{\chi,\eta}) - \bar{b}_1(\chi), \mathbb{E}^{Y_\zeta^{\chi,\eta}}[b_1(\chi, Y_{s-\zeta}^{\chi,\eta}) - \bar{b}_1(\chi)]\rangle\big]. \end{aligned}$$

Using Hölder's inequality first and (H1), Lemma 3.2 and [1, Section 3, (3.11)], we obtain

$$\begin{aligned} \mathcal{J}(s, \zeta, \chi, \eta) &\leq (\mathbb{E}[|b_1(\chi, Y_\zeta^{\chi,\eta}) - \bar{b}_1(\chi)|^2])^{\frac{1}{2}}\big(\mathbb{E}\big[|\mathbb{E}^{Y_\zeta^{\chi,\eta}}[b_1(\chi, Y_{s-\zeta}^{\chi,\eta}) - \bar{b}_1(\chi)]|^2\big]\big)^{\frac{1}{2}} \\ &\leq C(1 + \|\chi\|_\infty^2 + \|\eta\|_\infty^2)e^{-\frac{\rho}{2}(s-\zeta)}, \end{aligned}$$

where $C > 0$ is a constant. This completes the proof. □

Now, we recall the following result from [1].

Lemma 3.4. *Suppose that* (H1)-(H5) *hold. Then,* $\bar{b}_1 : \mathscr{L} \to \mathbb{R}^n$*, defined by (3.4), is Lipschitz.*

3.2. Main Result.

According to (3.4), we can formulate an averaged equation:

$$d\bar{X}(t) = \bar{b}_1(\bar{X}_t)dt + \sigma_1(\bar{X}(t-\tau))dB_t^H, \quad t > 0, \quad \bar{X}_0 = \xi \in \mathscr{L}. \tag{3.7}$$

By Lemma 3.4 and [33, Theorem 4.1], it is easy to know (3.7) has a unique strong solution $(\bar{X}(t))_{t\geq -\tau}$.

We now state our main result of averaging principle in the sense of convergence in the p-th moment uniformly in time.

Theorem 3.5. *Suppose that* (H1)-(H5) *hold, for any* $p > 0$*, one has*

$$\lim_{\varepsilon\to 0} \mathbb{E}\big[\|X^\varepsilon - \bar{X}\|^p_{\alpha,\infty(\tau)}\big] = 0.$$

The proof of Theorem 3.5 consists of the following steps: Firstly, we give some a priori estimate for the solution of (3.1). Secondly, following the discretization techniques inspired by Khasminskii in [16], we introduce the auxiliary process $(\hat{X}^\varepsilon, \hat{Y}^\varepsilon)$ and divide $[0, T]$ into intervals depending of size $\delta := \frac{\tau}{N} < 1$, for a positive integer N sufficiently large. For any $t \in [0, T]$, we construct $\hat{Y}^\varepsilon$ with initial value $\hat{Y}_0^\varepsilon = Y_0^\varepsilon = \eta \in \mathscr{L}$

$$\begin{aligned} d\hat{Y}^\varepsilon(t) &= \frac{1}{\varepsilon} b_2(X^\varepsilon_{t_\delta}, \hat{Y}^\varepsilon(t), \hat{Y}^\varepsilon(t-\tau))dt + \frac{1}{\sqrt{\varepsilon}}\sigma_2(X^\varepsilon_{t_\delta}, \hat{Y}^\varepsilon(t), \hat{Y}^\varepsilon(t-\tau))dW_t, \\ \hat{Y}^\varepsilon(t_\delta) &= Y^\varepsilon(t_\delta), \end{aligned}$$

where $t_\delta = \lfloor \frac{t}{\delta} \rfloor \delta$ is the nearest breakpoint preceding t and define the process $\hat{X}^\varepsilon$ by

$$d\hat{X}^\varepsilon(t) = b_1(X^\varepsilon_{t_\delta}, \hat{Y}^\varepsilon_t)dt + \sigma_1(X^\varepsilon(t-\tau))dB_t^H \tag{3.8}$$

with the initial value $\hat{X}_0^\varepsilon = \xi \in \mathscr{L}$. Then, we can derive uniform bounds $\|X^\varepsilon - \hat{X}^\varepsilon\|_{\alpha,\lambda(\tau)}$. Thirdly, based on the ergodic property of the frozen equation, we obtain appropriate control of $\|\hat{X}^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}$. Finally, we can estimate $\|X^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}$ and obtain the main result.

Step 1: A priori estimate for the solution of (3.1). We use techniques similar to those used in [32, Theorem 4.2] to give a priori estimate for the solution X^ε.

Lemma 3.6. *Suppose that* (H1), (H2) *and* (H5) *hold. Then, for any* $p \geq 1$*, there exists a constant* $C_p > 0$ *which is independent of* ε *such that*

$$\mathbb{E}[\|X^\varepsilon\|^p_{\alpha,\infty(\tau)}] \leq C_p.$$

Proof: For shortness, denote, $\Lambda := \Lambda_{\alpha,B^H} \vee 1$ and for any $\lambda \geq 1$ let

$$\begin{aligned} \|f\|_{\infty,\lambda(\tau),t} &:= \sup_{-\tau\leq s\leq t} e^{-\lambda s}|f(s)|, \\ \|f\|_{1,\lambda(\tau),t} &:= \sup_{-\tau\leq s\leq t} e^{-\lambda s}\int_{-\tau}^s \frac{|f(s)-f(r)|}{(s-r)^{\alpha+1}}dr. \end{aligned}$$

We start by estimating $\|X^\varepsilon\|_{\infty,\lambda(\tau),t}$. We have

$$\begin{aligned} \|X^\varepsilon\|_{\infty,\lambda(\tau),t} &\leq \sup_{-\tau\leq s\leq 0} e^{-\lambda s}|\xi(s)| + \sup_{0\leq s\leq t} e^{-\lambda s}|X^\varepsilon(s)| \\ &\leq \sup_{-\tau\leq s\leq 0} e^{-\lambda s}|\xi(s)| + \sup_{0\leq s\leq t} e^{-\lambda s}\left|\int_0^s b_1\left(X_r^\varepsilon, Y_r^\varepsilon\right)dr\right| \end{aligned}$$

$$
\begin{aligned}
&\quad + \sup_{0\le s\le t} e^{-\lambda s}\Big|\int_0^s \sigma_1(X^\varepsilon(r-\tau))dB_r^H\Big| \\
&=: \quad \mathbf{I}_1+\mathbf{I}_2+\mathbf{I}_3.
\end{aligned}
$$

First, for $\mathbf{I}_1, \mathbf{I}_2$, by (H1), one has

$$
\begin{aligned}
\mathbf{I}_1+\mathbf{I}_2 &\le \|\xi\|_{\alpha,\infty(-\tau,0)} + C\sup_{0\le s\le t} e^{-\lambda s}\int_0^s (1+\|X_r^\varepsilon\|_\infty)dr \\
&\le \|\xi\|_{\alpha,\infty(-\tau,0)} + C\sup_{0\le s\le t}\int_0^s e^{-\lambda(s-r)}(1+\sup_{-\tau\le q\le r} e^{-\lambda r}|X^\varepsilon(q)|)dr \\
&\le \|\xi\|_{\alpha,\infty(-\tau,0)} + C\sup_{0\le s\le t}\int_0^s e^{-\lambda(s-r)}(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})dr,
\end{aligned}
$$

where $C>0$ is a constant.

Next, for the third term $\mathbf{I}_3$, by (H2) and (2.3), we have

$$
\begin{aligned}
\mathbf{I}_3 &\le \Lambda_{\alpha,B^H}\sup_{0\le s\le t} e^{-\lambda s}\int_0^s \frac{|\sigma_1(X^\varepsilon(r-\tau))|}{r^\alpha}dr \\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t} e^{-\lambda s}\int_0^s \Big(\int_0^r \frac{|\sigma_1(X^\varepsilon(r-\tau))-\sigma_1(X^\varepsilon(q-\tau))|}{(r-q)^{1+\alpha}}dq\Big)dr \\
&\le \Lambda_{\alpha,B^H}\sup_{0\le s\le t} e^{-\lambda s}\int_{-\tau}^{s-\tau}\Big(\frac{1+|X^\varepsilon(r)|}{(r+\tau)^\alpha}+\int_{-\tau}^r \frac{|X^\varepsilon(r)-X^\varepsilon(q)|}{(r-q)^{1+\alpha}}dq\Big)dr \\
&\le \Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_{-\tau}^{s-\tau} e^{-\lambda(s-r)}\Big(\frac{1+e^{-\lambda r}|X^\varepsilon(r)|}{(r+\tau)^\alpha} \\
&\quad +e^{-\lambda r}\int_{-\tau}^r \frac{|X^\varepsilon(r)-X^\varepsilon(q)|}{(r-q)^{1+\alpha}}dq\Big)dr \\
&\le \Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_{-\tau}^{s-\tau} e^{-\lambda(s-r)}[(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})(r+\tau)^{-\alpha}+\|X^\varepsilon\|_{1,\lambda(\tau),t}]dr \\
&\le \Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_0^s e^{-\lambda(s-u+\tau)}[(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})u^{-\alpha}+\|X^\varepsilon\|_{1,\lambda(\tau),t}]du.
\end{aligned}
$$

Thus, by Lemma 2.1, it follows that

$$
\|X^\varepsilon\|_{\infty,\lambda(\tau),t} \le K\Lambda\left(1+\lambda^{\alpha-1}\|X^\varepsilon\|_{\infty,\lambda(\tau),t}+\lambda^{-1}\|X^\varepsilon\|_{1,\lambda(\tau),t}\right),
$$

with some constant K which is dependent on $\|\xi\|_{\alpha,\infty(-\tau,0)}$ and can be assumed to be greater than 1 without loss of generality.

To proceed, noting that for $t\in[-\tau,0]$, one has

$$
\int_{-\tau}^t \frac{|X^\varepsilon(t)-X^\varepsilon(s)|}{(t-s)^{1+\alpha}}ds = \int_{-\tau}^t \frac{|\xi(t)-\xi(s)|}{(t-s)^{1+\alpha}}ds,
$$

and for $t\in[0,T]$, one has

$$
\begin{aligned}
\int_{-\tau}^t \frac{|X^\varepsilon(t)-X^\varepsilon(s)|}{(t-s)^{1+\alpha}}ds &= \int_{-\tau}^0 \frac{|X^\varepsilon(t)-\xi(0)|}{(t-s)^{1+\alpha}}ds+\int_{-\tau}^0 \frac{|\xi(0)-\xi(s)|}{(-s)^{1+\alpha}}ds \\
&\quad +\int_0^t \frac{|X^\varepsilon(t)-X^\varepsilon(s)|}{(t-s)^{1+\alpha}}ds.
\end{aligned}
$$

Consequently, we have

$$
\|X^\varepsilon\|_{1,\lambda(\tau),t} \le \sup_{-\tau\le s\le 0} e^{-\lambda s}\int_{-\tau}^s \frac{|\xi(s)-\xi(r)|}{(s-r)^{1+\alpha}}dr+\sup_{0\le s\le t} e^{-\lambda s}\int_{-\tau}^0 \frac{|\xi(0)-\xi(r)|}{(-r)^{1+\alpha}}dr
$$

$$+\sup_{0\le s\le t} e^{-\lambda s}\int_{-\tau}^{0}\frac{|X^\varepsilon(s)-\xi(0)|}{(s-r)^{1+\alpha}}dr+\sup_{0\le s\le t} e^{-\lambda s}\int_{0}^{s}\frac{|X^\varepsilon(s)-X^\varepsilon(r)|}{(s-r)^{1+\alpha}}dr$$
$$=: \sum_{i=1}^{4}\mathbf{J}_i.$$

It is easy to obtain

$$\mathbf{J}_1+\mathbf{J}_2\le \|\xi\|_{\alpha,\infty(-\tau,0)}.$$

In what follows, by (2.3), using the same step as for the terms $\mathbf{I}_2$ and $\mathbf{I}_3$, we have

$$\begin{aligned}
\mathbf{J}_3 &\le \sup_{0\le s\le t}\frac{e^{-\lambda s}}{s^\alpha}|X^\varepsilon(s)-\xi(0)|\\
&\le \sup_{0\le s\le t}\frac{1}{s^\alpha}\int_0^s e^{-\lambda(s-r)}(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})dr\\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t}\frac{1}{s^\alpha}\int_0^s e^{-\lambda(s-u+\tau)}[(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})u^{-\alpha}+\|X^\varepsilon\|_{1,\lambda(\tau),t}]du\\
&\le \Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_0^s e^{-\lambda(s-u+\tau)}[(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})u^{-2\alpha}+(s-u)^{-\alpha}\|X^\varepsilon\|_{1,\lambda(\tau),t}]du\\
&\le K\Lambda(1+\lambda^{2\alpha-1}\|X^\varepsilon\|_{\infty,\lambda(\tau),t}+\lambda^{-\alpha}\|X^\varepsilon\|_{1,\lambda(\tau),t}).
\end{aligned}$$

For the term $\mathbf{J}_4$, applying Lemma 2.2, we derive that

$$\begin{aligned}
\mathbf{J}_4 &\le \sup_{0\le s\le t} e^{-\lambda s}\int_0^s\frac{\left|\int_r^s b_1(X_q^\varepsilon,Y_q^\varepsilon)dq\right|}{(s-r)^{1+\alpha}}dr\\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t} e^{-\lambda s}\bigg(\int_0^s (s-r)^{-2\alpha}|\sigma_1(X^\varepsilon(r-\tau))|dr\\
&\quad +\int_0^s (s-r)^{-\alpha}\bigg(\int_0^r\frac{|\sigma_1(X^\varepsilon(r-\tau))-\sigma_1(X^\varepsilon(q-\tau))|}{(r-q)^{1+\alpha}}dq\bigg)dr\bigg)\\
&\le \sup_{0\le s\le t} e^{-\lambda s}\int_0^s\frac{|b_1(X_q^\varepsilon,Y_q^\varepsilon)|}{(s-r)^\alpha}dr\\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t} e^{-\lambda s}\bigg(\int_{-\tau}^{s-\tau}(s-r-\tau)^{-2\alpha}(1+|X^\varepsilon(r)|)dr\\
&\quad +\int_{-\tau}^{s-\tau}(s-\tau-r)^{-\alpha}\int_{-\tau}^{r}\frac{|\sigma_1(X^\varepsilon(r))-\sigma_1(X^\varepsilon(q))|}{(r-q)^{1+\alpha}}dqdr\bigg)\\
&\le \sup_{0\le s\le t}\int_0^s e^{-\lambda(s-r)}\frac{e^{-\lambda r}(1+\sup_{0\le q\le r}|X^\varepsilon(q)|)}{(s-r)^\alpha}dr\\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_{-\tau}^{s-\tau} e^{-\lambda(s-r)}\frac{1+e^{-\lambda r}|X^\varepsilon(r)|}{(s-r-\tau)^{2\alpha}}dr\\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_{-\tau}^{s-\tau} e^{-\lambda(s-r)}(s-\tau-r)^{-\alpha}e^{-\lambda r}\int_{-\tau}^{r}\frac{|X^\varepsilon(r)-X^\varepsilon(q)|}{(r-q)^{1+\alpha}}dqdr\\
&\le \sup_{0\le s\le t}\int_0^s e^{-\lambda(s-r)}(s-r)^{-\alpha}(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})dr\\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_0^s e^{-\lambda(s-u+\tau)}(s-u)^{-2\alpha}(1+\|X^\varepsilon\|_{\infty,\lambda(\tau),t})du\\
&\quad +\Lambda_{\alpha,B^H}\sup_{0\le s\le t}\int_0^s e^{-\lambda(s-u+\tau)}(s-u)^{-\alpha}\|X^\varepsilon\|_{1,\lambda(\tau)t}dr.
\end{aligned}$$

Thus, by Lemma 2.1 again, we have

$$\|X^\varepsilon\|_{1,\lambda(\tau),t}\le K\Lambda(1+\lambda^{2\alpha-1}\|X^\varepsilon\|_{\infty,\lambda(\tau),t}+\lambda^{-\alpha}\|X^\varepsilon\|_{1,\lambda(\tau)t}). \tag{3.9}$$

Putting $\lambda = (4K\Lambda)^{\frac{1}{1-\alpha}}$, we get

$$\|X^\varepsilon\|_{\infty,\lambda(\tau),t} \leq \frac{4}{3}K\Lambda\left(1+\lambda^{-1}\|X^\varepsilon\|_{1,\lambda(\tau),t}\right). \tag{3.10}$$

Then, plugging this to the inequality (3.9) and making simple transformations, we arrive at

$$\|X^\varepsilon\|_{1,\lambda(\tau),t} \leq \frac{3}{2}K\Lambda + 2(K\Lambda)^{1/(1-\alpha)} \leq C\Lambda^{1/(1-\alpha)},$$

where $C>0$ is a constant which is independent of ε.

Substituting this into (3.10), we get

$$\|X^\varepsilon\|_{\infty,\lambda(\tau),t} \leq C\Lambda^{1/(1-\alpha)}.$$

Thus, we have

$$\begin{aligned}\|X^\varepsilon\|_{\alpha,\infty(\tau)} &\leq e^{\lambda T}\left(\|X^\varepsilon\|_{\infty,\lambda(\tau),T}+\|X^\varepsilon\|_{1,\lambda(\tau),T}\right)\\ &\leq Ce^{\Lambda^{1/(1-\alpha)}}\Lambda^{1/(1-\alpha)}\\ &\leq Ce^{\Lambda^{1/(1-\alpha)}}\left(1+\left(\Lambda_{\alpha,B^H}\right)^{1/(1-\alpha)}\right).\end{aligned}$$

Since $0<\frac{1}{1-\alpha}<2$, by (2.4), we have

$$\mathbb{E}\left[e^{(\Lambda_{\alpha,B^H})^{\vartheta}}\right] \leq C.$$

Then, the statement follows. □

Using similar techniques, we have the following remark. Here, we omit the proof.

Remark 3.7. *Suppose that* (H1)-(H5) *hold. Then, for any* $p\geq 1$*, there exist constants* $C, C_p>0$ *which are independent of* ε *such that*

$$\|\hat{X}^\varepsilon\|_{\alpha,\infty(\tau)}+\|\bar{X}\|_{\alpha,\infty(\tau)} \leq Ce^{(\Lambda_{\alpha,B^H})^{1/(1-\alpha)}}\left(1+(\Lambda_{\alpha,B^H})^{1/(1-\alpha)}\right) \tag{3.11}$$

and

$$\mathbb{E}[\|\bar{X}\|^p_{\alpha,\infty(\tau)}]+\mathbb{E}[\|\hat{X}^\varepsilon\|^p_{\alpha,\infty(\tau)}] \leq C_p. \tag{3.12}$$

Lemma 3.8. *Suppose that* (H1)-(H5) *hold. Then, for any* $p>\frac{1}{1-\alpha}$*, there exists a constant* $C_p>0$ *which is independent of* ε *such that*

$$\sup_{t\in[0,T]}\mathbb{E}[\|X^\varepsilon_t - X^\varepsilon_{t_\delta}\|^p_\infty] \leq C_p\delta^{p(1-\alpha)-1}.$$

Proof: We start to estimate $|X^\varepsilon(t)-X^\varepsilon(s)|$. For any $s,t\in[0,T]$, there exists $C>0$ such that

$$\begin{aligned}|X^\varepsilon(t)-X^\varepsilon(s)| &\leq \left|\int_s^t b_1\left(X^\varepsilon_r, Y^\varepsilon_r\right)dr\right| + \left|\int_s^t \sigma_1(X^\varepsilon(r-\tau))dB^H_r\right|\\ &\leq C\int_s^t(1+\|X^\varepsilon_r\|_\infty)dr + C\Lambda_{\alpha,B^H}\int_s^t\frac{|\sigma_1(X^\varepsilon(r-\tau))|}{(r-s)^\alpha}dr\\ &\quad +C\Lambda_{\alpha,B^H}\int_s^t\int_s^r\frac{|\sigma_1(X^\varepsilon(r-\tau))-\sigma_1(X^\varepsilon(q-\tau))|}{(r-q)^{1+\alpha}}dqdr\\ &\leq C\int_s^t(1+\sup_{-\tau\leq u\leq r}|X^\varepsilon(u)|)dr\\ &\quad +C\Lambda_{\alpha,B^H}\int_s^t\frac{(1+\sup_{-\tau\leq u\leq r}|X^\varepsilon(u)|)}{(r-s)^\alpha}dr\end{aligned}$$

$$
\begin{aligned}
&\quad + C\Lambda_{\alpha,B^H}\int_s^t\int_s^r \frac{|X^\varepsilon(r-\tau)-X^\varepsilon(q-\tau)|}{(r-q)^{1+\alpha}}dqdr \\
&\leq C(1+\|X^\varepsilon\|_{\alpha,\infty(\tau)})|t-s| + C\Lambda_{\alpha,B^H}\int_s^t \frac{(1+\|X^\varepsilon\|_{\alpha,\infty(\tau)})}{(r-s)^\alpha}dr \\
&\quad + C\Lambda_{\alpha,B^H}\int_{s-\tau}^{t-\tau}\int_{s-\tau}^v \frac{|X^\varepsilon(v)-X^\varepsilon(u)|}{(v-u)^{1+\alpha}}dudv \\
&\leq C(1+\Lambda_{\alpha,B^H})(1+\|X^\varepsilon\|^p_{\alpha,\infty(\tau)})|t-s|^{1-\alpha}. \qquad (3.13)
\end{aligned}
$$

Note that the same conclusion holds for $|\hat{X}^\varepsilon(t)-\hat{X}^\varepsilon(s)|$ and $|\bar{X}(t)-\bar{X}(s)|$.

Then, observe that

$$
\begin{aligned}
\mathbb{E}[\|X_t^\varepsilon - X_{t_\delta}^\varepsilon\|_\infty^p] &\leq \mathbb{E}\Big[\sum_{m=0}^{N-1}\sup_{-(m+1)\delta\leq\theta\leq -m\delta}|X^\varepsilon(t+\theta)-X^\varepsilon(t_\delta+\theta)|^p\Big] \\
&\leq N\max_{m=0,\cdots,N-1}\mathbb{E}\Big[\sup_{-(m+1)\delta\leq\theta\leq -m\delta}|X^\varepsilon(t+\theta)-X^\varepsilon(t_\delta+\theta)|^p\Big] \\
&=: N\max_{m=0,\cdots,N-1}J_p(t,m,\delta),
\end{aligned}
$$

where $N=\frac{\tau}{\delta}$ by the definition of δ. For any $t\in[0,T]$ and any $\theta\in[-\tau,0]$, there exist $k,m\geq 0$ such that $t\in[k\delta,(k+1)\delta)$ and $\theta\in[-(m+1)\delta,-m\delta]$. Thus, one has

$$
t+\theta\in[(k-m-1)\delta,(k+1-m)\delta] \quad \text{and} \quad t_\delta+\theta\in[(k-m-1)\delta,(k-m)\delta].
$$

We consider three cases.

Case 1. $m\leq k-1$. Involving Hölder's inequality, by (H1), (H2), (3.13) and Lemma 3.6, there exists a constant $C_p>0$ such that

$$
\begin{aligned}
J_p(t,m,\delta) &\leq C_p\mathbb{E}\Big[\sup_{-(m+1)\delta\leq\theta\leq -m\delta}\Big|\int_{k\delta+\theta}^{t+\theta}b_1(X_r^\varepsilon,Y_r^\varepsilon)dr + \int_{k\delta+\theta}^{t+\theta}\sigma_1(X^\varepsilon(r-\tau))dB_r^H\Big|^p\Big] \\
&\leq C_p\delta^{p-1}\int_{k\delta-(m+1)\delta}^{t-m\delta}\mathbb{E}\big[|b_1(X_r^\varepsilon,Y_r^\varepsilon)|^p\big]dr \\
&\quad + C_p\mathbb{E}\Big[\Big|\int_{k\delta-(m+1)\delta}^{t-(m+1)\delta}\sigma_1(X^\varepsilon(r-\tau))dB_r^H\Big|^p\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{-(m+1)\delta\leq\theta\leq -m\delta}\Big|\int_{t-(m+1)\delta}^{t+\theta}\sigma_1(X^\varepsilon(r-\tau))dB_r^H\Big|^p\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{-(m+1)\delta\leq\theta\leq -m\delta}\Big|\int_{k\delta-(m+1)\delta}^{k\delta+\theta}\sigma_1(X^\varepsilon(r-\tau))dB_r^H\Big|^p\Big] \\
&\leq C_p\delta^p + C_p\delta^{p(1-\alpha)}\mathbb{E}\Big[\big(1+\Lambda_{\alpha,B^H}\big)^p(1+\|X^\varepsilon\|_{\alpha,\infty(\tau)})^p\Big] \\
&\leq C_p\delta^{p(1-\alpha)}.
\end{aligned}
$$

Case 2. $m\geq k-1$. By (H5), there exists a constant $C_p>0$ such that

$$
|X^\varepsilon(t+\theta)-X^\varepsilon(t_\delta+\theta)|^p = |\xi(t+\theta)-\xi(t_\delta+\theta)|^p \leq C_p\delta^p.
$$

Case 3. $m=k$. By Hölder's inequality, we deduce from (H1), (H2) and (3.13) that

$$
\begin{aligned}
J_p(t,m,\delta) &= \mathbb{E}\Big[\sup_{-(m+1)\delta\leq\theta\leq -m\delta}|X^\varepsilon(t+\theta)-X^\varepsilon(t_\delta+\theta)|^p\Big] \\
&\leq C_p\delta^p + C_p\mathbb{E}\Big[\sup_{-(m+1)\delta\leq\theta\leq -m\delta}|X^\varepsilon(t+\theta)-X^\varepsilon(t_\delta+\theta)|^p\mathbf{1}_{\{t+\theta>0\}}\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\leq C_p\delta^p + C_p\mathbb{E}\Big[\sup_{-t\leq\theta\leq -k\delta}\Big|\int_0^{t+\theta} b_1(X_r^\varepsilon, Y_r^\varepsilon)dr\Big|^p\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{-t\leq\theta\leq -k\delta}\Big|\int_0^{t+\theta}\sigma_1(X^\varepsilon(r-\tau))dB_r^H\Big|^p\Big] \\
&\leq C_p\delta^{p(1-\alpha)},
\end{aligned}
$$

where $C_p > 0$ is a constant. Thus, the desired assertion is finished by taking the discussions above into account. □

To derive uniform bounds $\|X^\varepsilon - \hat{X}^\varepsilon\|_{\alpha,\lambda(\tau)}$, we will also need following estimate $\|Y_t^\varepsilon - \hat{Y}_t^\varepsilon\|_\infty$ which follows from [1, Lemma 4.2, Lemma 4.3] and to make our paper self-contained, the proofs will be given in Appendix.

Lemma 3.9. *Under* (H1)-(H5)*, for any $p > \frac{2}{1-\alpha}$, there exists $\beta > 0$ which is independent of ε such that*

$$
\sup_{t\in[0,T]}\mathbb{E}[\|Y_t^\varepsilon - \hat{Y}_t^\varepsilon\|_\infty^p] \leq C_p\varepsilon^{-1}\delta^{p(1-\alpha)-1}e^{\beta\frac{\delta}{\varepsilon}}. \tag{3.14}
$$

where $C_p > 0$ is independent of ε.

Lemma 3.10. *Under* (H1)-(H5)*, for any $p > \frac{2}{1-\alpha}$, we have*

$$
\sup_{t\in[0,T]}\mathbb{E}[\|Y_t^\varepsilon\|_\infty^p + \|\hat{Y}_t^\varepsilon\|_\infty^p] \leq C_p,
$$

where C_p is a constant which is independent of ε.

Lemma 3.11. *Suppose that* (H1)-(H5) *hold. Then, for any $p > \frac{2}{1-\alpha}$, there exists a constant $C_p > 0$ such that*

$$
\mathbb{E}[\|\hat{X}^\varepsilon - X^\varepsilon\|_{\alpha,\lambda(\tau)}^p] \leq C_p\delta^{p(1-\alpha)}(1 + \varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}}).
$$

Proof: From (3.1) and (3.8), we have

$$
\begin{aligned}
\mathbb{E}[\|\hat{X}^\varepsilon - X^\varepsilon\|_{\alpha,\lambda(\tau)}^p] &\leq \mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}|\hat{X}^\varepsilon(t) - X^\varepsilon(t)|^p\Big] \\
&\quad + \mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big(\int_0^t \frac{|\hat{X}^\varepsilon(t) - X^\varepsilon(t) - \hat{X}^\varepsilon(s) + X^\varepsilon(s)|}{(t-s)^{\alpha+1}}ds\Big)^p\Big] \\
&\quad + \mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big(\int_{-\tau}^0 \frac{|\hat{X}^\varepsilon(t) - X^\varepsilon(t) - \hat{X}^\varepsilon(s) + X^\varepsilon(s)|}{(t-s)^{\alpha+1}}ds\Big)^p\Big] \\
&=: \mathbf{A}_1 + \mathbf{A}_2 + \mathbf{A}_3.
\end{aligned}
$$

Firstly, for $\mathbf{A}_1, \mathbf{A}_2$, by (H1) and (H2), Lemma 2.2, Lemma 3.8 and Lemma 3.9, we have

$$
\begin{aligned}
\mathbf{A}_1 + \mathbf{A}_2 &\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (b_1(X_s^\varepsilon, Y_s^\varepsilon) - b_1(X_s^\varepsilon, \hat{Y}_s^\varepsilon))ds\Big\|_\alpha^p\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (b_1(X_s^\varepsilon, \hat{Y}_s^\varepsilon) - b_1(X_{s_\delta}^\varepsilon, \hat{Y}_s^\varepsilon))ds\Big\|_\alpha^p\Big] \\
&\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big(\int_0^t \frac{|b_1(X_s^\varepsilon, Y_s^\varepsilon) - b_1(X_s^\varepsilon, \hat{Y}_s^\varepsilon)|}{(t-s)^\alpha}ds\Big)^p\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big(\int_0^t \frac{|b_1(X_s^\varepsilon, \hat{Y}_s^\varepsilon) - b_1(X_{s_\delta}^\varepsilon, \hat{Y}_s^\varepsilon)|}{(t-s)^\alpha}ds\Big)^p\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\leq C_p \int_0^T (\mathbb{E}[\|X_s^\varepsilon - X_{s_\delta}^\varepsilon\|_\infty^p + \|Y_s^\varepsilon - \hat{Y}_s^\varepsilon\|_\infty^p])ds \\
&\leq C_p \delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}}).
\end{aligned}
$$

Secondly, for $\mathbf{A}_3$, we have

$$
\begin{aligned}
\mathbf{A}_3 &\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big(\int_{-\tau}^0 \frac{|\hat{X}^\varepsilon(t) - X^\varepsilon(t) - \hat{X}^\varepsilon(0) + X^\varepsilon(0)|}{(t-s)^{\alpha+1}}ds\Big)^p\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big(\int_{-\tau}^0 \frac{|\hat{X}^\varepsilon(0) - X^\varepsilon(0) - \xi(s) + \xi(s)|}{(-s)^{\alpha+1}}ds\Big)^p\Big] \\
&\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big(\int_{-\tau}^0 \frac{|\hat{X}^\varepsilon(t) - X^\varepsilon(t)|}{(t-s)^{\alpha+1}}ds\Big)^p\Big] \\
&\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big|\int_0^t (b_1(X_s^\varepsilon,\hat{Y}_s^\varepsilon) - b_1(X_{s_\delta}^\varepsilon,\hat{Y}_s^\varepsilon))ds\Big|^p\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big|\int_0^t (b_1(X_s^\varepsilon,Y_s^\varepsilon) - b_1(X_s^\varepsilon,\hat{Y}_s^\varepsilon))ds\Big|^p\Big] \\
&\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}e^{-\lambda t}\int_0^t \|X_s^\varepsilon - X_{s_\delta}^\varepsilon\|_\infty^p ds\Big] \\
&\quad + C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}e^{-\lambda t}\int_0^t \|Y_s^\varepsilon - \hat{Y}_s^\varepsilon\|_\infty^p ds\Big] \\
&\leq C_p \int_0^T (\mathbb{E}[\|X_s^\varepsilon - X_{s_\delta}^\varepsilon\|_\infty^p + \|Y_s^\varepsilon - \hat{Y}_s^\varepsilon\|_\infty^p])ds \\
&\leq C_p \delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}}).
\end{aligned}
$$

Thus, we have

$$
\mathbb{E}[\|\hat{X}^\varepsilon - X^\varepsilon\|_{\alpha,\lambda(\tau)}^p] \leq C_p\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}}).
$$

This completes the proof. □

Step 2: The estimate for $\|\bar{X} - \hat{X}^\varepsilon\|_{\alpha,\lambda(\tau)}$.

For each $R > 1$, we define the following stopping time τ_R,

$$
\tau_R := \inf\{t \geq 0 : \|B^H\|_{\alpha,0,t} \geq R\} \wedge T. \tag{3.15}
$$

Lemma 3.12. *Suppose that* (H1)-(H5) *hold. Then, for any* $p > \frac{2}{1-\alpha}$, *there exist positive constants* C_p *and* $C_{p,R}$ *such that*

$$
\mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}^p] \leq C_p\sqrt{R^{-1}\mathbb{E}[\|B^H\|_{\alpha,0,T}^2]} + C_{p,R}\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}}) + C_{p,R}(\delta + \varepsilon^{p'}\delta^{-p'}),
$$

where $p' \in (1,2)$.

Proof: From (3.7) and (3.8), we have

$$
\begin{aligned}
\mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}^p] &\leq \mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}^p \mathbf{1}_{\{\tau_R < T\}}] \\
&\quad + \mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}^p \mathbf{1}_{\{\tau_R \geq T\}}].
\end{aligned} \tag{3.16}
$$

For the first term on the right-hand side of inequality (3.16), by Chebyshev's inequality, we have

$$
\mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}^p \mathbf{1}_{\{\tau_R < T\}}] \leq \big(\mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|_{\alpha,\lambda(\tau)}^{2p}]\big)^{\frac{1}{2}}\big(\mathbb{P}(\tau_R < T)\big)^{\frac{1}{2}}. \tag{3.17}
$$

It is easy to obtain

$$\mathbb{P}\big(\tau_R < T\big) \leq \mathbb{P}\big(\|B^H\|_{\alpha,0,T} \geq R\big) \leq R^{-1}\mathbb{E}[\|B^H\|^2_{\alpha,0,T}].$$

Because $\|B^H\|_{\alpha,0,T}$ has moments of all order, see Lemma 7.5 in Nualart and Răşcanu [23], thus we have

$$\lim_{R\to\infty} R^{-1}\mathbb{E}[\|B^H\|^2_{\alpha,0,T}] = 0.$$

Then, summing up all bounds we obtain

$$\mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_{\{\tau_R<T\}}] \leq C_p\sqrt{R^{-1}\mathbb{E}[\|B^H\|^2_{\alpha,0,T}]}.$$

For the second term on the right-hand side of inequality (3.16), we have

$$\begin{aligned}
\mathbb{E}[\|\hat{X}^\varepsilon - \bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D] \quad &\leq \quad C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}|\hat{X}^\varepsilon(t) - \bar{X}(t)|^p\mathbf{1}_D\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big(\int_0^t \frac{|\hat{X}^\varepsilon(t) - \bar{X}(t) - \hat{X}^\varepsilon(s) + \bar{X}(s)|}{(t-s)^{\alpha+1}}ds\Big)^p\mathbf{1}_D\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big(\int_{-\tau}^0 \frac{|\hat{X}^\varepsilon(t) - \bar{X}(t) - \hat{X}^\varepsilon(s) + \bar{X}(s)|}{(t-s)^{\alpha+1}}ds\Big)^p\mathbf{1}_D\Big] \\
&=: \quad \mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_3,
\end{aligned}$$

where $D := \{\|B^H\|_{\alpha,0,T} \leq R\}$.

For the first two terms, we have

$$\begin{aligned}
\mathbf{B}_1 + \mathbf{B}_2 \quad &\leq \quad C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (b_1(X^\varepsilon_{s_\delta}, \hat{Y}^\varepsilon_s) - \bar{b}_1(X^\varepsilon_{s_\delta}))ds\Big\|^p_\alpha\mathbf{1}_D\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (\bar{b}_1(X^\varepsilon_{s_\delta}) - \bar{b}_1(X^\varepsilon_s))ds\Big\|^p_\alpha\mathbf{1}_D\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (\bar{b}_1(X^\varepsilon_s) - \bar{b}_1(\hat{X}^\varepsilon_s))ds\Big\|^p_\alpha\mathbf{1}_D\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (\bar{b}_1(\hat{X}^\varepsilon_s) - \bar{b}_1(\bar{X}_s))ds\Big\|^p_\alpha\mathbf{1}_D\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (\sigma_1(\hat{X}^\varepsilon(s-\tau)) - \sigma_1(\bar{X}(s-\tau)))dB^H_s\Big\|^p_\alpha\mathbf{1}_D\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} e^{-\lambda t}\Big\|\int_0^t (\sigma_1(X^\varepsilon(s-\tau)) - \sigma_1(\hat{X}^\varepsilon(s-\tau)))dB^H_s\Big\|^p_\alpha\mathbf{1}_D\Big] \\
&=: \quad \sum_{i=1}^6 \mathbf{C}_i.
\end{aligned}$$

It is easy to know

$$\begin{aligned}
\mathbf{C}_1 \quad &\leq \quad C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big|\int_0^t (b_1(X^\varepsilon_{s_\delta}, \hat{Y}^\varepsilon_s) - \bar{b}_1(X^\varepsilon_{s_\delta}))ds\Big|^p\Big] \\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big(\int_0^t \frac{|\int_s^t (b_1(X^\varepsilon_{r_\delta}, \hat{Y}^\varepsilon_r) - \bar{b}_1(X^\varepsilon_{r_\delta}))dr|}{(t-s)^{1+\alpha}}ds\Big)^p\Big] \\
&=: \quad \mathbf{C}_{11} + \mathbf{C}_{12}.
\end{aligned}$$

For $\mathbf{C}_{11}$, we have

$$\mathbf{C}_{11} \quad \leq \quad C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big|\sum_{k=0}^{\lfloor \frac{t}{\delta}\rfloor-1}\int_{k\delta}^{(k+1)\delta} (b_1(X^\varepsilon_{k\delta}, \hat{Y}^\varepsilon_s) - \bar{b}_1(X^\varepsilon_{k\delta}))ds\Big|^p\Big]$$

$$
\begin{aligned}
&+C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big|\int_{t_\delta}^t (b_1(X^\varepsilon_{s_\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{s_\delta}))ds\Big|^p\Big]\\
\leq\;& C_p\delta^p + C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\big(\lfloor\frac{t}{\delta}\rfloor\big)^{p-1}\sum_{k=0}^{\lfloor\frac{t}{\delta}\rfloor-1}\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{k\delta}))ds\Big|^p\Big]\\
\leq\;& C_p\delta^p + \frac{C_p}{\delta^p}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\mathbb{E}\Big[\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{k\delta}))ds\Big|^p\Big].
\end{aligned}
$$

Then, for $\mathbf{C}_{12}$, by Hölder's inequality and the fact that $\alpha<\frac{1}{2}$, we have

$$
\begin{aligned}
\mathbf{C}_{12}\;\leq\;& C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big(\int_0^t\frac{|\int_s^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|}{(t-s)^{1+\alpha}}ds\Big)^p\Big]\\
\leq\;& C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\Big(\int_0^t(t-s)^{\frac{(-1-\alpha)p+\frac{3}{2}+\alpha}{p-1}}ds\Big)^{p-1}\\
&\times\int_0^t\frac{|\int_s^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}ds\Big]\\
\leq\;& C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{|\int_s^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\ell^c}ds\Big]\\
&+C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{|\int_s^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\ell}ds\Big]\\
=:\;& \mathbf{C}_{121}+\mathbf{C}_{122},
\end{aligned}
$$

where $\mathbf{1}$. is an indicator function, $\ell:=\{t<(\lfloor\frac{s}{\delta}\rfloor+2)\delta\}$ and $\ell^c:=\{t\geq(\lfloor\frac{s}{\delta}\rfloor+2)\delta\}$.

By (H1) and the fact that $\lfloor\lambda_1\rfloor-\lfloor\lambda_2\rfloor\leq\lambda_1-\lambda_2+1$, for $\lambda_1\geq\lambda_2\geq 0$, we have

$$
\begin{aligned}
\mathbf{C}_{121}\;\leq\;& C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{|\int_s^{(\lfloor\frac{s}{\delta}\rfloor+1)\delta}(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\ell^c}ds\Big]\\
&+C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{|\int_{t_\delta}^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\ell^c}ds\Big]\\
&+C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{(\lfloor\frac{t}{\delta}\rfloor-\lfloor\frac{s}{\delta}\rfloor-1)^{p-1}}{(t-s)^{\frac{3}{2}+\alpha}}\\
&\times\sum_{k=\lfloor\frac{s}{\delta}\rfloor+1}^{\lfloor\frac{t}{\delta}\rfloor-1}\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Big|^p\mathbf{1}_{\ell^c}ds\Big]\\
\leq\;& C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{\big((\lfloor\frac{s}{\delta}\rfloor+1)\delta-s\big)^{p-1}\int_s^{(\lfloor\frac{s}{\delta}\rfloor+1)\delta}|(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))|^pdr}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\ell^c}ds\Big]\\
&+C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{\big(t-t_\delta\big)^{p-1}\int_{t_\delta}^t|(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))|^pdr}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\ell^c}ds\Big]\\
&+\frac{C_p}{\delta^{p-1}}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t(t-s)^{p-\frac{3}{2}-\alpha}\sum_{k=\lfloor\frac{s}{\delta}\rfloor+1}^{\lfloor\frac{t}{\delta}\rfloor-1}\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Big|^p\mathbf{1}_{\ell^c}ds\Big]\\
\leq\;& C_p\delta^{p-1}+\frac{C_p}{\delta^p}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\mathbb{E}\Big[\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Big|^p\Big].
\end{aligned}
$$

For $\mathbf{C}_{122}$, set $\jmath:=\{\lfloor\frac{t}{\delta}\rfloor>1\}$ and $\jmath^c:=\{\lfloor\frac{t}{\delta}\rfloor\leq 1\}$, by (H1) and the fact that $t-s<\lfloor\frac{s}{\delta}\rfloor\delta-s+2\delta\leq 2\delta$, we have

$$
\mathbf{C}_{122}\;\leq\; C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^{(\lfloor\frac{t}{\delta}\rfloor-1)\delta}\frac{|\int_s^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\jmath\bigcap\ell}ds\Big]
$$

$$
\begin{aligned}
&+C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{(\lfloor\frac{t}{\delta}\rfloor-1)\delta}^{t}\frac{|\int_s^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\mathcal{J}\bigcap\ell}ds\Big]\\
&+C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t\frac{|\int_s^t(b_1(X^\varepsilon_{r_\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{r_\delta}))dr|^p}{(t-s)^{\frac{3}{2}+\alpha}}\mathbf{1}_{\mathcal{J}^c\bigcap\ell}ds\Big]\\
\leq\;& C_p\delta^p\sup_{t\in[0,T]}\Big(\int_0^{(\lfloor\frac{t}{\delta}\rfloor-1)\delta}(t-s)^{p-\frac{3}{2}-\alpha}\mathbf{1}_{\mathcal{J}\bigcap\ell}ds\Big)\\
&+C_p\sup_{t\in[0,T]}\Big(\int_{(\lfloor\frac{t}{\delta}\rfloor-1)\delta}^{t}(t-s)^{p-\frac{3}{2}-\alpha}\mathbf{1}_{\mathcal{J}\bigcap\ell}ds\Big)\\
&+C_p\sup_{t\in[0,T]}\Big(\int_0^t(t-s)^{p-\frac{3}{2}-\alpha}\mathbf{1}_{\mathcal{J}^c\bigcap\ell}ds\Big)\\
\leq\;& C_p\delta^{p-\frac{1}{2}-\alpha}.
\end{aligned}
$$

Thus, for any $p'\in(1,2)$, we have

$$
\begin{aligned}
\mathbf{C}_1\;\leq\;& C_p\delta^{p-1}+\frac{C_p}{\delta^p}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\Bigg\{\Bigg(\mathbb{E}\Bigg[\Bigg|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Bigg|^2\Bigg]\Bigg)^{\frac{p'}{2}}\\
&\times\Bigg(\mathbb{E}\Bigg[\Bigg|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Bigg|^{\frac{2(p-p')}{2-p'}}\Bigg]\Bigg)^{\frac{2-p'}{2}}\Bigg\}\\
\leq\;& C_p\delta^{p-1}+\frac{C_p}{\delta^p}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\Bigg\{\Bigg(\mathbb{E}\Bigg[\Bigg|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Bigg|^2\Bigg]\Bigg)^{\frac{p'}{2}}\\
&\times\Bigg(\delta^{\frac{2(p-p')}{2-p'}-1}\mathbb{E}\Bigg[\int_{k\delta}^{(k+1)\delta}|(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta})|^{\frac{2(p-p')}{2-p'}}dr\Bigg]\Bigg)^{\frac{2-p'}{2}}\Bigg\}\\
\leq\;& C_p\delta^{p-1}+\frac{C_p}{\delta^{p'}}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\Bigg\{\Bigg(\mathbb{E}\Bigg[\Bigg|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Bigg|^2\Bigg]\Bigg)^{\frac{p'}{2}}\\
\leq\;& C_p\delta^{p-1}+\frac{C_p}{\delta^{p'}}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\Bigg(\int_0^{\frac{\delta}{\varepsilon}}\int_\zeta^{\frac{\delta}{\varepsilon}}\mathcal{J}_k(s,\zeta)dsd\zeta\Bigg)^{\frac{p'}{2}},
\end{aligned}
\tag{3.18}
$$

where $0\leq\zeta\leq s\leq\frac{\delta}{\varepsilon}$, and

$$
\mathcal{J}_k(s,\zeta)=\mathbb{E}[\langle b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_{s\varepsilon+k\delta})-\bar{b}_1(X^\varepsilon_{k\delta}),b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_{\zeta\varepsilon+k\delta})-\bar{b}_1(X^\varepsilon_{k\delta})\rangle]. \tag{3.19}
$$

Now, by the construction of $\hat{Y}^\varepsilon$ and a time shift transformation, for any fixed k and $s\in[0,\delta]$, we have

$$
\begin{aligned}
\hat{Y}^\varepsilon(s+k\delta)\;=\;&\hat{Y}^\varepsilon(k\delta)+\frac{1}{\varepsilon}\int_{k\delta}^{k\delta+s}b_2(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon(r),\hat{Y}^\varepsilon(r-\tau))dr\\
&+\frac{1}{\sqrt{\varepsilon}}\int_{k\delta}^{k\delta+s}\sigma_2(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon(r),\hat{Y}^\varepsilon(r-\tau))dW_r\\
=\;&\hat{Y}^\varepsilon(k\delta)+\frac{1}{\varepsilon}\int_0^s b_2\big(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon(r+k\delta),\hat{Y}^\varepsilon(r+k\delta-\tau)\big)dr\\
&+\frac{1}{\sqrt{\varepsilon}}\int_0^s\sigma_2\big(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon(r+k\delta),\hat{Y}^\varepsilon(r+k\delta-\tau)\big)dW^*_r,
\end{aligned}
$$

where $W^*_t=W_{t+k\delta}-W_{k\delta}$ is the shift version of W_t, and hence they have the same distribution.

For fixed $\varepsilon>0$ and $r\geq0$, let

$$
Y^{X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_{k\delta}}\big(\frac{r}{\varepsilon}+\theta\big)=\hat{Y}^\varepsilon(r+k\delta+\theta),\qquad\theta\in[-\tau,0].
$$

Let $\bar{W}$ be a Wiener process and independent of W. Construct a process $Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}$ by means of

$$\begin{aligned} Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(\frac{s}{\varepsilon}) &= \hat{Y}^{\varepsilon}(k\delta)+\int_0^{\frac{s}{\varepsilon}} b_2\big(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(r),Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(r-\tau)\big)dr \\ &\quad +\int_0^{\frac{s}{\varepsilon}} \sigma_2\big(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(r),Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(r-\tau)\big)d\bar{W}_r \\ &= \hat{Y}^{\varepsilon}(k\delta)+\frac{1}{\varepsilon}\int_0^{s} b_2\big(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(\frac{r}{\varepsilon}),Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(\frac{r}{\varepsilon}-\tau)\big)dr \\ &\quad +\frac{1}{\sqrt{\varepsilon}}\int_0^{s} \sigma_2\big(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(\frac{r}{\varepsilon}),Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}(\frac{r}{\varepsilon}-\tau)\big)d\bar{\bar{W}}^{\varepsilon}_r, \end{aligned} \tag{3.20}$$

where $\bar{\bar{W}}^{\varepsilon}_t=\sqrt{\varepsilon}\bar{W}_{t/\varepsilon}$ is the scaled version of $\bar{W}_t$. Because both W^* and $\bar{\bar{W}}$ are independent of $(X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta})$, by comparison, yields

$$\big(X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{s+k\delta}\big)_{s\in[0,\delta)}\sim\big(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}_{\frac{s}{\varepsilon}}\big)_{s\in[0,\delta)}, \tag{3.21}$$

where $\sim$ denotes coincidence in distribution sense.

Thus, for $s\in[0,\delta)$, from (3.19), we have

$$\mathcal{J}_k(s,\zeta)=\mathbb{E}[\langle b_1(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}_s)-\bar{b}_1(X^{\varepsilon}_{k\delta}),b_1(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}_{\zeta})-\bar{b}_1(k\delta,X^{\varepsilon}_{k\delta})\rangle].$$

Let $\mathscr{M}^{\varepsilon}_{k\delta}$ be the σ-field generated by $X^{\varepsilon}_{k\delta}$ and $\hat{Y}^{\varepsilon}_{k\delta}$ that is independent of $\{Y^{\chi,\eta}_r:r\geq 0\}$. By adopting the approach in [24, Theorem 7.1.2] . We can show

$$\begin{aligned} \mathcal{J}_k(s,\zeta) &= \mathbb{E}\big[\mathbb{E}[\langle b_1(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}_s)-\bar{b}_1(X^{\varepsilon}_{k\delta}),b_1(X^{\varepsilon}_{k\delta},Y^{X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta}}_{\zeta})-\bar{b}_1(X^{\varepsilon}_{k\delta})\rangle|\mathscr{M}^{\varepsilon}_{k\delta}]\big] \\ &= \mathbb{E}\big[\mathcal{J}(s,\zeta,\chi,\eta)|_{(\chi,\eta)=(X^{\varepsilon}_{k\delta},\hat{Y}^{\varepsilon}_{k\delta})}\big], \end{aligned}$$

which, with the aid of Lemma 3.3, yields

$$\mathcal{J}_k(s,\zeta)\leq C(1+\mathbb{E}\big[\|X^{\varepsilon}_{k\delta}\|^2_{\infty}]+\mathbb{E}\big[\|\hat{Y}^{\varepsilon}_{k\delta}\|^2_{\infty}])e^{-\frac{\rho}{2}(s-\zeta)},$$

where $C>0$ is a constant which is independent of $k,\varepsilon,\delta,s,\zeta$.

Then, by (3.18), one has

$$\mathbf{C}_1\leq C_p(\varepsilon^{p'}\delta^{-p'}+\delta^{p-1}).$$

Next, by Lemma 3.4, Lemma 3.8 and Lemma 3.11, it is easy to obtain

$$\begin{aligned} \sum_{i=2}^{4}\mathbf{C}_i &\leq C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}e^{-\lambda t}\int_0^t(t-s)^{-\alpha}|\bar{b}_1(X^{\varepsilon}_{s_\delta})-\bar{b}_1(X^{\varepsilon}_s)|^p ds\mathbf{1}_D\bigg] \\ &\quad +C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}e^{-\lambda t}\int_0^t(t-s)^{-\alpha}|\bar{b}_1(X^{\varepsilon}_s)-\bar{b}_1(\hat{X}^{\varepsilon}_s)|^p ds\mathbf{1}_D\bigg] \\ &\quad +C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}e^{-\lambda t}\int_0^t(t-s)^{-\alpha}|\bar{b}_1(\hat{X}^{\varepsilon}_s)-\bar{b}_1(\bar{X}_s)|^p\mathbf{1}_D ds\bigg] \\ &\leq C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}e^{-\lambda t}\Big(\int_0^t(t-s)^{-\frac{p}{p-1}\alpha}ds\Big)^{p-1}\int_0^t\|X^{\varepsilon}_{s_\delta}-X^{\varepsilon}_s\|^p_{\infty}ds\bigg] \\ &\quad +C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}\int_0^t e^{-\lambda(t-s)}(t-s)^{-\alpha}\sup_{-\tau\leq r\leq s}e^{-\lambda s}|\hat{X}^{\varepsilon}(r)-X^{\varepsilon}(r)|^p\mathbf{1}_D ds\bigg] \\ &\quad +C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}\int_0^t e^{-\lambda(t-s)}(t-s)^{-\alpha}\sup_{-\tau\leq r\leq s}e^{-\lambda s}|\hat{X}^{\varepsilon}(r)-\bar{X}(r)|^p\mathbf{1}_D ds\bigg] \end{aligned}$$

$$
\begin{aligned}
&\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t \|X^\varepsilon_{s_\delta}-X^\varepsilon_s\|^p_\infty ds\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{-\tau\leq t\leq T} e^{-\lambda t}|\hat{X}^\varepsilon(t)-X^\varepsilon(t)|^p\mathbf{1}_D\Big]\Big[\sup_{t\in[0,T]}\int_0^t e^{-\lambda(t-s)}(t-s)^{-\alpha}ds\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{-\tau\leq t\leq T} e^{-\lambda t}|\hat{X}^\varepsilon(t)-\bar{X}(t)|^p\mathbf{1}_D\Big]\Big[\sup_{t\in[0,T]}\int_0^t e^{-\lambda(t-s)}(t-s)^{-\alpha}ds\Big]\\
&\leq C_p\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}})+C_p\lambda^{\alpha-1}\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D]. \qquad (3.22)
\end{aligned}
$$

For $\mathbf{C}_5$, by Lemma 2.2, we have

$$
\begin{aligned}
\mathbf{C}_5 &\leq C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t e^{-\lambda t}[(t-r)^{-2\alpha}+r^{-\alpha}]\|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))\|^p_\alpha\mathbf{1}_D dr\Big]\\
&\leq C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t e^{-\lambda t}[(t-r)^{-2\alpha}+r^{-\alpha}]|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))|^p\mathbf{1}_D dr\Big]\\
&\quad +C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t e^{-\lambda t}[(t-r)^{-2\alpha}+r^{-\alpha}]\Big(\int_0^r (r-q)^{-1-\alpha}\\
&\qquad \times|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))-\sigma_1(\hat{X}^\varepsilon(q-\tau))+\sigma_1(\bar{X}(q-\tau))|dq\Big)^p\mathbf{1}_D dr\Big]\\
&=: \mathbf{C}_{51}+\mathbf{C}_{52}.
\end{aligned}
$$

For $\mathbf{C}_{51}$, by Hölder inequality and Lemma 2.1, we have

$$
\begin{aligned}
\mathbf{C}_{51} &\leq C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t e^{-\lambda(t-r)}[(t-r)^{-2\alpha}+r^{-\alpha}]\sup_{-\tau\leq q\leq r} e^{-\lambda r}|\hat{X}^\varepsilon(q)-\bar{X}(q)|^p\mathbf{1}_D dr\Big]\\
&\leq C_{p,R}\lambda^{2\alpha-1}\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D].
\end{aligned}
$$

By (H2) and Lemma 7.1 in Nualart and Răşcanu [23], there exists a constant $C>0$ such that

$$
\begin{aligned}
|\sigma(x_1)-\sigma(x_2)-\sigma(x_3)+\sigma(x_4)| &\leq C|x_1-x_2-x_3+x_4|+C|x_1-x_3|\\
&\quad \times(|x_1-x_2|+|x_3-x_4|). \qquad (3.23)
\end{aligned}
$$

Then, for $\mathbf{C}_{52}$, we have

$$
\begin{aligned}
\mathbf{C}_{52} &\leq C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_0^t e^{-\lambda t}[(t-r)^{-2\alpha}+r^{-\alpha}]\Big(\int_{-\tau}^{r-\tau}(r-\tau-u)^{-1-\alpha}\\
&\qquad \times|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))-\sigma_1(\hat{X}^\varepsilon(u))+\sigma_1(\bar{X}(u))|du\Big)^p\mathbf{1}_D dr\Big]\\
&\leq C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{-\tau}^{t-\tau} e^{-\lambda t}[(t-s-\tau)^{-2\alpha}+(s+\tau)^{-\alpha}]\Big(\int_{-\tau}^{s}(s-u)^{-1-\alpha}\\
&\qquad \times|\sigma_1(\hat{X}^\varepsilon(s))-\sigma_1(\bar{X}(s))-\sigma_1(\hat{X}^\varepsilon(u))+\sigma_1(\bar{X}(u))|du\Big)^p\mathbf{1}_D ds\Big]\\
&\leq C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{-\tau}^{t-\tau} e^{-\lambda t}[(t-s-\tau)^{-2\alpha}+(s+\tau)^{-\alpha}]\\
&\qquad \times\Big(\int_{-\tau}^{s}\frac{|\hat{X}^\varepsilon(s)-\bar{X}(s)-\hat{X}^\varepsilon(u)+\bar{X}(u)|}{(s-u)^{1+\alpha}}du\Big)^p\mathbf{1}_D ds\Big]\\
&\quad +C_{p,R}\mathbb{E}\Big[\sup_{t\in[0,T]}\int_{-\tau}^{t-\tau} e^{-\lambda t}[(t-s-\tau)^{-2\alpha}+(s+\tau)^{-\alpha}]\\
&\qquad \times\Big(|\hat{X}^\varepsilon(s)-\bar{X}(s)|\int_{-\tau}^{s}\frac{|\hat{X}^\varepsilon(s)-\hat{X}^\varepsilon(u)|}{(s-u)^{1+\alpha}}du\Big)^p\mathbf{1}_D ds\Big]
\end{aligned}
$$

$$
\begin{aligned}
&+C_{p,R}\mathbb{E}\bigg[\sup_{t\in[0,T]}\int_{-\tau}^{t-\tau} e^{-\lambda t}[(t-s-\tau)^{-2\alpha}+(s+\tau)^{-\alpha}] \\
&\quad\times\bigg(|\hat{X}^\varepsilon(s)-\bar{X}(s)|\int_{-\tau}^{s}\frac{|\bar{X}(s)-\bar{X}(u)|}{(s-u)^{1+\alpha}}du\bigg)^p\mathbf{1}_D ds\bigg] \\
\leq\ & C_{p,R}\mathbb{E}\bigg[\sup_{t\in[0,T]}\int_{-\tau}^{t-\tau} e^{-\lambda(t-s)}[(t-s-\tau)^{-2\alpha}+(s+\tau)^{-\alpha}] \\
&\quad\times(1+\Delta(\hat{X}^\varepsilon)+\Delta(\bar{X}))^p\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D ds\bigg],
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta(\hat{X}^\varepsilon) &= \sup_{-\tau\leq s\leq T}\int_{-\tau}^{s}\frac{|\hat{X}^\varepsilon(s)-\hat{X}^\varepsilon(q)|}{(s-q)^{1+\alpha}}dq, \\
\Delta(\bar{X}) &= \sup_{-\tau\leq s\leq T}\int_{-\tau}^{s}\frac{|\bar{X}(s)-\bar{X}(q)|}{(s-q)^{1+\alpha}}dq.
\end{aligned}
$$

Then, by (3.11) and (3.13), under the condition that $\|B^H\|_{\alpha,0,T}\leq R$, there exists a constant C_R, such

$$
\begin{aligned}
\Delta(\hat{X}^\varepsilon)+\Delta(\bar{X}) \leq\ & C\Lambda_{\alpha,B^H}(1+\|\hat{X}^\varepsilon\|_{\alpha,\infty(\tau)})\sup_{-\tau\leq s\leq T}\int_{-\tau}^{s}(s-r)^{(1-\alpha)-1-\alpha}dr \\
&+C\Lambda_{\alpha,B^H}(1+\|\bar{X}\|_{\alpha,\infty(\tau)})\sup_{-\tau\leq s\leq T}\int_{-\tau}^{s}(s-r)^{(1-\alpha)-1-\alpha}dr \\
\leq\ & C\Lambda_{\alpha,B^H}(1+\|\hat{X}^\varepsilon\|_{\alpha,\infty(\tau)}+\|\bar{X}\|_{\alpha,\infty(\tau)}) \\
\leq\ & C_R. \qquad (3.24)
\end{aligned}
$$

Thus, by (3.24), we obtain

$$
\mathbf{C}_5\leq\mathbf{C}_{51}+\mathbf{C}_{52}\leq C_{p,R}\lambda^{2\alpha-1}\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D]. \qquad (3.25)
$$

Now, let us consider $\mathbf{B}_3$. Clearly, we have

$$
\begin{aligned}
\mathbf{B}_3 =\ & \mathbb{E}\bigg[\sup_{t\in[0,T]}e^{-\lambda t}\bigg(\int_{-\tau}^{0}\frac{|\hat{X}^\varepsilon(t)-\bar{X}(t)|}{(t-s)^{\alpha+1}}ds\bigg)^p\bigg] \\
\leq\ & \mathbb{E}\bigg[\sup_{t\in[0,T]}t^{-p\alpha}e^{-\lambda t}|\hat{X}^\varepsilon(t)-\bar{X}(t)|^p\bigg] \\
\leq\ & C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}t^{-p\alpha}e^{-\lambda t}\bigg|\int_0^t(b_1(X^\varepsilon_{s_\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{s_\delta}))ds\bigg|^p\mathbf{1}_D\bigg] \\
&+C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}t^{-p\alpha}e^{-\lambda t}\bigg|\int_0^t(\bar{b}_1(X^\varepsilon_{s_\delta})-\bar{b}_1(X^\varepsilon_s))ds\bigg|^p\mathbf{1}_D\bigg] \\
&+C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}t^{-p\alpha}e^{-\lambda t}\bigg|\int_0^t(\bar{b}_1(X^\varepsilon_s)-\bar{b}_1(\hat{X}^\varepsilon_s))ds\bigg|^p\mathbf{1}_D\bigg] \\
&+C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}t^{-p\alpha}e^{-\lambda t}\bigg|\int_0^t(\bar{b}_1(\hat{X}^\varepsilon_s)-\bar{b}_1(\bar{X}_s))ds\bigg|^p\mathbf{1}_D\bigg] \\
&+C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}t^{-p\alpha}e^{-\lambda t}\bigg|\int_0^t(\sigma_1(\hat{X}^\varepsilon(s-\tau))-\sigma_1(\bar{X}(s-\tau)))dB^H_s\bigg|^p\mathbf{1}_D\bigg] \\
&+C_p\mathbb{E}\bigg[\sup_{t\in[0,T]}t^{-p\alpha}e^{-\lambda t}\bigg|\int_0^t(\sigma_1(X^\varepsilon(s-\tau))-\sigma_1(\hat{X}^\varepsilon(s-\tau)))dB^H_s\bigg|^p\mathbf{1}_D\bigg]
\end{aligned}
$$

$$=: \sum_{j=1}^{6} \mathbf{B}_{3i}.$$

First, for $\mathbf{B}_{31}$, one has

$$\begin{aligned}
\mathbf{B}_{31} &\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big|\int_0^t (b_1(X^\varepsilon_{s_\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{s_\delta}))ds\Big|^p\mathbf{1}_D\Big]\\
&\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big|\int_{t_\delta}^t (b_1(X^\varepsilon_{s_\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{s_\delta}))ds\Big|^p\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}\big(\lfloor\frac{t}{\delta}\rfloor\big)^{p-1}\sum_{k=0}^{\lfloor\frac{t}{\delta}\rfloor-1}\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{k\delta}))ds\Big|^p\Big]\\
&\leq C_p\delta+\frac{C_p}{\delta^{p-1}}\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}\sum_{k=0}^{\lfloor\frac{t}{\delta}\rfloor-1}\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{k\delta}))ds\Big|^p\Big]\\
&\leq C_p\delta+\frac{C_p}{\delta^p}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\mathbb{E}\Big[\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_s)-\bar{b}_1(X^\varepsilon_{k\delta}))ds\Big|^p\Big]\\
&\leq C_p\delta+\frac{C_p}{\delta^{p'}}\max_{0\leq k\leq\lfloor\frac{T}{\delta}\rfloor-1}\Big(\mathbb{E}\Big[\Big|\int_{k\delta}^{(k+1)\delta}(b_1(X^\varepsilon_{k\delta},\hat{Y}^\varepsilon_r)-\bar{b}_1(X^\varepsilon_{k\delta}))dr\Big|^2\Big]\Big)^{\frac{p'}{2}}\\
&\leq C_p\delta+C_p\varepsilon^{p'}\delta^{-p'}.
\end{aligned}$$

On the other hand, for $\mathbf{B}_{32},\mathbf{B}_{33}$ and $\mathbf{B}_{34}$, we obtain that

$$\begin{aligned}
\mathbf{B}_{32}+\mathbf{B}_{33}+\mathbf{B}_{34} &\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big|\int_0^t(\bar{b}_1(X^\varepsilon_{s_\delta})-\bar{b}_1(X^\varepsilon_s))ds\Big|^p\mathbf{1}_D\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big|\int_0^t(\bar{b}_1(X^\varepsilon_s)-\bar{b}_1(\hat{X}^\varepsilon_s))ds\Big|^p\mathbf{1}_D\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big|\int_0^t(\bar{b}_1(\hat{X}^\varepsilon_s)-\bar{b}_1(\bar{X}_s))ds\Big|^p\mathbf{1}_D\Big]\\
&\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}\int_0^t e^{-\lambda t}\|X^\varepsilon_{s_\delta}-X^\varepsilon_s\|^p_\infty ds\mathbf{1}_D\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}\int_0^t e^{-\lambda(t-s)}\sup_{-\tau\leq r\leq s}e^{-\lambda s}|\hat{X}^\varepsilon(r)-X^\varepsilon(r)|^p ds\mathbf{1}_D\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}\int_0^t e^{-\lambda(t-s)}\sup_{-\tau\leq r\leq s}e^{-\lambda s}|\hat{X}^\varepsilon(r)-\bar{X}(r)|^p ds\mathbf{1}_D\Big]\\
&\leq C_p\int_0^T\mathbb{E}[\|X^\varepsilon_{s_\delta}-X^\varepsilon_s\|^p_\infty\mathbf{1}_D]ds+C_p\lambda^{-1}\mathbb{E}[\|\hat{X}^\varepsilon-X^\varepsilon\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D]\\
&\quad +C_p\lambda^{-1}\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D]\\
&\leq C_p\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}})+C_p\lambda^{-1}\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D].
\end{aligned}$$

Then, for $\mathbf{B}_{35}$, we have

$$\begin{aligned}
\mathbf{B}_{35} &\leq C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big(\int_0^t\frac{|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))|}{r^\alpha}dr\Big)^p\mathbf{1}_D\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{-p\alpha}e^{-\lambda t}\Big(\int_0^t\Big(\int_0^r(r-q)^{-1-\alpha}\\
&\qquad\times|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))-\sigma_1(\hat{X}^\varepsilon(q-\tau))+\sigma_1(\bar{X}(q-\tau))|dq\Big)dr\Big)^p\mathbf{1}_D\Big]
\end{aligned}$$

$$
\begin{aligned}
&\le C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{(p-1)(1-2\alpha)}e^{-\lambda t}\int_0^t \frac{|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))|^p}{r^{2\alpha}}\mathbf{1}_D dr\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}e^{-\lambda t}\int_0^t\Big(\int_{-\tau}^{r-\tau}(r-\tau-u)^{-1-\alpha}\\
&\qquad \times|\sigma_1(\hat{X}^\varepsilon(r-\tau))-\sigma_1(\bar{X}(r-\tau))-\sigma_1(\hat{X}^\varepsilon(u))+\sigma_1(\bar{X}(u))|du\Big)^p\mathbf{1}_D dr\Big]\\
&\le C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{(p-1)(1-2\alpha)}\int_0^t e^{-\lambda(t-r)}\frac{\sup_{-\tau\le q\le r}e^{-\lambda r}|\hat{X}^\varepsilon(q)-\bar{X}(q)|^p}{r^{2\alpha}}\mathbf{1}_D dr\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}\int_{-\tau}^{t-\tau}e^{-\lambda(t-s)}\\
&\qquad \times e^{-\lambda s}\Big(\int_{-\tau}^s\frac{|\hat{X}^\varepsilon(s)-\bar{X}(s)-\hat{X}^\varepsilon(u)+\bar{X}(u)|}{(s-u)^{1+\alpha}}du\Big)^p\mathbf{1}_D ds\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}\int_{-\tau}^{t-\tau}e^{-\lambda(t-s)}\\
&\qquad \times e^{-\lambda s}\Big(|\hat{X}^\varepsilon(s)-\bar{X}(s)|\int_{-\tau}^s\frac{|\hat{X}^\varepsilon(s)-\hat{X}^\varepsilon(u)|}{(s-u)^{1+\alpha}}du\Big)^p\mathbf{1}_D ds\Big]\\
&\quad +C_p\mathbb{E}\Big[\sup_{t\in[0,T]} t^{p-1-p\alpha}\int_{-\tau}^{t-\tau}e^{-\lambda(t-s)}\\
&\qquad \times e^{-\lambda s}\Big(|\hat{X}^\varepsilon(s)-\bar{X}(s)|\int_{-\tau}^s\frac{|\bar{X}(s)-\bar{X}(u)|}{(s-u)^{1+\alpha}}du\Big)^p\mathbf{1}_D ds\Big]\\
&\le C_{p,R}\lambda^{2\alpha-1}\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D].
\end{aligned}
$$

In the same way as for the term $\mathbf{B}_{35}$, we have

$$
\mathbf{B}_{36}\le C_{p,R}\lambda^{2\alpha-1}\mathbb{E}[\|\hat{X}^\varepsilon-X^\varepsilon\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D].
$$

Now, by taking the discussions above into account, we obtain

$$
\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}\mathbf{1}_D]\le C_{p,R}\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}})+C_{p,R}(\delta+\varepsilon^{p'}\delta^{-p'}).
$$

Finally, we obtain that

$$
\begin{aligned}
\mathbb{E}[\|\hat{X}^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}] &\le C_{p,R}\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}})+C_{p,R}(\delta+\varepsilon^{p'}\delta^{-p'})\\
&\quad +C_p\sqrt{R^{-1}\mathbb{E}[\|B^H\|^2_{\alpha,0,T}]}.
\end{aligned}
$$

Then, the statement follows. □

Step 3: The estimate for $\|\bar{X}-X^\varepsilon\|_{\alpha,\lambda(\tau)}$.

By Lemma 3.11 and Lemma 3.12, we have

$$
\begin{aligned}
\mathbb{E}[\|X^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}] &\le C_{p,R}\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}})+C_{p,R}(\delta+\varepsilon^{p'}\delta^{-p'})\\
&\quad +C_p\sqrt{R^{-1}\mathbb{E}[\|B^H\|^2_{\alpha,0,T}]}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
\mathbb{E}[\|X^\varepsilon-\bar{X}\|^p_{\alpha,\infty(\tau)}] &\le e^{\lambda T}\mathbb{E}[\|X^\varepsilon-\bar{X}\|^p_{\alpha,\lambda(\tau)}]\\
&\le C_{p,R}\delta^{p(1-\alpha)-1}(1+\varepsilon^{-1}e^{\beta\frac{\delta}{\varepsilon}})+C_{p,R}(\delta+\varepsilon^{p'}\delta^{-p'})\\
&\quad +C_p\sqrt{R^{-1}\mathbb{E}[\|B^H\|^2_{\alpha,0,T}]}.
\end{aligned}
$$

Thus, if $\delta = \varepsilon\sqrt{-\ln\varepsilon}$, then, for any $p > \frac{2}{1-\alpha}$, as $R \to \infty$, one see that

$$\lim_{\varepsilon\to 0} \mathbb{E}[\|X^\varepsilon - \bar{X}\|^p_{\alpha,\infty(\tau)}] = 0.$$

The Hölder's inequality yields that above conclusion also holds for $0 < p \le \frac{2}{1-\alpha}$.
This completes the proof. □

Appendix

Note that similar to the proofs of [1, Lemma 4.2, Lemma 4.3], the proofs of Lemma 3.9 and Lemma 3.10 in this paper can be obtained. To make this paper self-contained, we present the modified proof here.

The Proof of Lemma 3.9: In what follows, we verify (3.14) by an induction argument. For any $t \in [0,\tau)$, due to $Y^\varepsilon_0 = \hat{Y}^\varepsilon_0 = \eta$, it is easy to check that

$$\mathbb{E}[\|Y^\varepsilon_t - \hat{Y}^\varepsilon_t\|^p_\infty] \le \sum_{j=0}^{\lfloor t/\delta\rfloor} \mathbb{E}\Big[\sup_{j\delta\le s\le((j+1)\delta)\wedge t} |Y^\varepsilon(s) - \hat{Y}^\varepsilon(s)|^p\Big] =: I(t,\delta).$$

By means of Itô's formula and B-D-G's inequality, together with $Y^\varepsilon(t_\delta) = \hat{Y}^\varepsilon(t_\delta)$, we obtain from (H3) that

$$\begin{aligned}
&\mathbb{E}\big[\sup_{j\delta\le s\le((j+1)\delta)\wedge t} |Y^\varepsilon(s) - \hat{Y}^\varepsilon(s)|^p\big]\\
&\le \frac{C}{\varepsilon}\int_{j\delta}^{((j+1)\delta)\wedge t} \mathbb{E}[\|X^\varepsilon_s - X^\varepsilon_{s_\delta}\|^p_\infty + |Y^\varepsilon(s) - \hat{Y}^\varepsilon(s)|^p]ds\\
&\quad + \frac{1}{2}\mathbb{E}\big[\sup_{j\delta\le s\le((j+1)\delta)\wedge t} |Y^\varepsilon(s) - \hat{Y}^\varepsilon(s)|^p\big], \qquad t \in [0,\tau).
\end{aligned}$$

Consequently, we conclude that

$$\begin{aligned}
I(t,\delta) &\le \frac{1}{\varepsilon}\int_0^t \mathbb{E}[\|X^\varepsilon_s - X^\varepsilon_{s_\delta}\|^p_\infty]ds\\
&\quad + \frac{1}{\varepsilon}\sum_{j=0}^{\lfloor t/\delta\rfloor}\int_{j\delta}^{((j+1)\delta)\wedge t} \mathbb{E}[|Y^\varepsilon(s) - \hat{Y}^\varepsilon(s)|^p]ds\\
&\le \frac{1}{\varepsilon}\int_0^t \mathbb{E}\big[\|X^\varepsilon_s - X^\varepsilon_{s_\delta}\|^p_\infty\big]ds\\
&\quad + \frac{1}{\varepsilon}\int_0^\delta \sum_{j=0}^{\lfloor t/\delta\rfloor} \mathbb{E}\big[\sup_{j\delta\le r\le(j\delta+s)\wedge t} |Y^\varepsilon(r) - \hat{Y}^\varepsilon(r)|^p\big]ds\\
&\le \frac{1}{\varepsilon}\int_0^t \mathbb{E}[\|X^\varepsilon_s - X^\varepsilon_{s_\delta}\|^p_\infty]ds + \frac{1}{\varepsilon}\int_0^\delta I(t,s)ds.
\end{aligned} \tag{3.26}$$

This, combined with Lemma 3.8 and Gronwall's inequality, gives that

$$\mathbb{E}[\|Y^\varepsilon_t - \hat{Y}^\varepsilon_t\|^p_\infty] \le C_p\varepsilon^{-1}\delta^{p(1-\alpha)-1}e^{\beta\frac{\delta}{\varepsilon}}, \qquad t \in [0,\tau), \tag{3.27}$$

for some $C_p > 0$.

Next, for any $t \in [0,\tau)$, thanks to (3.27), it is immediately to obtain

$$\mathbb{E}[\|Y^\varepsilon_t - \hat{Y}^\varepsilon_t\|^p_\infty] \quad \le \quad \mathbb{E}[\|Y^\varepsilon_\tau - \hat{Y}^\varepsilon_\tau\|^p_\infty] + \mathbb{E}\big[\sup_{\tau\le s\le t} |Y^\varepsilon(s) - \hat{Y}^\varepsilon(s)|^p\big]$$

$$
\begin{aligned}
&\le C_p\varepsilon^{-1}\delta^{p(1-\alpha)-1}e^{\beta\frac{\delta}{\varepsilon}} \\
&\quad +C_p\sum_{j=0}^{\lfloor t-\tau\rfloor}\mathbb{E}\big[\sup_{(N+j)\delta\le s\le((N+j+1)\delta)\wedge t}|Y^\varepsilon(s)-\hat{Y}^\varepsilon(s)|^p\big] \\
&=: C_p\varepsilon^{-1}\delta^{p(1-\alpha)-1}e^{\beta\frac{\delta}{\varepsilon}}+C_pM(t,\tau,\delta).
\end{aligned}
$$

Carrying out a similar argument to derive (3.26), we deduce from (3.27) that

$$
\begin{aligned}
M(t,\tau,\delta) &\le \frac{1}{\varepsilon}\int_\tau^t\mathbb{E}[\|X_s^\varepsilon-X_{s_\delta}^\varepsilon\|_\infty^p]ds \\
&\quad +\frac{1}{\varepsilon}\int_0^\delta\sum_{j=0}^{\lfloor t-\tau\rfloor}\mathbb{E}\big[\sup_{(N+j)\delta\le r\le((N+j)\delta+s)\wedge t}|Y^\varepsilon(r)-\hat{Y}^\varepsilon(r)|^p\big]ds \\
&\quad +\frac{1}{\varepsilon}\int_0^\delta\sum_{j=0}^{\lfloor t-\tau\rfloor}\mathbb{E}\big[\sup_{j\delta\le s\le((j+1)\delta)\wedge(t-\tau)}|Y^\varepsilon(s)-\hat{Y}^\varepsilon(s)|^p\big]ds \\
&\le C_p\left(\varepsilon^{-1}\delta^{p(1-\alpha)-1}+\frac{\delta}{\varepsilon}\frac{\delta^{p(1-\alpha)-1}}{\varepsilon}e^{\frac{\beta\delta}{\varepsilon}}\right)+\frac{1}{\varepsilon}\int_0^\delta M(t,\tau,s)ds.
\end{aligned}
$$

Thus, the Gronwall's inequality yields

$$
M(t,\tau,\delta)\le C_p\frac{\delta^{p(1-\alpha)-1}}{\varepsilon}e^{\frac{\beta\delta}{\varepsilon}},
$$

where we have used $\frac{\varepsilon}{\delta}\in(0,1)$. Finally, (3.27) follows by repeating the previous procedure. □

The Proof of Lemma 3.10: From (3.2), it follows that,

$$
\begin{aligned}
Y^\varepsilon(t) &= \eta(0)+\int_0^{t/\varepsilon}b_2(X_{\varepsilon s}^\varepsilon,Y^\varepsilon(\varepsilon s),Y^\varepsilon(\varepsilon s-\tau))dt \\
&\quad +\int_0^{t/\varepsilon}\sigma_2(X_{\varepsilon s}^\varepsilon,Y^\varepsilon(\varepsilon s),Y^\varepsilon(\varepsilon s-\tau))d\bar{W}_s, \qquad t>0, \qquad (3.28)
\end{aligned}
$$

where we used the fact that $\bar{W}_t:=\frac{1}{\sqrt{\varepsilon}}W_{\varepsilon t}$ is a Brownian motion. For fixed $\varepsilon>0$ and $t\ge0$, let $\bar{Y}^\varepsilon(t+\theta)=Y^\varepsilon(\varepsilon t+\theta),\theta\in[-\tau,0]$. So, one has $\bar{Y}_t^\varepsilon=Y_{\varepsilon t}^\varepsilon$. Observe that (3.28) can be rewritten as follows.

$$
\begin{aligned}
\bar{Y}^\varepsilon(t/\varepsilon) &= \eta(0)+\int_0^{t/\varepsilon}b_2(X_{\varepsilon s}^\varepsilon,\bar{Y}^\varepsilon(s),\bar{Y}^\varepsilon(s-\tau))dt \\
&\quad +\int_0^{t/\varepsilon}\sigma_2(X_{\varepsilon s}^\varepsilon,\bar{Y}^\varepsilon(s),\bar{Y}^\varepsilon(s-\tau))d\bar{W}_s.
\end{aligned}
$$

Then, following the argument which has been obtained in [1, Section 3, (3.11)], for any $s>0$ we deduce that

$$
\mathbb{E}[\|\bar{Y}_s^\varepsilon\|_\infty^2]\le1+\|\eta\|_\infty^2e^{-\rho s}+\mathbb{E}\big[\sup_{0\le r\le\varepsilon s}\|X_r^\varepsilon\|_\infty^2\big].
$$

This, together with $\bar{Y}_t^\varepsilon=Y_{\varepsilon t}^\varepsilon$, gives that

$$
\mathbb{E}[\|Y_{\varepsilon s}^\varepsilon\|_\infty^2]\le1+\|\eta\|_\infty^2e^{-\rho s}+\mathbb{E}\big[\sup_{0\le r\le\varepsilon s}\|X_r^\varepsilon\|_\infty^2\big].
$$

In particular, taking $s=t/\varepsilon$ we arrive at,

$$
\mathbb{E}[\|Y_t^\varepsilon\|_\infty^2]\le1+\|\eta\|_\infty^2+\mathbb{E}\big[\sup_{0\le r\le t}\|X_r^\varepsilon\|_\infty^2\big].
$$

This, together with Lemma 3.6, yields that,

$$\sup_{t\in[0,T]} \mathbb{E}[\|Y_t^\varepsilon\|_\infty^2] \le C$$

for some $C > 0$. Observe from Lemma 3.9 and Hölder's inequality that

$$\begin{aligned} \mathbb{E}[\|\hat{Y}_t^\varepsilon\|_\infty^2] &\le 2\mathbb{E}[\|Y_t^\varepsilon - \hat{Y}_t^\varepsilon\|_\infty^2] + 2\mathbb{E}[\|Y_t^\varepsilon\|_\infty^2] \\ &\le C + C\big(\varepsilon^{-1}\delta^{p(1-\alpha)-1}e^{\frac{\beta\delta}{\varepsilon}}\big)^{2/p}, \qquad p > 2(1-\alpha)^{-1}. \end{aligned}$$

Next, taking $\delta = \varepsilon(-\ln\varepsilon)^{\frac{1}{2}}$ in the estimate above and letting $y = (\ln\varepsilon)^{\frac{1}{2}}$, we have

$$\mathbb{E}[\|Y_t^\varepsilon\|_\infty^2] \le 1 + \big(e^{y^2}(e^{-y^2}y)^{p(1-\alpha)-1}e^{\beta y}\big)^{2/p}, \qquad p > 2(1-\alpha)^{-1}.$$

Then, the desired assertion follows since the leading term

$$e^{y^2}(e^{-y^2}y)^{p(1-\alpha)-1}e^{\beta y} \to 0$$

as $y \uparrow \infty$ whenever $p > 2(1-\alpha)^{-1}$. □

Acknowledgments

B. Pei was partially supported by National Natural Science Foundation (NSF) of China under Grants No.12172285, NSF of Chongqing under Grant No.cstc2021jcyj-msxmX0296,Guangdong Basic and Applied Basic Research Foundation under Grant No.2214050001158, Young Talent fund of University Association for Science and Technology in Shaanxi and Fundamental Research Funds for the Central Universities.

Y. Xu was partially supported by NSF of China under Grant No.12072264, Key International (Regional) Joint Research Program of NSF of China under Grant No.12120101002, Research Funds for Interdisciplinary Subject of Northwestern Polytechnical University, and Shaanxi Provincial Key R&D Program under Grants No.2019TD-010 and No.2020KW-013.

References

[1] J. Bao, Q. Song, G. Yin, and C. Yuan, Ergodicity and strong limit results for two-time-scale functional stochastic differential equations. *Stochastic Analysis and Applications,* 35(6):1030-1046, 2017.

[2] F. Biagini, Y. Hu, B. Øksendal, and T. Zhang, *Stochastic Calculus for Fractional Brownian Motion and Applications.* Springer Science & Business Media, 2008.

[3] C. Bréhier, Strong and weak orders in averaging for SPDEs. *Stochastic Processes and Their Applications*, 122(7):2553–2593, 2012.

[4] S. Cerrai and M. Freidlin, Averaging principle for a class of stochastic reaction–diffusion equations. *Probability Theory and Related Fields*, 144(1):137–177, 2009.

[5] P. Cheridito, Mixed fractional Brownian motion. *Bernoulli* 7(6), 913–934, 2001.

[6] L. Decreusefond, A. Üstünel, Fractional Brownian motion: Theory and applications. *ESAIM: Proceedings*, 5:75–86, 1998.

[7] J. Duan and W. Wang, *Effective Dynamics of Stochastic Partial Differential Equations.* Elsevier, 2014.

[8] M. Ferrante, C. Rovira, Convergence of delay differential equations driven by fractional Brownian motion. *Journal of Evolution Equations*, 10(4): 761-783, 2010.

[9] M. Freidlin, A. Wentzell, *Random Perturbations of Dynamical Systems*, Springer, New York, 2012.

[10] H. Fu, L. Wan, J. Liu, Strong convergence in averaging principle for stochastic hyperbolic–parabolic equations with two time-scales. *Stochastic Processes and Their Applications*, 125(8): 3255–3279, 2015.

[11] M. Garrido-Atienza, K. Lu, B. Schmalfuss, Random dynamical systems for stochastic partial differential equations driven by a fractional Brownian motion, *Discrete & Continuous Dynamical Systems-B,* 14(2): 473-493, 2010.

[12] D. Givon, Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems. *Multiscale Modeling & Simulation*, 6(2): 577-594, 2007.

[13] J. Guerra, D. Nualart, Stochastic differential equations driven by fractional Brownian motion and standard Brownian motion. *Stochastic Analysis and Applications,* 26(5): 1053-1075,2008.

[14] M. Hairer, X.-M. Li, Averaging dynamics driven by fractional Brownian motion. *Annals of Probability.* 48:1826–1860, 2020.

[15] J. Hu, C. Yuan, Strong convergence of neutral stochastic functional differential equations with two time-scales. *Discrete & Continuous Dynamical Systems-B*, 24(11): 5831, 2019.

[16] R. Khasminskii, On an averaging principle for Itô stochastic differential equations. *Kibernetica*, 4: (1968), 260-279.

[17] W. Liu, M. Röckner, X. Sun, Y. Xie, Averaging principle for slow-fast stochastic differential equations with time dependent locally Lipschitz coefficients. *Journal of Differential Equations*, 268(6):2910-2948, 2020.

[18] D. Liu, Strong convergence of principle of averaging for multiscale stochastic dynamical systems. *Communications in Mathematical Sciences,* 8(4): 999-1020, 2010.

[19] X. Mao, *Stochastic Differential Equations and Applications.* Elsevier, 2007.

[20] Y. Mishura, *Stochastic Calculus for Fractional Brownian Motion and Related Processes*, Springer, Berlin, 2008.

[21] S. Mohammed, *Stochastic Functional Differential Equations.* Boston, Pitman, 1984.

[22] D. Nualart, *The Malliavin Calculus and Related Topics, Second edition.* Springer-Verlag, Berlin, 2006.

[23] D. Nualart, A. Răşcanu, Differential equations driven by fractional Brownian motion. *Collectanea Mathematica,* 53(1): 55-81, 2002.

[24] B. Øksendal, *Stochastic Differential Equations.* Springer, Heidelberg, 2003.

[25] B. Pei, Y. Inahama, Y. Xu, Averaging principles for mixed fast-slow systems driven by fractional Brownian motion. *Kyoto Journal of Mathematics*, In Press.

[26] B. Pei, Y. Inahama, Y. Xu, Averaging principle for fast-slow system driven by mixed fractional Brownian rough path. *Journal of Differential Equations*, 301:202–235, 2020.

[27] B. Pei, Y. Xu, Y. Bai, Convergence of p-th mean in an averaging principle for stochastic partial differential equations driven by fractional Brownian motion. *Discrete & Continuous Dynamical Systems-Series B*, 25:1141–1158, 2020.

[28] B. Pei, Y. Xu, G. Yin, Stochastic averaging for a class of two-time-scale systems of stochastic partial differential equations. *Nonlinear Analysis*, 160: 159-176, 2017.

[29] B. Pei, Y. Xu, G. Yin, Averaging principles for SPDEs driven by fractional Brownian motions with random delays modulated by two-time-scale Markov switching processes. *Stochastics and Dynamics*, 18(3):1850023, 2018.

[30] B. Pei, Y. Xu, G. Yin, X. Zhang, Averaging principles for functional stochastic partial differential equations driven by a fractional Brownian motion modulated by two-time-scale Markovian switching processes, *Nonlinear Analysis: Hybrid Systems*, 27: 107-124, 2018.

[31] J. Schoenmakers and P. Kloeden, Robust option replication for a Black-Scholes model extended with nondeterministic trends. *Journal of Applied Mathematics and Sochastic Analysis* 12(2):113–120, 1999.

[32] G. Shevchenko, Mixed fractional stochastic differential equations with jumps. *Stochastics*, 86(2): 203-217, 2014.

[33] G. Shevchenko, Mixed stochastic delay differential equations. *Theory of Probability and Mathematical Statistics*, 89: 181-195, 2014.

[34] X. Sun, J. Zhai, Averaging principle for stochastic real Ginzburg-Landau equation driven by α-stable process. *Communications on Pure & Applied Analysis*, 19: 1291-1319, 2020.

[35] W. Thompson, R. Kuske, A. Monahan, Stochastic averaging of dynamical systems with multiple time scales forced with α-stable noise. *Multiscale Modeling & Simulation*, 13(4): 1194-1223, 2015.

[36] F. Wu and G. Yin, An averaging principle for two-time-scale stochastic functional differential equations. *Journal of Differential Equations*, 269 (1): 1037-1077, 2020.

[37] J. Xu, Y. Miao, $L^p(p > 2)$-strong convergence of an averaging principle for two-time-scales jump-diffusion stochastic differential equations. *Nonlinear Analysis: Hybrid Systems*, 18: 33-47, 2015.

[38] Y. Xu, J. Duan, W. Xu, An averaging principle for stochastic dynamical systems with Lévy noise. *Physica D: Nonlinear Phenomena*, 240(17):1395–1401, 2011.

[39] Y. Xu, B. Pei, R. Guo, Stochastic averaging for slow-fast dynamical systems with fractional Brownian motion. *Discrete & Continuous Dynamical Systems-Series B*, 20(7):2257–2267, 2015.

[40] Y. Xu, B. Pei, J. Wu, Stochastic averaging principle for differential equations with non-Lipschitz coefficients driven by fractional Brownian motion. *Stochastics & Dynamics*, 17(02): 1750013,2017.

[41] M. Zähle, Integration with respect to fractal functions and stochastic calculus. I. *Probability theory and related fields*, 111(3): 333-374,1998.