

# Averaging principle for rough slow-fast systems of level 3

Yuzuru INAHAMA

## Abstract

The averaging principle for slow-fast systems of various kind of stochastic (partial) differential equations has been extensively studied. An analogous result was shown for slow-fast systems of rough differential equations driven by random rough paths a few years ago and the study of “rough slow-fast systems” seems to be gaining momentum now. In all known results, however, the driving rough paths are of level 2. In this paper we formulate rough slow-fast systems driven by random rough paths of level 3 and prove the strong averaging principle of Khas’minskii-type.

**Keywords.** Slow-fast system, Averaging principle, Rough path theory.

**Mathematics subject classification.** 60L90, 70K65, 70K70, 60F99.

## 1 Introduction

Let  $(w_t)$  and  $(b_t)$  be two independent standard (finite-dimensional) Brownian motions (BMs). A slow-fast system of (finite-dimensional) stochastic differential equations (SDEs) of Itô-type are given by

$$\begin{cases} X_t^\varepsilon &= x_0 + \int_0^t f(X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon) db_s, \\ Y_t^\varepsilon &= y_0 + \varepsilon^{-1} \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) ds + \varepsilon^{-1/2} \int_0^t h(X_s^\varepsilon, Y_s^\varepsilon) dw_s, \end{cases}$$

where  $0 < \varepsilon \ll 1$  is a small parameter. The processes  $X^\varepsilon$  and  $Y^\varepsilon$  are called the slow component and the fast component, respectively. Suitable conditions are imposed on  $g$  and  $h$  so that the following so-called frozen SDE satisfies certain ergodicity for every  $x$ :

$$Y_t^{x,y} = y + \int_0^t g(x, Y_s^{x,y}) dt + \int_0^t h(x, Y_s^{x,y}) dw_s.$$

An associated unique invariant probability measure is denoted by  $\mu^x$ . We set  $\bar{f}(x) = \int f(x, y) \mu^x(dy)$  and set  $\bar{\sigma}(x)$  in a similar way. Consider the following averaged SDE:

$$\bar{X}_t = x_0 + \int_0^t \bar{f}(\bar{X}_s) ds + \int_0^t \bar{\sigma}(\bar{X}_s) db_s.$$

The averaging principle of this type, which was initiated by Khas'minskii [9], claims that  $X^\varepsilon$  converges to  $\bar{X}$  in an appropriate sense as  $\varepsilon \searrow 0$ . Early results are well-summarized in [2, Chapter 7]. Since then the study of averaging principle has been developed greatly. Today, so many kinds of slow-fast systems of stochastic (partial) differential equations are known to satisfy the averaging principle.

Rough differential equations (RDEs) are generalized controlled ordinary differential equations, which can be regarded as “de-randomization” of usual SDEs. The generalized controls (i.e. drivers) are called rough paths (RPs). When the RPs are random, we have systems quite similar to SDEs even when there is no (semi)martingale property. For basic information on rough path theory, see [3, 4] among others.

The study of slow-fast systems of random RDEs looks quite natural and interesting. The first averaging result was obtained by [13], which adopted fractional calculus approach to RP theory. It was then followed by [8], which used the controlled path theory to prove the averaging principle. Within just two years after that, four papers already appeared along this research direction. A large deviation principle associated with the averaging principle was shown in [15]. A convergence rate of the averaging principle was obtained in [14]. The averaging principle for slow-fast systems of semilinear rough partial differential equations was proved in [10]. In these three works, the fast component is driven by Brownian RP, while the slow component is driven by fractional Brownian RP or more general random RP. In a recent paper [11] the authors studied a slow-fast system whose fast component is driven by fractional Brownian noise and proved an almost-sure version of averaging principle.

In the theory of Gaussian RPs, a prominent example is fractional Brownian RP (i.e. a canonical lift of fractional Brownian motion) with Hurst parameter  $H \in (1/4, 1/2]$ . When  $H \in (1/4, 1/3]$ , the fractional Brownian RP is a (random) RP of level 3. In all the existing works on slow-fast systems of RDEs, however, the driving RPs are of level 2. It is therefore quite natural and important to generalize the theory of rough slow-fast systems to the case of level 3. Our main purpose in this paper is to construct rough slow-fast systems of level 3 and prove the strong averaging principle, which can be regarded as a level-3 version of the preceding work [8].

The structure of this paper is as follows. In Section 2 we provide assumptions on the coefficients and driving RP and then state our main theorem (Theorem 2.1). In Section 3 we carefully explain controlled path theory of level 3. Although this is basically known among experts, there seem to be few papers which actually elaborate on the subject. Section 4 is devoted to constructing random mixed RPs which drive our slow-fast systems. Unlike the other sections, this section is specific to the level 3 case and has no counterpart in the preceding work [8]. Hence, for those who already understand the level 2 case in [8], this section is the most important. Section 5 consists of two parts. In the first half we precisely formulate our rough slow-fast system in a deterministic way. The second half discusses probabilistic aspects of the system including the proof of our main theorem. However, since the second half is quite similar to a counterpart in the level 2 case in [8], our exposition is a little bit sketchy.

## Notation:

Before closing Introduction, we introduce the notation which will be used throughout the paper. We write  $\mathbb{N} = \{1, 2, \dots\}$  and  $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ . We set  $[\![1, k]\!] := \{1, \dots, k\}$  for  $k \in \mathbb{N}$ . Let  $T \in (0, \infty)$  be arbitrary and we work on the time interval  $[0, T]$  unless otherwise specified. For a subinterval  $[a, b] \subset [0, T]$ , we set  $\Delta_{[a, b]} = \{(s, t) \in \mathbb{R}^2 \mid a \leq s \leq t \leq b\}$ . When  $[a, b] = [0, T]$ , we simply write  $\Delta_T$  for this set.

Below,  $\mathcal{V}, \mathcal{V}_i$  ( $i \in \mathbb{N}$ ) and  $\mathcal{W}$  are Euclidean spaces. The set of bounded linear maps from  $\mathcal{V}$  to  $\mathcal{W}$  is denoted by  $L(\mathcal{V}, \mathcal{W})$ , which coincides with the set of all real  $(\dim \mathcal{W}) \times (\dim \mathcal{V})$  matrices. For  $k \geq 2$ , the set of  $k$ -bounded linear maps from  $\mathcal{V}_1 \times \dots \times \mathcal{V}_k$  to  $\mathcal{W}$  is denoted by  $L^{(k)}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathcal{W})$ . There are natural identification as follows;  $L(\mathcal{V}, \mathcal{W}) \cong \mathcal{V}^* \otimes \mathcal{W}$  and  $L^{(k)}(\mathcal{V}_1, \dots, \mathcal{V}_k; \mathcal{W}) \cong L(\mathcal{V}_1 \otimes \dots \otimes \mathcal{V}_k, \mathcal{W})$ . As usual, the truncated tensor algebra of degree  $k$  ( $k \in \mathbb{N}$ ) over  $\mathcal{V}$  is defined by  $T^k(\mathcal{V}) := \bigoplus_{i=0}^k \mathcal{V}^{\otimes i}$ , where we set  $\mathcal{V}^{\otimes 0} := \mathbb{R}$ .

- The set of all continuous path  $\varphi: [a, b] \rightarrow \mathcal{V}$  is denoted by  $\mathcal{C}([a, b], \mathcal{V})$ . With the usual sup-norm  $\|\varphi\|_{\infty, [a, b]}$  on the  $[a, b]$ -interval,  $\mathcal{C}([a, b], \mathcal{V})$  is a Banach space. The difference of  $\varphi$  is frequently denoted by  $\varphi^1$ , that is,  $\varphi_{s, t}^1 := \varphi_t - \varphi_s$  for  $(s, t) \in \Delta_{[a, b]}$ .
- Let  $0 < \gamma \leq 1$ . For a path  $\varphi: [a, b] \rightarrow \mathcal{V}$ , the  $\gamma$ -Hölder seminorm is defined by

$$\|\varphi\|_{\gamma, [a, b]} := \sup_{a \leq s < t \leq b} \frac{|\varphi_t - \varphi_s|_{\mathcal{V}}}{(t - s)^{\gamma}}.$$

If the right hand side is finite, we say  $\varphi$  is  $\gamma$ -Hölder continuous on  $[a, b]$ . The space of all  $\gamma$ -Hölder continuous paths on  $[a, b]$  is denoted by  $\mathcal{C}^{\gamma}([a, b], \mathcal{V})$ . The norm on this Banach space is  $|\varphi_a|_{\mathcal{V}} + \|\varphi\|_{\gamma, [a, b]}$ .

- Let  $\gamma > 0$ . For a continuous map  $\eta: \Delta_{[a, b]} \rightarrow \mathcal{V}$ , we set

$$\|\eta\|_{\gamma, [a, b]} := \sup_{a \leq s < t \leq b} \frac{|\eta_{s, t}|_{\mathcal{V}}}{(t - s)^{\gamma}}.$$

If this is finite, then  $\eta$  vanishes on the diagonal. The set of all such  $\eta$  with  $\|\eta\|_{\gamma, [a, b]} < \infty$  is denoted by  $\mathcal{C}_{(2)}^{\gamma}([a, b], \mathcal{V})$ , which is a Banach space with  $\|\eta\|_{\gamma, [a, b]}$ .

- When  $[a, b] = [0, T]$ , we write  $\mathcal{C}(\mathcal{V}), \mathcal{C}^{\gamma}(\mathcal{V}), \mathcal{C}_{(2)}^{\gamma}(\mathcal{V})$  for these spaces and  $\|\cdot\|_{\infty}, \|\cdot\|_{\gamma}, \|\cdot\|_{\gamma}$  for the corresponding (semi)norms for simplicity of notation. For  $z \in \mathcal{V}$ , we set  $\mathcal{C}_z(\mathcal{V}) := \{\varphi \in \mathcal{C}(\mathcal{V}) \mid \varphi_0 = z\}$ . We also set  $\mathcal{C}_z^{\gamma}(\mathcal{V})$  in a similar way.
- Let  $U$  be an open set of  $\mathcal{V}$ . For  $k \in \mathbb{N}_0$ ,  $C^k(U, \mathcal{W})$  stands for the set of  $C^k$ -functions from  $U$  to  $\mathcal{W}$ . (When  $k = 0$ , we simply write  $C(U, \mathcal{W})$  instead of  $C^0(U, \mathcal{W})$ .) The set of bounded  $C^k$ -functions  $f: U \rightarrow \mathcal{W}$  whose derivatives up to order  $k$  are all bounded is denoted by  $C_b^k(U, \mathcal{W})$ , which is a Banach space with the norm  $\|f\|_{C_b^k} := \sum_{i=0}^k \|\nabla^i f\|_{\infty}$ . (Here,  $\|\cdot\|_{\infty}$  stands for the usual sup-norm on  $U$ .)

- Let  $\gamma \in (1/2, 1]$  and  $m \in \mathbb{N}$ . If  $w$  belongs to  $\mathcal{C}_0^\gamma(\mathcal{V})$ , then we can define

$$S(w)_{s,t}^m := \int_{s \leq t_1 \leq \dots \leq t_m \leq t} dw_{t_1} \otimes \dots \otimes dw_{t_m}, \quad (s, t) \in \Delta_T$$

as an iterated Young integral. We call  $S(w)^m$  the  $m$ th signature of  $w$ . It is well-known that  $S_k(w)_{s,t} := (1, S(w)_{s,t}^1, \dots, S(w)_{s,t}^k) \in T^k(\mathcal{V})$  and Chen's relation holds, that is,

$$S_k(w)_{s,t} = S_k(w)_{s,u} \otimes S_k(w)_{u,t}, \quad s \leq u \leq t.$$

Here,  $\otimes$  stands for the multiplication in  $T^k(\mathcal{V})$ .

- Let  $\alpha \in (1/4, 1/2]$  and write  $k := \lfloor 1/\alpha \rfloor$ . We recall the definition of  $\alpha$ -Hölder rough path ( $\alpha$ -RP or RP). A continuous map  $X = (1, X^1, \dots, X^k) : \Delta_T \rightarrow T^k(\mathcal{V})$  is called  $\mathcal{V}$ -valued  $\alpha$ -RP if  $\|X^i\|_{i\alpha} < \infty$  for all  $i \in \llbracket 1, k \rrbracket$  and

$$X_{s,t} = X_{s,u} \otimes X_{u,t}, \quad s \leq u \leq t. \quad (1.1)$$

holds in  $T^k(\mathcal{V})$ . (This is called Chen's relation.) The set of all  $\mathcal{V}$ -valued  $\alpha$ -RPs is denoted by  $\Omega_\alpha(\mathcal{V})$ . With the distance  $d_\alpha(X, \hat{X}) := \sum_{i=1}^k \|X^i - \hat{X}^i\|_{i\alpha}$ ,  $\Omega_\alpha(\mathcal{V})$  is a complete metric space. The homogeneous norm of  $X$  is denoted by  $\|X\|_\alpha := \sum_{i=1}^k \|X^i\|_{i\alpha}^{1/i}$ . The dilation by  $\delta \in \mathbb{R}$  is defined by  $\delta X = (1, \delta X^1, \dots, \delta^k X^k)$ . It is clear that  $\|\delta X\|_\alpha = |\delta| \cdot \|X\|_\alpha$ . A typical example of RP is  $S_k(w)$  for  $w \in \mathcal{C}_0^\gamma(\mathcal{V})$  with  $\gamma \in (1/2, 1]$ . This is called a natural lift of  $w$ . We view  $S_k$  as a continuous map from  $\mathcal{C}_0^\gamma(\mathcal{V})$  to  $\Omega_\alpha(\mathcal{V})$  and call it the lift map.

- Let  $\alpha \in (1/4, 1/2]$  and write  $k := \lfloor 1/\alpha \rfloor$ . We define  $G\Omega_\alpha(\mathcal{V})$  to be the  $d_\alpha$ -closure of  $S_k(\mathcal{C}_0^1(\mathcal{V}))$ . It is called the  $\alpha$ -Hölder geometric RP space over  $\mathcal{V}$  and is a complete and separable metric space. It also coincides with the  $d_\alpha$ -closure of  $S_k(\mathcal{C}_0^\gamma(\mathcal{V}))$  for any  $1/2 < \gamma \leq 1$ . A geometric RP  $X \in G\Omega_\alpha(\mathcal{V})$  satisfies another important algebraic property called the Shuffle relations. To explain it, we identify  $\mathcal{V} = \mathbb{R}^d$  ( $d = \dim \mathcal{V}$ ) and the coordinate of  $X^i$ 's are denoted by  $X^{1,p}$ ,  $X^{2,pq}$  and  $X^{3,pqr}$  ( $p, q, r \in \llbracket 1, d \rrbracket$ ). Then we have

$$X_{s,t}^{1,p} X_{s,t}^{1,q} = X_{s,t}^{2,pq} + X_{s,t}^{2,qp}, \quad X_{s,t}^{1,p} X_{s,t}^{2,qr} = X_{s,t}^{3,pqr} + X_{s,t}^{3,qpr} + X_{s,t}^{3,grp} \quad (1.2)$$

for all  $p, q, r \in \llbracket 1, d \rrbracket$  and  $(s, t) \in \Delta_T$ . (When  $1/3 < \alpha \leq 1/2$ , we only have the first formula.) For basic information on  $\alpha$ -Hölder geometric RPs, the reader is referred to [4, Chapter 9].

## 2 Assumptions and main result

In this section we first introduce natural assumptions on the coefficients and driving random RP of the following slow-fast system and then state our main theorem.

Our slow-fast system of RDEs is given by

$$\begin{cases} X_t^\varepsilon &= x_0 + \int_0^t f(X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t \sigma(X_s^\varepsilon) dB_s, \\ Y_t^\varepsilon &= y_0 + \varepsilon^{-1} \int_0^t g(X_s^\varepsilon, Y_s^\varepsilon) ds + \varepsilon^{-1/2} \int_0^t h(X_s^\varepsilon, Y_s^\varepsilon) dW_s. \end{cases} \quad (2.1)$$

The time interval is  $[0, T]$ . Here,  $0 < \varepsilon \leq 1$  is a small parameter and (the first level path of)  $(X^\varepsilon, Y^\varepsilon)$  takes values in  $\mathbb{R}^m \times \mathbb{R}^n$ . The starting point  $(x_0, y_0)$  is always deterministic and arbitrary. (We will not keep track of the dependence on  $T, x_0, y_0$ .) At the first stage, (2.1) is a deterministic system of RDEs driven by an  $(d+e)$ -dimensional third-level RP which is denoted by  $(B, W)$ . A precise definition of the system (2.1) will be given in Subsection 5.1.

When we consider this slow-fast system of RDEs, the following are imposed as our standing assumptions:

- $\sigma \in C^4(\mathbb{R}^m, L(\mathbb{R}^d, \mathbb{R}^m))$  and  $h \in C^4(\mathbb{R}^m \times \mathbb{R}^n, L(\mathbb{R}^e, \mathbb{R}^n))$ ,
- $f \in C(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^m)$  and  $g \in C(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^n)$  are locally Lipschitz continuous.

These guarantee the existence of a unique local solution of (2.1). (This fact is well-known. See also Remark 3.4 below.) Since we show the strong version of the averaging principle in this work, we assume that  $\sigma$  depends only on the slow component.

We set

$$\tilde{g}(x, y) := g(x, y) + \frac{1}{2} \left\{ \sum_{i=1}^n \sum_{j=1}^e \frac{\partial h^{kj}}{\partial y_i}(x, y) h^{ij}(x, y) \right\}_{1 \leq k \leq n}, \quad (x, y) \in \mathbb{R}^m \times \mathbb{R}^n. \quad (2.2)$$

Here, we wrote  $h = \{h^{ij}\}_{1 \leq i \leq n, 1 \leq j \leq e}$ . The second term on the right hand side above is the standard Itô-Stratonovich correction term when  $x$  is viewed as a parameter.

To formulate our main theorem, we introduce more assumptions on these coefficients.

- (H1)  $\sigma$  is of  $C_b^4$ .
- (H2)  $f$  is bounded and globally Lipschitz continuous.
- (H3)  $h$  is globally Lipschitz continuous.
- (H4)  $\tilde{g}$  is globally Lipschitz continuous.
- (H5) There exist constants  $\gamma_1 > 0$  and  $C > 0$  such that, for all  $x \in \mathbb{R}^m$  and  $y \in \mathbb{R}^n$ ,

$$2\langle y, \tilde{g}(x, y) \rangle + |h(x, y)|^2 \leq -\gamma_1 |y|^2 + C(|x|^2 + 1).$$

- (H6) There exists a constant  $\gamma_2 > 0$  such that, for all  $x \in \mathbb{R}^m$  and  $y_1, y_2 \in \mathbb{R}^n$ ,

$$2\langle y_1 - y_2, \tilde{g}(x, y_1) - \tilde{g}(x, y_2) \rangle + |h(x, y_1) - h(x, y_2)|^2 \leq -\gamma_2 |y_1 - y_2|^2.$$

Let  $\frac{1}{4} < \alpha_0 \leq \frac{1}{3}$  and let  $(\Omega, \mathcal{F}, \mathbb{P}; \{\mathcal{F}_t\}_{0 \leq t \leq T})$  be a filtered probability space satisfying the usual condition. On this probability space, the following two independent random

variables  $w$  and  $B = (B^1, B^2, B^3)$  are defined. The former,  $w = (w_t)_{0 \leq t \leq T}$ , is a standard  $e$ -dimensional  $\{\mathcal{F}_t\}$ -BM. The Stratonovich RP lift of  $w$  is denoted by  $W = (W^1, W^2)$ . The latter,  $B = \{(B_{s,t}^1, B_{s,t}^2, B_{s,t}^3)\}_{(s,t) \in \Delta_T}$ , is an  $G\Omega_\alpha(\mathbb{R}^d)$ -valued random variable (i.e., random RP) for every  $\alpha \in (1/4, \alpha_0)$ . Here,  $G\Omega_\alpha(\mathbb{R}^d)$  is the space of  $\alpha$ -Hölder geometric RPs over  $\mathbb{R}^d$ . We assume that  $(B_{s,t}^1, B_{s,t}^2, B_{s,t}^3)$  is  $\mathcal{F}_t$ -measurable for every  $(s, t) \in \Delta_T$ .

We assume the following condition on the integrability of  $B$ . Below,  $\|B\|_\alpha := \|B^1\|_\alpha + \|B^2\|_{2\alpha}^{1/2} + \|B^3\|_{3\alpha}^{1/3}$  denotes the  $\alpha$ -Hölder homogeneous RP norm over the time interval  $[0, T]$ .

(A) For every  $\alpha \in (1/4, \alpha_0)$  and  $p \in [1, \infty)$ , we have  $\mathbb{E}[\|B\|_\alpha^p] < \infty$ .

Under this assumption, the mixed random RP  $(B, W)$  and the slow-fast system (2.1) of RDEs driven by it can be defined in a natural way (see Subsection 5.2 for precise definitions). We will show the averaging principle for (2.1) when it is driven by this random RP. Note that, unlike in the second-level case in the preceding work [8], the “Brownian component” of the mixed random RP is of Stratonovich-type.

Next, we introduce the frozen SDE and the averaged RDE associated with the slow-fast system (2.1) in the usual way. The frozen SDE is given as follows:

$$Y_t^{x,y} = y + \int_0^t \tilde{g}(x, Y_s^{x,y}) dt + \int_0^t h(x, Y_s^{x,y}) d^I w_t,$$

Here,  $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$  are deterministic and arbitrary and  $d^I w_t$  stands for the standard Itô integral with respect to a standard  $e$ -dimensional BM ( $w_t$ ). We are only interested in the law of  $Y^{x,y}$  and hence any realization of BM will do. Under the assumptions of Theorem 2.1 below, the Markov semigroup  $(P_t^x)_{t \geq 0}$  defined by  $P_t^x \varphi(y) = \mathbb{E}[\varphi(Y_t^{x,y})]$  for a bounded measurable function  $\varphi$  has a unique invariant probability measure, which is denoted by  $\mu^x$ . (This fact is well-known. See [12, Appendix A] among many others.)

Define the averaged drift by  $\bar{f}(x) = \int_{\mathbb{R}^n} f(x, y) \mu^x(dy)$  for  $x \in \mathbb{R}^m$ . The averaged RDE is given as follows:

$$\bar{X}_t = x_0 + \int_0^t \bar{f}(\bar{X}_s) ds + \int_0^t \sigma(\bar{X}_s) dB_s \quad (2.3)$$

Here,  $x_0 \in \mathbb{R}^m$  is the same as in (2.1). Under the assumptions of Theorem 2.1 below,  $\bar{f}$  is again bounded and globally Lipschitz. (This fact is also well-known. See [12, Lemma A.1] for example.) Therefore, this RDE has a unique global solution for every realization of  $B = (B^1, B^2, B^3)$ . (See Propositions 3.3 below for details.)

Now we are in a position to state our main result, whose proof will be provided at the end of Section 5. It claims that (the first level path of) the slow component of the slow-fast system (2.1) of RDEs converges to (the first level path of) the averaged RDE (2.3) in  $L^p$ -sense as  $\varepsilon \searrow 0$ . Here,  $\|\cdot\|_\beta$  stands for the  $\beta$ -Hölder (semi)norm of a usual path over the time interval  $[0, T]$ .

**Theorem 2.1.** *Assume (A) and (H1)–(H6). Then, for every  $p \in [1, \infty)$  and  $\beta \in (\frac{1}{4}, \alpha_0)$ , we have*

$$\lim_{\varepsilon \searrow 0} \mathbb{E}[\|X^\varepsilon - \bar{X}\|_\beta^p] = 0.$$

**Remark 2.2.** The law of  $X^\varepsilon - \bar{X}$  is uniquely determined by the law of  $B = (B^1, B^2, B^3)$  and the  $e$ -dimensional Wiener measure. In fact,  $X^\varepsilon - \bar{X}$  is obtained as a functional of  $B$  and  $w$ . So, the choice of a filtered probability space that carries  $B$  and  $w$  does not matter. (Verifying the existence of such a filtered probability space is easy.)

**Example 2.3.** A prominent example of  $B = (B^1, B^2, B^3)$  satisfying Assumption **(A)** in Theorem 2.1 is fractional Brownian RP with Hurst parameter  $H \in (\frac{1}{4}, \frac{1}{3}]$ . In this case,  $\alpha_0 = H$  in **(A)**. This is a canonical RP lift of  $d$ -dimensional fractional BM with Hurst parameter  $H$ . For more information, see [4, Chapter 15]. Concerning this example, we will provide more explanations in Remark 4.11.

**Remark 2.4.** The assumptions on the coefficients  $\sigma, h, f, g$  in Theorem 2.1 are stronger than those in the author's preceding work [8] on the level-two case. In particular,  $\tilde{g}$  is not allowed to be of super-linear growth in the present paper. This is not because of theoretical limitation, but simply because computations become quite involved in the level-three case. (The assumptions in Theorem 2.1 are basically similar to those in [12, 13].) It could be interesting to relax these assumptions.

### 3 Controlled path theory of level 3

In this section we collect basic results from the theory of controlled paths of level 3. It should be noted that they are basically known. For instance, [1] studied the theory of controlled paths of any level. For the level 3 case, the unpublished work [7] could be useful. However, they are few works which really elaborate somewhat tedious computations. Moreover, our RDEs are slightly more general than standard ones. We will therefore provide a detailed explanation below.

Throughout this section  $T > 0$  and  $\alpha \in (1/4, 1/3]$  and we let  $\mathcal{V}$  and  $\mathcal{W}$  be Euclidean spaces.

#### 3.1 Definition of controlled paths

First we recall the definition of a controlled path (CP) with respect to a geometric RP  $X = (1, X^1, X^2, X^3) \in G\Omega_\alpha(\mathcal{V})$ . Let  $[a, b] \subset [0, T]$  be a subinterval. We say that  $(Y, Y^\dagger, Y^{\dagger\dagger}, Y^\#, Y^{\#\#})$  is a  $\mathcal{W}$ -valued CP with respect to  $X$  on  $[a, b]$  if

$$(Y, Y^\dagger, Y^{\dagger\dagger}, Y^\#, Y^{\#\#}) \in \mathcal{C}^\alpha([a, b], \mathcal{W}) \times \mathcal{C}^\alpha([a, b], L(\mathcal{V}, \mathcal{W})) \\ \times \mathcal{C}^\alpha([a, b], L(\mathcal{V}, L(\mathcal{V}, \mathcal{W}))) \times \mathcal{C}_{(2)}^{3\alpha}([a, b], \mathcal{W}) \times \mathcal{C}_{(2)}^{2\alpha}([a, b], L(\mathcal{V}, \mathcal{W}))$$

and, for all  $(s, t) \in \Delta_{[a, b]}$ ,

$$Y_t - Y_s = Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\#, \quad (3.1)$$

$$Y_t^\dagger - Y_s^\dagger = Y_s^{\dagger\dagger} X_{s,t}^1 + Y_{s,t}^{\#\#}. \quad (3.2)$$

Notice the natural identifications  $L(\mathcal{V}, L(\mathcal{V}, \mathcal{W})) \cong L^2(\mathcal{V} \times \mathcal{V}; \mathcal{W}) \cong L(\mathcal{V}^{\otimes 2}, \mathcal{W})$ . The set of all such CPs with respect to  $X$  is denoted by  $\mathcal{Q}_X^\alpha([a, b], \mathcal{W})$ . ( $X$  is often referred to as a reference RP.) For simplicity,  $(Y, Y^\dagger, Y^{\dagger\dagger}, Y^\sharp, Y^{\sharp\sharp})$  will often be written as  $(Y, Y^\dagger, Y^{\dagger\dagger})$ . Obviously, (3.1) and (3.2) imply that both  $Y^\sharp$  and  $Y^{\sharp\sharp}$  must vanish on the diagonal. Note that in fact  $X^3$  is not involved in the definition of a CP.

A natural seminorm on  $\mathcal{Q}_X^\alpha([a, b], \mathcal{W})$  is defined by

$$\|(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\alpha, [a, b]} = \|Y^{\dagger\dagger}\|_{\alpha, [a, b]} + \|Y^\sharp\|_{3\alpha, [a, b]} + \|Y^{\sharp\sharp}\|_{2\alpha, [a, b]}$$

Then,  $\mathcal{Q}_X^\alpha([a, b], \mathcal{W})$  is a Banach space with the norm

$$|Y_a| + |Y_a^\dagger| + |Y_a^{\dagger\dagger}| + \|(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\alpha, [a, b]}.$$

(When  $[a, b] = [0, T]$ , we write  $\mathcal{Q}_X^\alpha(\mathcal{W})$  and  $\|\cdot\|_{\mathcal{Q}_X^\alpha}$  for simplicity.) Then, there exist positive constants  $C$  and  $C'$  depending only on  $\alpha$  and  $b - a$  such that

$$\begin{aligned} \|Y^\dagger\|_{\alpha, [a, b]} &\leq \|Y^{\dagger\dagger}\|_{\infty, [a, b]} \|X^1\|_{\alpha, [a, b]} + \|Y^{\sharp\sharp}\|_{\alpha, [a, b]} \\ &\leq \{|Y_a^{\dagger\dagger}| + (b - a)^\alpha \|Y^{\dagger\dagger}\|_{\alpha, [a, b]}\} \|X^1\|_{\alpha, [a, b]} + (b - a)^\alpha \|Y^{\sharp\sharp}\|_{2\alpha, [a, b]} \\ &\leq C(1 + \|X^1\|_\alpha) (|Y_a^{\dagger\dagger}| + \|(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\alpha, [a, b]}) \end{aligned} \quad (3.3)$$

and

$$\begin{aligned} \|Y\|_{\alpha, [a, b]} &\leq \|Y^\dagger\|_{\infty, [a, b]} \|X^1\|_{\alpha, [a, b]} + \|Y^{\dagger\dagger}\|_{\infty, [a, b]} \|X^2\|_{\alpha, [a, b]} + \|Y^\sharp\|_{\alpha, [a, b]} \\ &\leq \{|Y_a^\dagger| + (b - a)^\alpha \|Y^\dagger\|_{\alpha, [a, b]}\} \|X^1\|_{\alpha, [a, b]} \\ &\quad + \{|Y_a^{\dagger\dagger}| + (b - a)^\alpha \|Y^{\dagger\dagger}\|_{\alpha, [a, b]}\} (b - a)^\alpha \|X^2\|_{2\alpha, [a, b]} + (b - a)^{2\alpha} \|Y^\sharp\|_{3\alpha, [a, b]} \\ &\leq C'(1 + \|X^1\|_\alpha^2 + \|X^2\|_{2\alpha}^2) (|Y_a^\dagger| + |Y_a^{\dagger\dagger}| + \|(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\alpha, [a, b]}). \end{aligned} \quad (3.4)$$

**Example 3.1.** Here are a few typical examples of CPs for a given RP  $X \in G\Omega_\alpha(\mathcal{V})$ . (In the first three examples the time interval is  $[0, T]$  just for simplicity. It can be replaced by any subinterval  $[a, b]$ .)

1. For  $\xi \in \mathcal{W}$ ,  $\sigma \in L(\mathcal{V}, \mathcal{W})$  and  $\eta \in L(\mathcal{V}, L(\mathcal{V}, \mathcal{W}))$ ,

$$t \mapsto (\xi + \sigma X_{0,t}^1 + \eta X_{0,t}^2, \sigma + \eta X_{0,t}^1, \eta)$$

belongs to  $\mathcal{Q}_X^\alpha(\mathcal{W})$ . Note that  $\sharp$ - and  $\sharp\sharp$ -components of this CP are zero. Hence, the  $\mathcal{Q}_X^\alpha$ -seminorm of this CP is zero, too.

2. If  $\varphi \in \mathcal{C}^{3\alpha}(\mathcal{W})$ , then obviously  $(\varphi, 0, 0) \in \mathcal{Q}_X^\alpha(\mathcal{W})$  with  $\|\varphi\|_{3\alpha} = \|(\varphi, 0, 0)\|_{\mathcal{Q}_X^\alpha}$ . In this way, we have a natural continuous embedding  $\mathcal{C}^{3\alpha}(\mathcal{W}) \hookrightarrow \mathcal{Q}_X^\alpha(\mathcal{W})$ .
3. Suppose that  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\alpha(\mathcal{W})$  and  $g: \mathcal{W} \rightarrow \mathcal{W}'$  is a  $C^3$ -function from  $\mathcal{W}$  to another Euclidean space  $\mathcal{W}'$ . Then  $(g(Y), g(Y)^\dagger, g(Y)^{\dagger\dagger}) \in \mathcal{Q}_X^\alpha(\mathcal{W}')$  if we set

$$\bullet \quad g(Y)_t := g(Y_t).$$

- $g(Y)_t^\dagger := \nabla g(Y_t)Y_t^\dagger$ , where the right hand side is the composition of the two linear maps  $\nabla g(Y_t) \in L(\mathcal{W}, \mathcal{W}')$  and  $Y_t^\dagger \in L(\mathcal{V}, \mathcal{W})$ .
- $g(Y)_t^{\dagger\dagger} := \nabla g(Y_t)Y_t^{\dagger\dagger} + \nabla^2 g(Y_t)\langle Y_t^\dagger \bullet, Y_t^\dagger \star \rangle$ , where the first term on the right hand side is the composition of the two linear maps  $\nabla g(Y_t) \in L(\mathcal{W}, \mathcal{W}')$  and  $Y_t^{\dagger\dagger} \in L(\mathcal{V}, L(\mathcal{V}, \mathcal{W}))$ . (Note that the second term is symmetric in  $\bullet$  and  $\star$ .)

Using Taylor's theorem and (3.1), we can verify this fact as follows:

$$\begin{aligned}
g(Y_t) - g(Y_s) &= \nabla g(Y_s)\langle Y_{s,t}^1 \rangle + \frac{1}{2}\nabla^2 g(Y_s)\langle Y_{s,t}^1, Y_{s,t}^1 \rangle \\
&\quad + \frac{1}{2}\int_0^1 d\theta (1-\theta)^2 \nabla^3 g(Y_s + \theta Y_{s,t}^1)\langle Y_{s,t}^1, Y_{s,t}^1, Y_{s,t}^1 \rangle \\
&= \nabla g(Y_s)\langle Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2 \rangle \\
&\quad + \frac{1}{2}\nabla^2 g(Y_s)\langle Y_s^\dagger X_{s,t}^1, Y_s^\dagger X_{s,t}^1 \rangle + g(Y)_{s,t}^\sharp,
\end{aligned} \tag{3.5}$$

where

$$\begin{aligned}
g(Y)_{s,t}^\sharp &:= \nabla g(Y_s)Y_{s,t}^\sharp + \nabla^2 g(Y_s)\langle Y_s^\dagger X_{s,t}^1, Y_s^{\dagger\dagger} X_{s,t}^2 \rangle + \frac{1}{2}\nabla^2 g(Y_s)\langle Y_s^{\dagger\dagger} X_{s,t}^2, Y_s^{\dagger\dagger} X_{s,t}^2 \rangle \\
&\quad + \nabla^2 g(Y_s)\langle Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2, Y_{s,t}^\sharp \rangle + \frac{1}{2}\nabla^2 g(Y_s)\langle Y_{s,t}^\sharp, Y_{s,t}^\sharp \rangle \\
&\quad + \frac{1}{2}\int_0^1 d\theta (1-\theta)^2 \nabla^3 g(Y_s + \theta Y_{s,t}^1)\langle Y_{s,t}^1, Y_{s,t}^1, Y_{s,t}^1 \rangle.
\end{aligned} \tag{3.6}$$

It is easy to see from (3.6) that  $g(Y)^\sharp \in \mathcal{C}_{(2)}^{3\alpha}(\mathcal{W}')$ . Due to the shuffle relation (1.2) and the symmetry of the bilinear mapping, it holds that

$$\frac{1}{2}\nabla^2 g(Y_s)\langle Y_s^\dagger X_{s,t}^1, Y_s^\dagger X_{s,t}^1 \rangle = \nabla^2 g(Y_t)\langle Y_t^\dagger \bullet, Y_t^\dagger \star \rangle|_{(\bullet, \star)=X_{s,t}^2}.$$

Here, the right hand should be understood as  $X_{s,t}^2$  being plugged into an element of  $L(\mathcal{V}^{\otimes 2}, \mathcal{W}')$ . Thus, the composition  $g(Y)$  satisfies (3.1).

Keeping (3.1) and (3.2) in mind, we can see that

$$\begin{aligned}
g(Y)_t^\dagger - g(Y)_s^\dagger &= \nabla g(Y_t)Y_t^\dagger - \nabla g(Y_s)Y_s^\dagger \\
&= \nabla g(Y_s)Y_t^\dagger + \nabla^2 g(Y_s)\langle Y_{s,t}^1, Y_t^\dagger \rangle \\
&\quad + \int_0^1 d\theta (1-\theta) \nabla^3 g(Y_s + \theta Y_{s,t}^1)\langle Y_{s,t}^1, Y_{s,t}^1, Y_t^\dagger \rangle - \nabla g(Y_s)Y_s^\dagger \\
&= \nabla g(Y_s)\langle Y_s^{\dagger\dagger} X_{s,t}^1 \rangle + \nabla^2 g(Y_s)\langle Y_s^\dagger X_{s,t}^1, Y_s^\dagger \rangle + g(Y)_{s,t}^{\sharp\sharp},
\end{aligned} \tag{3.7}$$

where

$$g(Y)_{s,t}^{\sharp\sharp} := \nabla g(Y_s)\langle Y_{s,t}^{\sharp\sharp} \rangle$$

$$\begin{aligned}
& + \nabla^2 g(Y_s) \langle Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\#, Y_s^\dagger \rangle \\
& + \nabla^2 g(Y_s) \langle Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\#, Y_s^{\dagger\dagger} X_{s,t}^1 + Y_{s,t}^{\#\#} \rangle \\
& + \int_0^1 d\theta (1-\theta) \nabla^3 g(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_{s,t}^1, Y_t^\dagger \rangle.
\end{aligned} \tag{3.8}$$

It is easy to see from (3.8) that  $g(Y)^{\#\#} \in \mathcal{C}_{(2)}^{2\alpha}(\mathcal{W}')$ . Thus, the composition satisfies (3.2), too.

4. Concatenation of two CPs is also a CP. Let  $0 \leq a < b < c \leq T$ . For  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\alpha([a, b], \mathcal{W})$  and  $(\hat{Y}, \hat{Y}^\dagger, \hat{Y}^{\dagger\dagger}) \in \mathcal{Q}_X^\alpha([b, c], \mathcal{W})$  with  $(Y_b, Y_b^\dagger, Y_b^{\dagger\dagger}) = (\hat{Y}_b, \hat{Y}_b^\dagger, \hat{Y}_b^{\dagger\dagger})$ , their concatenation  $(Z, Z^\dagger, Z^{\dagger\dagger}) := (Y * \hat{Y}, Y^\dagger * \hat{Y}^\dagger, Y^{\dagger\dagger} * \hat{Y}^{\dagger\dagger})$  can naturally be defined and belongs to  $\mathcal{Q}_X^\alpha([a, c], \mathcal{W})$ . (Here,  $*$  stands for the usual concatenation operation for two continuous paths.)

It is clear that  $Z, Z^\dagger, Z^{\dagger\dagger}$  are  $\alpha$ -Hölder continuous on  $[a, c]$ . To prove that  $Z^\#, Z^{\#\#} \in \mathcal{C}_{(2)}^{2\alpha}(\mathcal{W})$ , it is sufficient to observe the following: For  $a \leq s \leq b \leq t \leq c$ , we have

$$\begin{aligned}
Z_{s,t}^{\#\#} &= Z_{s,t}^{\dagger,1} - Z_s^{\dagger\dagger} X_{s,t}^1 \\
&= (Z_{s,b}^{\dagger,1} - Z_s^{\dagger\dagger} X_{s,b}^1) + (Z_{b,t}^{\dagger,1} - Z_b^{\dagger\dagger} X_{b,t}^1) + Z_{s,b}^{\dagger\dagger,1} X_{b,t}^1 \\
&= Y_{s,b}^{\#\#} + \hat{Y}_{b,t}^{\#\#} + Y_{s,b}^{\dagger\dagger,1} X_{b,t}^1.
\end{aligned} \tag{3.9}$$

and

$$\begin{aligned}
Z_{s,t}^\# &= Z_{s,t}^1 - Z_s^\dagger X_{s,t}^1 - Z_s^{\dagger\dagger} X_{s,t}^2 \\
&= (Z_{s,b}^1 + Z_{b,t}^1) - \{Z_s^\dagger X_{s,b}^1 + Z_b^\dagger X_{b,t}^1 - Z_{s,b}^{\dagger,1} X_{b,t}^1\} \\
&\quad - \{Z_s^{\dagger\dagger} X_{s,b}^2 + Z_b^{\dagger\dagger} X_{b,t}^2 - Z_{s,b}^{\dagger\dagger,1} X_{b,t}^2 + Z_s^{\dagger\dagger} X_{s,b}^1 \otimes X_{b,t}^1\} \\
&= Y_{s,b}^\# + \hat{Y}_{b,t}^\# + Y_{s,b}^{\dagger\dagger,1} X_{b,t}^2 + Y_{s,b}^{\#\#} X_{b,t}^1.
\end{aligned} \tag{3.10}$$

Here, we have used (3.1)–(3.2) and Chen's relation for  $X$ . The right hand side of (3.9) and (3.10) are clearly dominated by a constant multiple of  $(t-s)^{2\alpha}$  and of  $(t-s)^{3\alpha}$ , respectively.

## 3.2 Integration of controlled paths against rough paths

Next we discuss integration of a CP  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\alpha([a, b], L(\mathcal{V}, \mathcal{W}))$  against a reference RP  $X \in G\Omega_\alpha(\mathcal{V})$ , where  $[a, b] \subset [0, T]$ . Note that  $Y^\dagger$  and  $Y^{\dagger\dagger}$  take values in

$$L(\mathcal{V}, L(\mathcal{V}, \mathcal{W})) \cong L^{(2)}(\mathcal{V} \times \mathcal{V}, \mathcal{W}) \cong L(\mathcal{V}^{\otimes 2}, \mathcal{W}),$$

and

$$L(\mathcal{V}, L(\mathcal{V}, L(\mathcal{V}, \mathcal{W}))) \cong L^{(3)}(\mathcal{V} \times \mathcal{V} \times \mathcal{V}, \mathcal{W}) \cong L(\mathcal{V}^{\otimes 3}, \mathcal{W}),$$

respectively. Note also that  $Y^\#$  and  $Y^{\#\#}$  take values in  $L(\mathcal{V}, \mathcal{W})$  and  $L(\mathcal{V}^{\otimes 2}, \mathcal{W})$ , respectively.

First, we define

$$J_{s,t} = Y_s X_{s,t}^1 + Y_s^\dagger X_{s,t}^2 + Y_s^{\dagger\dagger} X_{s,t}^3, \quad (s, t) \in \Delta_{[a,b]}.$$

From Chen's relation for  $X$  and (3.1)–(3.2), it easily follows that

$$J_{s,u} + J_{u,t} - J_{s,t} = Y_{s,u}^\# X_{u,t}^1 + Y_{s,u}^{\#\#} X_{u,t}^2 + Y_{s,u}^{\dagger\dagger,1} X_{u,t}^3, \quad a \leq s \leq u \leq t \leq b, \quad (3.11)$$

where we set  $Y_{s,u}^{\dagger\dagger,1} = Y_u^{\dagger\dagger} - Y_s^{\dagger\dagger}$ . Note that the right hand side above is dominated by a constant multiple of  $(t-s)^{4\alpha}$ .

Let  $\mathcal{P} = \{s = t_0 < t_1 < \dots < t_N = t\}$  be a partition of  $[s, t] \subset [a, b]$ . Its mesh size is denoted by  $|\mathcal{P}|$ . We define  $J_{s,t}(\mathcal{P}) = \sum_{i=1}^N J_{t_{i-1},t_i}$  and

$$\int_s^t Y_u dX_u = \lim_{|\mathcal{P}| \searrow 0} J_{s,t}(\mathcal{P}), \quad (s, t) \in \Delta_{[a,b]}. \quad (3.12)$$

The limit above is known to exist, and is called a RP integral. It will turn out in the next proposition that an (indefinite) RP integral against  $X$  is again a CP with respect to  $X$ . By the way it is defined, this RP integral clearly has additivity with respect to the interval  $[s, t]$ .

**Proposition 3.2.** *Let  $\frac{1}{4} < \alpha \leq \frac{1}{3}$  and  $[a, b] \subset [0, T]$ . Suppose that  $X \in G\Omega_\alpha(\mathcal{V})$  and  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\alpha([a, b], L(\mathcal{V}, \mathcal{W}))$ . Then, the limit in (3.12) exists for all  $(s, t) \in \Delta_{[a,b]}$ . Moreover, we have*

$$\left( \int_a^t Y_u dX_u, Y, Y^\dagger \right) \in \mathcal{Q}_X^\alpha([a, b], \mathcal{W}) \quad (3.13)$$

with the following estimate:

$$\begin{aligned} & \left| \int_s^t Y_u dX_u - (Y_s X_{s,t}^1 + Y_s^\dagger X_{s,t}^2 + Y_s^{\dagger\dagger} X_{s,t}^3) \right|_{\mathcal{W}} \\ & \leq \kappa_\alpha (t-s)^{4\alpha} (\|Y^\#\|_{3\alpha,[a,b]} \|X^1\|_{\alpha,[a,b]} + \|Y^{\#\#}\|_{2\alpha,[a,b]} \|X^2\|_{2\alpha,[a,b]} \\ & \quad + \|Y^{\dagger\dagger}\|_{\alpha,[a,b]} \|X^3\|_{3\alpha,[a,b]}), \quad (s, t) \in \Delta_{[a,b]}. \end{aligned} \quad (3.14)$$

Here, we set  $\kappa_\alpha = 2^{4\alpha} \zeta(4\alpha)$  with  $\zeta$  being the Riemann zeta function.

*Proof.* In this proof the norm of  $\mathcal{W}$  is denoted by  $|\cdot|$ . First we prove the convergence. For  $\mathcal{P}$  given as above, we can find  $i$  ( $1 \leq i \leq N-1$ ) such that  $t_{i+1} - t_{i-1} \leq 2(t-s)/(N-1)$ . Then, we see that

$$\begin{aligned} |J_{s,t}(\mathcal{P}) - J_{s,t}(\mathcal{P} \setminus \{t_i\})| &= |J_{t_{i-1},t_i} + J_{t_i,t_{i+1}} - J_{t_{i-1},t_{i+1}}| \\ &= |Y_{t_{i-1},t_i}^\# X_{t_i,t_{i+1}}^1 + Y_{t_{i-1},t_i}^{\#\#} X_{t_i,t_{i+1}}^2 + Y_{t_{i-1},t_i}^{\dagger\dagger,1} X_{t_i,t_{i+1}}^3| \\ &\leq (\|Y^\#\|_{3\alpha,[a,b]} \|X^1\|_{\alpha,[a,b]} + \|Y^{\#\#}\|_{2\alpha,[a,b]} \|X^2\|_{2\alpha,[a,b]} \\ & \quad + \|Y^{\dagger\dagger}\|_{\alpha,[a,b]} \|X^3\|_{3\alpha,[a,b]}) \left( \frac{2(t-s)}{N-1} \right)^{4\alpha}. \end{aligned}$$

Extracting points one by one from  $\mathcal{P}$  in this way until the partition becomes the trivial one  $\{s, t\}$ , we have

$$\begin{aligned} |J_{s,t}(\mathcal{P}) - J_{s,t}| &\leq \kappa_\alpha(t-s)^{4\alpha} (\|Y^\sharp\|_{3\alpha,[a,b]} \|X^1\|_{\alpha,[a,b]} \\ &\quad + \|Y^{\sharp\sharp}\|_{2\alpha,[a,b]} \|X^2\|_{2\alpha,[a,b]} + \|Y^{\dagger\dagger}\|_{\alpha,[a,b]} \|X^3\|_{3\alpha,[a,b]}). \end{aligned} \quad (3.15)$$

Note that the condition  $4\alpha > 1$  is used here. By standard argument in RP theory, (3.15) implies that  $\{J_{s,t}(\mathcal{P})\}_{\mathcal{P}}$  is Cauchy as  $|\mathcal{P}| \searrow 0$ . Therefore, the limit in (3.12) exists and (3.14) holds.

Since it obviously holds that

$$|Y_s^{\dagger\dagger} X_{s,t}^3| \leq \{|Y_a^{\dagger\dagger}| + (b-a)^\alpha \|Y^{\dagger\dagger}\|_{\alpha,[a,b]}\} \|X^3\|_{3\alpha,[a,b]} (t-s)^{3\alpha}$$

and

$$\begin{aligned} &|Y_t - Y_s - Y_s^\dagger X_{s,t}^1| \\ &\leq |Y_s^{\dagger\dagger} X_{s,t}^2| + |Y_{s,t}^\sharp| \\ &\leq \left\{ (|Y_a^{\dagger\dagger}| + (b-a)^\alpha \|Y^{\dagger\dagger}\|_{\alpha,[a,b]}) \|X^2\|_{2\alpha,[a,b]} + (b-a)^\alpha \|Y^{\sharp\sharp}\|_{2\alpha,[a,b]} \right\} (t-s)^{2\alpha}, \end{aligned}$$

the path  $(\int_a^t Y_u dX_u, Y, Y^\dagger)$  satisfies (3.1)–(3.2). Hence, (3.14) implies (3.13).  $\square$

### 3.3 Rough differential equations with bounded and globally Lipschitz drift

Now we discuss RDEs in the framework of controlled path theory. We basically follow [7, 1], but our RDE has a drift term. In this subsection,  $\mathcal{S}$  is a metric space and we assume  $\frac{1}{4} < \beta < \alpha \leq \frac{1}{3}$  and  $X \in G\Omega_\alpha(\mathcal{V}) \subset G\Omega_\beta(\mathcal{V})$ .

We set conditions on the coefficients of our RDE. Let  $\sigma: \mathcal{W} \rightarrow L(\mathcal{V}, \mathcal{W})$  be of  $C_b^4$  and let  $f: \mathcal{W} \times \mathcal{S} \rightarrow \mathcal{W}$  be a continuous map satisfying the following condition:

$$\sup_{y \in \mathcal{W}, z \in \mathcal{S}} |f(y, z)|_{\mathcal{W}} + \sup_{y, y' \in \mathcal{W}, y \neq y', z \in \mathcal{S}} \frac{|f(y, z) - f(y', z)|_{\mathcal{W}}}{|y - y'|_{\mathcal{W}}} < \infty. \quad (3.16)$$

The first and the second term above will be denoted by  $\|f\|_\infty$  and  $L_f$ , respectively.

For an  $\mathcal{S}$ -valued continuous path  $\psi: [0, T] \rightarrow \mathcal{S}$ , we consider the following RDE driven by  $X$  with the initial value  $\xi \in \mathcal{W}$ : For all  $t \in [0, T]$ ,

$$Y_t = \xi + \int_0^t f(Y_s, \psi_s) ds + \int_0^t \sigma(Y_s) dX_s, \quad Y_t^\dagger = \sigma(Y_t), \quad Y_t^{\dagger\dagger} = \nabla \sigma(Y_t) Y_t^\dagger. \quad (3.17)$$

For every  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\beta(\mathcal{W})$ , the right hand side of this system of equations also belongs to  $\mathcal{Q}_X^\beta(\mathcal{W})$ , due to Example 3.1 and Proposition 3.2. Therefore, (3.17) should be understood as an equality in  $\mathcal{Q}_X^\beta(\mathcal{W})$ . (Following the preceding works [7, 1], we slightly relax the Hölder topology of the space of CPs for quick proofs.)

One should note that if  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\beta(\mathcal{W})$  solves (3.17),  $Y_t^\dagger = \sigma(Y_t)$  and  $Y_t^{\dagger\dagger} = \nabla\sigma \cdot \sigma(Y_t)$ , in particular,  $Y^\dagger$  and  $Y^{\dagger\dagger}$  are completely determined by the first level part  $Y$ . Precisely,  $\nabla\sigma \cdot \sigma(y)$  above stands for the element of  $L(\mathcal{V} \otimes \mathcal{V}, \mathcal{W})$  defined by

$$\mathcal{V} \otimes \mathcal{V} \ni v \otimes v' \mapsto \nabla\sigma(y) \langle \sigma(y)v, v' \rangle \in \mathcal{W}.$$

Here we state our main result in this section.

**Proposition 3.3.** *Let the assumptions be as above. Then, for every  $X \in G\Omega_\alpha(\mathcal{V})$ ,  $\xi \in \mathcal{W}$  and  $\psi$ , there exists a unique global solution  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\beta(\mathcal{W})$  of RDE (3.17). Moreover, it satisfies the following estimate: there exist positive constants  $c$  and  $\nu$  independent of  $X, \xi, \psi, \sigma, f$  such that*

$$\|Y\|_\beta \leq c\{(K+1)(\|X\|_\alpha + 1)\}^\nu, \quad X \in G\Omega_\alpha(\mathcal{V}).$$

Here, we set  $K := \|\sigma\|_{C_b^4} \vee \|f\|_\infty \vee L_f$ .

*Proof.* In this proof the norm of finite-dimensional spaces is denoted by  $|\cdot|$ . Without loss of generality we may assume  $T = 1$ . Let  $\tau \in (0, 1]$  and  $\xi \in \mathcal{W}$ . For simplicity, we write  $\eta(y) := \nabla\sigma \cdot \sigma(y)$ . We denote by  $c_i > 0$  and  $\nu_i \in \mathbb{N}$  ( $i = 1, 2, \dots$ ) certain constants independent of  $X, \xi, \tilde{\xi}, \psi, \sigma, f, \tau$  and  $(s, t) \in \Delta_\tau$ .

We define  $\mathcal{M}_{[0, \tau]}^\xi, \mathcal{M}_{[0, \tau]}^1, \mathcal{M}_{[0, \tau]}^2: \mathcal{Q}_X^\beta([0, \tau], \mathcal{W}) \rightarrow \mathcal{Q}_X^\beta([0, \tau], \mathcal{W})$  by

$$\begin{aligned} \mathcal{M}_{[0, \tau]}^1(Y, Y^\dagger, Y^{\dagger\dagger}) &= \left( \int_0^\cdot \sigma(Y_s) dX_s, \sigma(Y), \nabla\sigma(Y) Y^\dagger \right), \\ \mathcal{M}_{[0, \tau]}^2(Y, Y^\dagger, Y^{\dagger\dagger}) &= \left( \int_0^\cdot f(Y_s, \psi_s) ds, 0, 0 \right), \\ \mathcal{M}_{[0, \tau]}^\xi(Y, Y^\dagger, Y^{\dagger\dagger}) &= (\xi, 0, 0) + \mathcal{M}_{[0, \tau]}^1(Y, Y^\dagger, Y^{\dagger\dagger}) + \mathcal{M}_{[0, \tau]}^2(Y, Y^\dagger, Y^{\dagger\dagger}). \end{aligned} \quad (3.18)$$

If  $(Y, Y^\dagger, Y^{\dagger\dagger})$  starts at  $(\xi, \sigma(\xi), \eta(\xi))$ , so does  $\mathcal{M}_{[0, \tau]}^\xi(Y, Y^\dagger, Y^{\dagger\dagger})$ . A fixed point of  $\mathcal{M}_{[0, \tau]}^\xi$  is a solution of RDE (3.17) on the interval  $[0, \tau]$ .

We also set

$$\begin{aligned} B_{[0, \tau]}^\xi &= \{(Y, Y^\dagger, Y^{\dagger\dagger}) \in \mathcal{Q}_X^\beta([0, \tau], \mathcal{W}) \mid \|(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta, [0, \tau]} \leq 1, \\ &\quad Y_0 = \xi, Y_0^\dagger = \sigma(\xi), Y_0^{\dagger\dagger} = \eta(\xi)\}. \end{aligned}$$

This set is something like a ball of radius 1 centered at

$$t \mapsto (\xi + \sigma(\xi)X_{0,t}^1 + \eta(\xi)X_{0,t}^2, \sigma(\xi) + \eta(\xi)X_{0,t}^1, \eta(\xi))$$

(see Example 3.1). Since the initial point  $(Y_0, Y_0^\dagger, Y_0^{\dagger\dagger})$  is fixed,  $\|\cdot\|_{\mathcal{Q}_X^\beta, [0, \tau]}$  defines a distance on this set.

For a while from now, we will work only on  $[0, \tau]$  and therefore omit  $[0, \tau]$  from the subscript for notational simplicity. We will often write  $\varphi_{s,t}^1 := \varphi_t - \varphi_s$  for a usual path  $\varphi$  that takes values in a vector space.

Let  $\xi, \tilde{\xi} \in \mathcal{W}$  and pick  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in B^\xi$  and  $(\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger}) \in B^{\tilde{\xi}}$  arbitrarily. For simplicity we write  $\Delta := \| (Y, Y^\dagger, Y^{\dagger\dagger}) - (\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger}) \|_{\mathcal{Q}_X^\beta}$ . One can show the following estimates for all  $(s, t) \in \Delta_\tau$  by repeatedly using (3.1) and (3.2):

$$\|Y^{\dagger\dagger}\|_\infty \leq |\eta(\xi)| + \sup_{s \leq \tau} |Y_s^{\dagger\dagger} - Y_0^{\dagger\dagger}| \leq K^2 + \|Y^{\dagger\dagger}\|_\beta \tau^\beta \leq K^2 + 1, \quad (3.19)$$

$$\|Y^{\dagger\dagger} - \tilde{Y}^{\dagger\dagger}\|_\infty \leq |\eta(\xi) - \eta(\tilde{\xi})| + \|Y^{\dagger\dagger} - \tilde{Y}^{\dagger\dagger}\|_\beta \tau^\beta \leq 2K^2 |\xi - \tilde{\xi}| + \Delta, \quad (3.20)$$

$$\begin{aligned} |Y_{s,t}^{\dagger,1}| &\leq |Y_s^{\dagger\dagger} X_{s,t}^1| + |Y_{s,t}^{\sharp\sharp}| \\ &\leq (K^2 + 1) \|X^1\|_\alpha (t-s)^\alpha + \|Y^{\sharp\sharp}\|_{2\beta} (t-s)^{2\beta} \\ &\leq (K^2 + 1) (\|X^1\|_\alpha + 1) (t-s)^\alpha, \end{aligned} \quad (3.21)$$

$$\begin{aligned} |Y_{s,t}^{\dagger,1} - \tilde{Y}_{s,t}^{\dagger,1}| &\leq |(Y_s^{\dagger\dagger} - \tilde{Y}_s^{\dagger\dagger}) X_{s,t}^1| + |Y_{s,t}^{\sharp\sharp} - \tilde{Y}_{s,t}^{\sharp\sharp}| \\ &\leq (2K^2 |\xi - \tilde{\xi}| + \Delta) \|X^1\|_\alpha (t-s)^\alpha + \|Y^{\sharp\sharp} - \tilde{Y}^{\sharp\sharp}\|_{2\beta} (t-s)^{2\beta} \\ &\leq \{2K^2 \|X^1\|_\alpha |\xi - \tilde{\xi}| + (\|X^1\|_\alpha + 1) \Delta\} (t-s)^\alpha, \end{aligned} \quad (3.22)$$

$$\begin{aligned} \|Y^\dagger\|_\infty &\leq |\sigma(\xi)| + \sup_{s \leq \tau} |Y_{0,s}^{\dagger,1}| \\ &\leq K + (K^2 + 1) (\|X^1\|_\alpha + 1) \leq 2(K^2 + 1) (\|X^1\|_\alpha + 1), \end{aligned} \quad (3.23)$$

$$\begin{aligned} \|Y^\dagger - \tilde{Y}^\dagger\|_\infty &\leq |\sigma(\xi) - \sigma(\tilde{\xi})| + \sup_{s \leq \tau} |Y_{0,s}^{\dagger,1} - \tilde{Y}_{0,s}^{\dagger,1}| \\ &\leq (K + 2K^2 \|X^1\|_\alpha) |\xi - \tilde{\xi}| + (\|X^1\|_\alpha + 1) \Delta. \end{aligned} \quad (3.24)$$

From these, we also have

$$\begin{aligned} |Y_{s,t}^1| &\leq |Y_s^\dagger X_{s,t}^1| + |Y_s^{\dagger\dagger} X_{s,t}^2| + |Y_{s,t}^\sharp| \\ &\leq 2(K^2 + 1) (\|X^1\|_\alpha + 1) \|X^1\|_\alpha (t-s)^\alpha \\ &\quad + (K^2 + 1) \|X^2\|_{2\alpha} (t-s)^{2\alpha} + \|Y^\sharp\|_{3\beta} (t-s)^{3\beta} \\ &\leq 4(K^2 + 1) (\|X\|_\alpha^2 + 1) (t-s)^\alpha, \end{aligned} \quad (3.25)$$

$$\begin{aligned} |Y_{s,t}^1 - \tilde{Y}_{s,t}^1| &\leq |(Y_s^\dagger - \tilde{Y}_s^\dagger) X_{s,t}^1| + |(Y_s^{\dagger\dagger} - \tilde{Y}_s^{\dagger\dagger}) X_{s,t}^2| + |Y_{s,t}^\sharp - \tilde{Y}_{s,t}^\sharp| \\ &\leq \|Y^\dagger - \tilde{Y}^\dagger\|_\infty \|X^1\|_\alpha (t-s)^\alpha \\ &\quad + \|Y^{\dagger\dagger} - \tilde{Y}^{\dagger\dagger}\|_\infty \|X^2\|_{2\alpha} (t-s)^{2\alpha} + \|Y^\sharp - \tilde{Y}^\sharp\|_{3\beta} (t-s)^{3\beta} \\ &\leq \{(K + 2K^2 \|X^1\|_\alpha) |\xi - \tilde{\xi}| \|X^1\|_\alpha + (1 + \|X^1\|_\alpha) \Delta \|X^1\|_\alpha \\ &\quad + 2K^2 |\xi - \tilde{\xi}| \|X^2\|_{2\alpha} + \Delta \|X^2\|_{2\alpha} + \Delta\} (t-s)^\alpha \\ &\leq 2\{(K^2 + 1) (\|X\|_\alpha^2 + 1) |\xi - \tilde{\xi}| + (\|X\|_\alpha^2 + 1) \Delta\} (t-s)^\alpha. \end{aligned} \quad (3.26)$$

Note that  $Y^\dagger$  and  $Y$  are in fact  $\alpha$ -Hölder continuous. Hence, if  $\tau$  is small, the  $\beta$ -Hölder seminorms of  $Y^\dagger$ ,  $Y^\dagger - \tilde{Y}^\dagger$ ,  $Y$  and  $Y - \tilde{Y}$  can be made very small. From (3.26) we can easily see that

$$\|Y - \tilde{Y}\|_\infty \leq |Y_0 - \tilde{Y}_0| + \sup_{s \leq \tau} |Y_{0,s}^1 - \tilde{Y}_{0,s}^1|$$

$$\leq \{1 + 2(K^2 + 1)(\|X\|_\alpha^2 + 1)\tau^\alpha\}|\xi - \tilde{\xi}| + 2(\|X\|_\alpha^2 + 1)\Delta\tau^\alpha. \quad (3.27)$$

Next we calculate  $(\sigma(Y), \sigma(Y)^\dagger, \sigma(Y)^{\dagger\dagger})$ , whose explicit form was provided in the third item of Example 3.1. We can easily see from (3.19), (3.21), (3.23) and (3.25) that

$$\begin{aligned} & |\sigma(Y)_t^{\dagger\dagger} - \sigma(Y)_s^{\dagger\dagger}| \\ & \leq |\nabla\sigma(Y_t)Y_t^{\dagger\dagger} - \nabla\sigma(Y_s)Y_s^{\dagger\dagger}| + |\nabla^2\sigma(Y_t)\langle Y_t^\dagger \bullet, Y_t^\dagger \star \rangle - \nabla^2\sigma(Y_s)\langle Y_s^\dagger \bullet, Y_s^\dagger \star \rangle| \\ & \leq \|\nabla^2\sigma\|_\infty |Y_{s,t}^1| \|Y^{\dagger\dagger}\|_\infty + \|\nabla\sigma\|_\infty |Y_{s,t}^{\dagger\dagger,1}| \\ & \quad + \|\nabla^3\sigma\|_\infty |Y_{s,t}^1| \|Y^\dagger\|_\infty^2 + 2\|\nabla^2\sigma\|_\infty |Y_{s,t}^{\dagger,1}| \|Y^\dagger\|_\infty \\ & \leq c_1(K+1)^7(\|X\|_\alpha + 1)^4(t-s)^\beta. \end{aligned} \quad (3.28)$$

From the explicit form of  $\sigma(Y)_{s,t}^{\sharp\sharp}$  in (3.8), we have

$$\begin{aligned} |\sigma(Y)_{s,t}^{\sharp\sharp}| & \leq |\nabla\sigma(Y_s)\langle Y_{s,t}^{\sharp\sharp} \rangle| \\ & \quad + |\nabla^2\sigma(Y_s)\langle Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\sharp, Y_s^\dagger \rangle| \\ & \quad + |\nabla^2\sigma(Y_s)\langle Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\sharp, Y_s^{\dagger\dagger} X_{s,t}^1 + Y_{s,t}^{\sharp\sharp} \rangle| \\ & \quad + \left| \int_0^1 d\theta (1-\theta) \nabla^3\sigma(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_{s,t}^1, Y_t^\dagger \rangle \right| \\ & \leq c_2(K+1)^7(\|X\|_\alpha + 1)^5(t-s)^{2\beta}, \end{aligned} \quad (3.29)$$

where we have used (3.19), (3.21), (3.23) and (3.25) again. Similarly, we see from (3.6) that

$$\begin{aligned} |\sigma(Y)_{s,t}^\sharp| & \leq |\nabla\sigma(Y_s)Y_{s,t}^\sharp| + |\nabla^2\sigma(Y_s)\langle Y_s^\dagger X_{s,t}^1, Y_s^{\dagger\dagger} X_{s,t}^2 \rangle| \\ & \quad + \frac{1}{2} |\nabla^2\sigma(Y_s)\langle Y_s^{\dagger\dagger} X_{s,t}^2, Y_s^{\dagger\dagger} X_{s,t}^2 \rangle| \\ & \quad + |\nabla^2\sigma(Y_s)\langle Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2, Y_{s,t}^\sharp \rangle| + \frac{1}{2} |\nabla^2\sigma(Y_s)\langle Y_{s,t}^\sharp, Y_{s,t}^\sharp \rangle| \\ & \quad + \frac{1}{2} \left| \int_0^1 d\theta (1-\theta)^2 \nabla^3\sigma(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_{s,t}^1, Y_{s,t}^1 \rangle \right| \\ & \leq c_3(K+1)^7(\|X\|_\alpha + 1)^6(t-s)^{3\beta}, \end{aligned} \quad (3.30)$$

where we have used (3.19), (3.21), (3.23) and (3.25) again. It immediately follows from (3.28)–(3.30) that

$$\|(\sigma(Y), \sigma(Y)^\dagger, \sigma(Y)^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} \leq c_4(K+1)^7(\|X\|_\alpha + 1)^6. \quad (3.31)$$

We will check that  $\mathcal{M}^\xi$  maps  $B^\xi$  to itself if  $\tau$  is small enough. As before, we simply write  $\nabla\sigma(y)\langle y' \rangle$  for  $\nabla\sigma(y)\langle y', \bullet \rangle = \nabla_{y'}\sigma(y) \in L(\mathcal{V}, \mathcal{W})$  when  $y, y' \in \mathcal{W}$ . It immediately follows from Example 3.1 that

$$\|\mathcal{M}^2(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} \leq \left\| \int_0^\cdot f(Y_s, \psi_s) ds \right\|_{3\beta} \leq K\tau^{1-3\beta} \leq K\tau^{\alpha-\beta}. \quad (3.32)$$

We estimate the norm of  $\mathcal{M}^1$ . The  $\dagger\dagger$ -component of  $\mathcal{M}^1$  satisfies

$$\begin{aligned} |\nabla\sigma(Y_t)Y_t^\dagger - \nabla\sigma(Y_s)Y_s^\dagger| &\leq |\nabla\sigma(Y_t)||Y_t^\dagger - Y_s^\dagger| + |\nabla\sigma(Y_t) - \nabla\sigma(Y_s)||Y_s^\dagger| \\ &\leq K\{|Y_{s,t}^{\dagger,1}| + |Y_{s,t}^1|\|Y^\dagger\|_\infty\} \\ &\leq 9(K+1)^5(\|X\|_\alpha + 1)^3(t-s)^\alpha, \end{aligned} \quad (3.33)$$

where we have used (3.21), (3.23) and (3.25). The  $\sharp\sharp$ -component of  $\mathcal{M}^1$  reads:

$$\begin{aligned} &\sigma(Y_t) - \sigma(Y_s) - \nabla\sigma(Y_s)\langle Y_s^\dagger X_{s,t}^1 \rangle \\ &= \{\sigma(Y_t) - \sigma(Y_s) - \nabla\sigma(Y_s)\langle Y_{s,t}^1 \rangle\} + \nabla\sigma(Y_s)\langle Y_{s,t}^1 - Y_s^\dagger X_{s,t}^1 \rangle \\ &= \int_0^1 d\theta(1-\theta)\nabla^2\sigma(Y_s + \theta Y_{s,t}^1)\langle Y_{s,t}^1, Y_{s,t}^1 \rangle + \nabla\sigma(Y_s)\langle Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\sharp \rangle. \end{aligned} \quad (3.34)$$

This implies that

$$|\sigma(Y_t) - \sigma(Y_s) - \nabla\sigma(Y_s)\langle Y_s^\dagger X_{s,t}^1 \rangle| \leq 9(K+1)^5(\|X\|_\alpha + 1)^4(t-s)^{2\alpha}, \quad (3.35)$$

where we have used (3.25) and (3.21).

By (3.14) in Proposition 3.2, we can estimate the  $\sharp$ -component of  $\mathcal{M}^1$  as follows:

$$\begin{aligned} &\left| \int_s^t \sigma(Y_u) dX_u - \sigma(Y_s)X_{s,t}^1 - \sigma(Y)_s^\dagger X_{s,t}^2 \right| \\ &\leq |\sigma(Y)_s^{\dagger\dagger} X_{s,t}^3| + \left| \int_s^t \sigma(Y_u) dX_u - \{\sigma(Y_s)X_{s,t}^1 + \sigma(Y)_s^\dagger X_{s,t}^2 + \sigma(Y)_s^{\dagger\dagger} X_{s,t}^3\} \right| \\ &\leq \{|\nabla\sigma(Y_s)\langle Y_s^{\dagger\dagger} \langle \bullet, \star \rangle, \star \rangle| + |\nabla^2\sigma(Y_s)\langle Y_s^\dagger \bullet, Y_s^\dagger \star, \star \rangle|\}|X_{s,t}^3| \\ &\quad + \kappa_\beta(t-s)^{4\beta}\{\|\sigma(Y)^\sharp\|_{3\beta}\|X^1\|_\beta + \|\sigma(Y)^{\sharp\sharp}\|_{2\beta}\|X^2\|_{2\beta} + \|\sigma(Y)^{\dagger\dagger}\|_\beta\|X^3\|_{3\beta}\} \\ &\leq c_5(K+1)^7(\|X\|_\alpha + 1)^7(t-s)^{3\alpha}, \end{aligned} \quad (3.36)$$

where we have also used (3.19), (3.23), (3.29) and (3.30).

Combining (3.33)–(3.35), we obtain

$$\begin{aligned} \|\mathcal{M}^1(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} &\leq (c_5 \vee 9) \left\{ (K+1)^4(\|X\|_\alpha + 1)^2\tau^{\alpha-\beta} \right. \\ &\quad \left. + (K+1)^5(\|X\|_\alpha + 1)^4\tau^{2(\alpha-\beta)} + (K+1)^7(\|X\|_\alpha + 1)^7\tau^{3(\alpha-\beta)} \right\}. \end{aligned}$$

From this and (3.32) we can estimate  $\mathcal{M}^\xi(Y, Y^\dagger, Y^{\dagger\dagger})$ . If

$$\tau^{\alpha-\beta} \leq [4(c_5 \vee 9)(K+1)^4(\|X\|_\alpha + 1)^3]^{-1},$$

then we have

$$\|\mathcal{M}^\xi(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} \leq \|\mathcal{M}^1(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} + \|\mathcal{M}^2(Y, Y^\dagger, Y^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} \leq 1,$$

which is equivalent to  $\mathcal{M}^\xi(B^\xi) \subset B^\xi$ .

Now we will prove that  $\mathcal{M}^\xi$  is a contraction on  $B^\xi$  if we take  $\tau$  smaller. We can easily see from (3.27) that

$$\begin{aligned}
& \|\mathcal{M}^2(Y, Y^\dagger, Y^{\dagger\dagger}) - \mathcal{M}^2(\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger})\|_{Q_X^\beta} \\
& \leq \left\| \int_0^\cdot \{f(Y_s, \psi_s) - f(\tilde{Y}_s, \psi_s)\} ds \right\|_{3\beta} \\
& \leq K \|Y - \tilde{Y}\|_\infty \tau^{1-3\beta} \\
& \leq K [1 + 2(K^2 + 1)(\|X\|_\alpha^2 + 1)\tau^\alpha] |\xi - \tilde{\xi}| + 2(\|X\|_\alpha^2 + 1)\Delta\tau^\alpha] \tau^{\alpha-\beta} \\
& \leq 3(K^2 + 1)(\|X\|_\alpha^2 + 1) |\xi - \tilde{\xi}| + 2K(\|X\|_\alpha^2 + 1)\tau^{2\alpha-\beta}\Delta. \tag{3.37}
\end{aligned}$$

We will estimate the difference of  $\mathcal{M}^1$ . To do so, we will first show that there are constants  $c_6 > 0$  and  $\nu_1 \in \mathbb{N}$  such that

$$\begin{aligned}
& \|(\sigma(Y), \sigma(Y)^\dagger, \sigma(Y)^{\dagger\dagger}) - (\sigma(\tilde{Y}), \sigma(\tilde{Y})^\dagger, \sigma(\tilde{Y})^{\dagger\dagger})\|_{Q_X^\beta} \\
& \leq c_6(K + 1)^{\nu_1} (\|X\|_\alpha + 1)^{\nu_1} (|\xi - \tilde{\xi}| + \Delta) \tag{3.38}
\end{aligned}$$

holds for all  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in B^\xi$  and  $(\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger}) \in B^{\tilde{\xi}}$ .

We can estimate the difference of the  $\dagger\dagger$ -components as follows:

$$\begin{aligned}
& |\{\sigma(Y)_t^{\dagger\dagger} - \sigma(Y)_s^{\dagger\dagger}\} - \{\sigma(\tilde{Y})_t^{\dagger\dagger} - \sigma(\tilde{Y})_s^{\dagger\dagger}\}| \\
& \leq |\{\nabla\sigma(Y_t)Y_t^{\dagger\dagger} - \nabla\sigma(Y_s)Y_s^{\dagger\dagger}\} - \{\nabla\sigma(\tilde{Y}_t)\tilde{Y}_t^{\dagger\dagger} - \nabla\sigma(\tilde{Y}_s)\tilde{Y}_s^{\dagger\dagger}\}| \\
& \quad + |\{\nabla^2\sigma(Y_t)\langle Y_t^\dagger \bullet, Y_t^\dagger \star \rangle - \nabla^2\sigma(Y_s)\langle Y_s^\dagger \bullet, Y_s^\dagger \star \rangle\} \\
& \quad \quad - \{\nabla^2\sigma(\tilde{Y}_t)\langle \tilde{Y}_t^\dagger \bullet, \tilde{Y}_t^\dagger \star \rangle - \nabla^2\sigma(\tilde{Y}_s)\langle \tilde{Y}_s^\dagger \bullet, \tilde{Y}_s^\dagger \star \rangle\}| \\
& \leq \left| \int_0^1 d\theta \{\nabla^2\sigma(Y_s + \theta Y_{s,t}^1)\langle Y_{s,t}^1, Y_t^{\dagger\dagger} \rangle - \nabla^2\sigma(\tilde{Y}_s + \theta \tilde{Y}_{s,t}^1)\langle \tilde{Y}_{s,t}^1, \tilde{Y}_t^{\dagger\dagger} \rangle\} \right| \\
& \quad + |\nabla\sigma(Y_s)\langle Y_{s,t}^{\dagger\dagger,1} \rangle - \nabla\sigma(\tilde{Y}_s)\langle \tilde{Y}_{s,t}^{\dagger\dagger,1} \rangle| \\
& \quad + \left| \int_0^1 d\theta \{\nabla^3\sigma(Y_s + \theta Y_{s,t}^1)\langle Y_{s,t}^1, Y_t^\dagger \bullet, Y_t^\dagger \star \rangle - \nabla^3\sigma(\tilde{Y}_s + \theta \tilde{Y}_{s,t}^1)\langle \tilde{Y}_{s,t}^1, \tilde{Y}_t^\dagger \bullet, \tilde{Y}_t^\dagger \star \rangle\} \right| \\
& \quad + |\nabla^2\sigma(Y_s)\langle Y_{s,t}^{\dagger,1} \bullet, Y_t^\dagger \star \rangle - \nabla^2\sigma(\tilde{Y}_s)\langle \tilde{Y}_{s,t}^{\dagger,1} \bullet, \tilde{Y}_t^\dagger \star \rangle| \\
& \quad + |\nabla^2\sigma(Y_s)\langle Y_s^\dagger \bullet, Y_{s,t}^{\dagger,1} \star \rangle - \nabla^2\sigma(\tilde{Y}_s)\langle \tilde{Y}_s^\dagger \bullet, \tilde{Y}_{s,t}^{\dagger,1} \star \rangle| \\
& \leq c_7(K + 1)^{\nu_2} (\|X\|_\alpha + 1)^{\nu_2} (|\xi - \tilde{\xi}| + \Delta) (t - s)^\beta, \tag{3.39}
\end{aligned}$$

where we have used (3.19)–(3.27) many times.

Using (3.8), we can estimate the difference of the  $\#\#$ -components in a similar way as follows:

$$\begin{aligned}
& |\sigma(Y)_{s,t}^{\#\#} - \sigma(\tilde{Y})_{s,t}^{\#\#}| \\
& \leq |\nabla\sigma(Y_s)\langle Y_{s,t}^{\#\#} \rangle - \nabla\sigma(\tilde{Y}_s)\langle \tilde{Y}_{s,t}^{\#\#} \rangle| \\
& \quad + |\nabla^2\sigma(Y_s)\langle Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\# \bullet, Y_s^\dagger \rangle - \nabla^2\sigma(\tilde{Y}_s)\langle \tilde{Y}_s^{\dagger\dagger} X_{s,t}^2 + \tilde{Y}_{s,t}^\# \bullet, \tilde{Y}_s^\dagger \rangle|
\end{aligned}$$

$$\begin{aligned}
& + |\nabla^2 \sigma(Y_s) \langle Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\sharp, Y_s^{\dagger\dagger} X_{s,t}^1 + Y_{s,t}^{\sharp\sharp} \rangle \\
& \quad - \nabla^2 \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^\dagger X_{s,t}^1 + \tilde{Y}_s^{\dagger\dagger} X_{s,t}^2 + \tilde{Y}_{s,t}^\sharp, \tilde{Y}_s^{\dagger\dagger} X_{s,t}^1 + \tilde{Y}_{s,t}^{\sharp\sharp} \rangle| \\
& + \left| \int_0^1 d\theta (1-\theta) \{ \nabla^3 \sigma(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_{s,t}^1, Y_t^\dagger \rangle - \nabla^3 \sigma(\tilde{Y}_s + \theta \tilde{Y}_{s,t}^1) \langle \tilde{Y}_{s,t}^1, \tilde{Y}_{s,t}^1, \tilde{Y}_t^\dagger \rangle \} \right| \\
& \leq c_8 (K+1)^{\nu_3} (\|X\|_\alpha + 1)^{\nu_3} (|\xi - \tilde{\xi}| + \Delta) (t-s)^{2\beta}, \tag{3.40}
\end{aligned}$$

where we have used (3.19)–(3.27) again. Note that  $\nabla^5 \sigma$  is not involved in the above estimation.

Similarly, we see from (3.6) that

$$\begin{aligned}
& |\sigma(Y)_{s,t}^\sharp - \sigma(\tilde{Y})_{s,t}^\sharp| \\
& \leq |\nabla \sigma(Y_s) Y_{s,t}^\sharp - \nabla \sigma(\tilde{Y}_s) \tilde{Y}_{s,t}^\sharp| \\
& \quad + |\nabla^2 \sigma(Y_s) \langle Y_s^\dagger X_{s,t}^1, Y_s^{\dagger\dagger} X_{s,t}^2 \rangle - \nabla^2 \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^\dagger X_{s,t}^1, \tilde{Y}_s^{\dagger\dagger} X_{s,t}^2 \rangle| \\
& \quad + \frac{1}{2} |\nabla^2 \sigma(Y_s) \langle Y_s^{\dagger\dagger} X_{s,t}^2, Y_s^{\dagger\dagger} X_{s,t}^2 \rangle - \nabla^2 \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^{\dagger\dagger} X_{s,t}^2, \tilde{Y}_s^{\dagger\dagger} X_{s,t}^2 \rangle| \\
& \quad + |\nabla^2 \sigma(Y_s) \langle Y_s^\dagger X_{s,t}^1 + Y_s^{\dagger\dagger} X_{s,t}^2, Y_{s,t}^\sharp \rangle - \nabla^2 \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^\dagger X_{s,t}^1 + \tilde{Y}_s^{\dagger\dagger} X_{s,t}^2, \tilde{Y}_{s,t}^\sharp \rangle| \\
& \quad + \frac{1}{2} |\nabla^2 \sigma(Y_s) \langle Y_{s,t}^\sharp, Y_{s,t}^\sharp \rangle - \nabla^2 \sigma(\tilde{Y}_s) \langle \tilde{Y}_{s,t}^\sharp, \tilde{Y}_{s,t}^\sharp \rangle| \\
& \quad + \frac{1}{2} \left| \int_0^1 d\theta (1-\theta)^2 \{ \nabla^3 \sigma(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_{s,t}^1, Y_{s,t}^1 \rangle - \nabla^3 \sigma(\tilde{Y}_s + \theta \tilde{Y}_{s,t}^1) \langle \tilde{Y}_{s,t}^1, \tilde{Y}_{s,t}^1, \tilde{Y}_{s,t}^1 \rangle \} \right| \\
& \leq c_9 (K+1)^{\nu_4} (\|X\|_\alpha + 1)^{\nu_4} (|\xi - \tilde{\xi}| + \Delta) (t-s)^{3\beta}, \tag{3.41}
\end{aligned}$$

where we have used (3.19)–(3.27) again. Note that  $\nabla^5 \sigma$  is not involved in the above estimation. Combining (3.39)–(3.41), we obtain (3.38).

We estimate the difference of  $\mathcal{M}^1$ 's. Let us first calculate the  $\dagger\dagger$ -component.

$$\begin{aligned}
& |\{ \nabla \sigma(Y_t) Y_t^\dagger - \nabla \sigma(Y_s) Y_s^\dagger \} - \{ \nabla \sigma(\tilde{Y}_t) \tilde{Y}_t^\dagger - \nabla \sigma(\tilde{Y}_s) \tilde{Y}_s^\dagger \}| \\
& \leq \left| \int_0^1 d\theta \{ \nabla^2 \sigma(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_t^\dagger \rangle - \nabla^2 \sigma(\tilde{Y}_s + \theta \tilde{Y}_{s,t}^1) \langle \tilde{Y}_{s,t}^1, \tilde{Y}_t^\dagger \rangle \} \right. \\
& \quad \left. + |\nabla \sigma(Y_s) \langle Y_{s,t}^{\dagger,1} \rangle - \nabla \sigma(\tilde{Y}_s) \langle \tilde{Y}_{s,t}^{\dagger,1} \rangle| \right| \\
& \leq c_{10} (K+1)^{\nu_5} (\|X\|_\alpha + 1)^{\nu_5} (|\xi - \tilde{\xi}| + \Delta) (t-s)^\alpha. \tag{3.42}
\end{aligned}$$

Note that the right hand side of (3.42) has  $(t-s)^\alpha$ , not  $(t-s)^\beta$ .

Recall the explicit expression of the  $\sharp\sharp$ -component of  $\mathcal{M}^1$  in (3.34). Using it, we can estimate the difference as follows:

$$\begin{aligned}
& |\{ \sigma(Y_t) - \sigma(Y_s) - \nabla \sigma(Y_s) \langle Y_s^\dagger X_{s,t}^1 \rangle \} - \{ \sigma(\tilde{Y}_t) - \sigma(\tilde{Y}_s) - \nabla \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^\dagger X_{s,t}^1 \rangle \}| \\
& \leq \left| \int_0^1 d\theta (1-\theta) \{ \nabla^2 \sigma(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_{s,t}^1 \rangle - \nabla^2 \sigma(\tilde{Y}_s + \theta \tilde{Y}_{s,t}^1) \langle \tilde{Y}_{s,t}^1, \tilde{Y}_{s,t}^1 \rangle \} \right. \\
& \quad \left. + |\nabla \sigma(Y_s) \langle Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\sharp \rangle - \nabla \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^{\dagger\dagger} X_{s,t}^2 + \tilde{Y}_{s,t}^\sharp \rangle| \right|
\end{aligned}$$

$$\leq c_{11}(K+1)^{\nu_6}(\|X\|_{\alpha}+1)^{\nu_6}(|\xi-\tilde{\xi}|+\Delta)(t-s)^{2\alpha}. \quad (3.43)$$

Note that the right hand side of (3.43) has  $(t-s)^{2\alpha}$ , not  $(t-s)^{2\beta}$ .

We now turn to  $\sharp$ -component. Since the rough path integration is linear when the driving RP is fixed, our calculation is quite similar to that in (3.36).

$$\begin{aligned} & \left| \int_s^t \{\sigma(Y_u) - \sigma(\tilde{Y}_u)\} dX_u - \{\sigma(Y_s) - \sigma(\tilde{Y}_s)\} X_{s,t}^1 - \{\sigma(Y)_s^\dagger - \sigma(\tilde{Y})_s^\dagger\} X_{s,t}^2 \right| \\ & \leq |\{\sigma(Y)_s^{\dagger\dagger} - \sigma(\tilde{Y})_s^{\dagger\dagger}\} X_{s,t}^3| \\ & \quad + \left| \int_s^t \{\sigma(Y_u) - \sigma(\tilde{Y}_u)\} dX_u \right. \\ & \quad \left. - \left[ \{\sigma(Y_s) - \sigma(\tilde{Y}_s)\} X_{s,t}^1 + \{\sigma(Y)_s^\dagger - \sigma(\tilde{Y})_s^\dagger\} X_{s,t}^2 + \{\sigma(Y)_s^{\dagger\dagger} - \sigma(\tilde{Y})_s^{\dagger\dagger}\} X_{s,t}^3 \right] \right| \\ & \leq |\nabla \sigma(Y_s) \langle Y_s^{\dagger\dagger} \langle \bullet, \star \rangle, * \rangle - \nabla \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^{\dagger\dagger} \langle \bullet, \star \rangle, * \rangle| \cdot |X_{s,t}^3| \\ & \quad + |\nabla^2 \sigma(Y_s) \langle Y_s^\dagger \bullet, Y_s^\dagger \star, * \rangle - \nabla^2 \sigma(\tilde{Y}_s) \langle \tilde{Y}_s^\dagger \bullet, \tilde{Y}_s^\dagger \star, * \rangle| \cdot |X_{s,t}^3| \\ & \quad + \kappa_\beta (t-s)^{4\beta} \{ \|Y^\sharp - \tilde{Y}^\sharp\|_{3\beta} \|X^1\|_\beta + \|Y^{\sharp\sharp} - \tilde{Y}^{\sharp\sharp}\|_{2\beta} \|X^2\|_{2\beta} + \|Y^{\dagger\dagger} - \tilde{Y}^{\dagger\dagger}\|_\beta \|X^3\|_{3\beta} \} \\ & \leq c_{12}(K+1)^{\nu_7}(\|X\|_{\alpha}+1)^{\nu_7}(|\xi-\tilde{\xi}|+\Delta)(t-s)^{3\alpha}. \end{aligned} \quad (3.44)$$

Note that the right hand side of (3.44) has  $(t-s)^{3\alpha}$ , not  $(t-s)^{3\beta}$ . Combining (3.42)–(3.44), we easily obtain

$$\begin{aligned} & \|\mathcal{M}^1(Y, Y^\dagger, Y^{\dagger\dagger}) - \mathcal{M}^1(\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} \\ & \leq c_{13}(K+1)^{\nu_8}(\|X\|_{\alpha}+1)^{\nu_8}(|\xi-\tilde{\xi}|+\Delta)\tau^{\alpha-\beta}. \end{aligned} \quad (3.45)$$

From (3.45) and (3.37) we can estimate  $\mathcal{M}^\xi(Y, Y^\dagger, Y^{\dagger\dagger})$ . There are constants  $c_{14} \geq 4$  and  $\nu_9 \geq 2$  such that if

$$\tau \leq \lambda \quad \text{with} \quad \lambda := [c_{14}(K+1)^{\nu_9}(\|X\|_{\alpha}+1)^{\nu_9}]^{-1/(\alpha-\beta)} \in (0, 1), \quad (3.46)$$

then we have  $\mathcal{M}^\xi(B^\xi) \subset B^\xi$  for every  $\xi$  and

$$\|\mathcal{M}^\xi(Y, Y^\dagger, Y^{\dagger\dagger}) - \mathcal{M}^\xi(\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta} \leq \frac{1}{2}(|\xi-\tilde{\xi}|+\Delta). \quad (3.47)$$

In particular,  $\mathcal{M}^\xi$  is a contraction on  $B^\xi = B_{[0,\tau]}^\xi$  for every  $\xi$  and therefore has a unique fixed point in this ball. Thus, we have obtained a local solution of RDE (3.17) on  $[0, \lambda]$ . Note that  $\lambda$  is determined by  $\|X\|_{\alpha}$  and  $K$ , but independent of  $\xi$  and  $\psi$ .

Next, we do the same thing on the second interval  $[\lambda, (2\lambda) \wedge 1]$  with the initial condition  $\xi$  at  $t=0$  being replaced by  $Y_\lambda$  at  $t=\lambda$ . Since all the estimates above is independent of  $\xi$  and  $\psi$ ,  $(Y_s, Y_s^\dagger, Y_s^{\dagger\dagger})_{s \in [\lambda, (2\lambda) \wedge 1]}$  satisfies the same estimates as those for  $(Y_s, Y_s^\dagger, Y_s^{\dagger\dagger})_{s \in [0, \lambda]}$ . By concatenating them as in the fourth item of Example 3.1, we obtain a solution on  $[0, (2\lambda) \wedge 1]$ .

We can continue this procedure to obtain a global  $(Y_s, Y_s^\dagger, Y_s^{\dagger\dagger})_{s \in [0,1]}$ . There are (at most)  $\lfloor \lambda^{-1} \rfloor + 1$  subintervals, where  $\lfloor \cdot \rfloor$  stands for the integer part. Except (perhaps) the last one, the length of each interval equals  $\lambda$ . On each subinterval,  $(Y, Y^\dagger, Y^{\dagger\dagger})$  satisfies the same estimates. In particular, Inequality (3.25) implies that  $\beta$ -Hölder norm of  $Y$  on each subinterval is dominated by 1. By Hölder's inequality for finite sums, we can easily see that

$$\|Y\|_{\beta, [0,1]} \leq (\lfloor \lambda^{-1} \rfloor + 1)^{1-\beta} \leq c_{15} \{ (K+1) (\|X\|_\alpha + 1) \}^{\nu_{10}},$$

which is the desired estimate for a global solution.

We now check that any solution  $(Y, Y^\dagger, Y^{\dagger\dagger}) = (Y, \sigma(Y), \nabla\sigma \cdot \sigma(Y))$  of RDE (3.17) defined on  $[0, \tau']$  ( $0 < \tau' \leq 1$ ) actually belongs to  $\mathcal{Q}_X^\alpha([0, \tau'], \mathcal{W})$ . By the estimate (3.13), we have

$$\left| \int_s^t \sigma(Y_u) dX_u - (\sigma(Y_s) X_{s,t}^1 + \sigma(Y_s)^\dagger X_{s,t}^2 + \sigma(Y_s)^{\dagger\dagger} X_{s,t}^3) \right|_{\mathcal{W}} \leq C_\beta (t-s)^{4\beta},$$

where the constants  $C_\beta > 0$  depends only on  $\beta$ -Hölder RP norm of  $X$  and  $\beta$ -Hölder seminorm of  $(Y, Y^\dagger, Y^{\dagger\dagger})$ . Since  $X$  is  $\alpha$ -Hölder RP and  $\sigma(Y)^\dagger = \nabla\sigma \cdot \sigma(Y)$ , the above inequality implies that  $Y^\#$  is of  $3\alpha$ -Hölder. Likewise,  $Y$  is  $\alpha$ -Hölder continuous and so are  $Y^{\dagger\dagger}$ . We can compute  $Y^{\#\#}$  in a similar way as before as follows:

$$\begin{aligned} & Y_t^\dagger - Y_s^\dagger - Y_s^{\dagger\dagger} X_{s,t}^1 \\ &= \sigma(Y_t) - \sigma(Y_s) - \nabla\sigma(Y_s) \langle Y_{s,t}^1 \rangle + \nabla\sigma(Y_s) \langle Y_{s,t}^1 - \sigma(Y_s) X_{s,t}^1 \rangle \\ &= \int_0^1 d\theta \nabla^2 \sigma(Y_s + \theta Y_{s,t}^1) \langle Y_{s,t}^1, Y_{s,t}^1 \rangle + \nabla\sigma(Y_s) \langle Y_s^{\dagger\dagger} X_{s,t}^2 + Y_{s,t}^\# \rangle. \end{aligned}$$

The right hand side is dominated by a constant multiple of  $(t-s)^{2\alpha}$ , which implies that  $Y^{\#\#}$  is of  $2\alpha$ -Hölder. Thus, we have seen that any solution is actually  $\alpha$ -Hölder CP.

Finally, we show the uniqueness of solution. The uniqueness is a time-local issue, so it suffices to prove that any two solutions,  $(Y, \sigma(Y), \nabla\sigma \cdot \sigma(Y))$  and  $(\tilde{Y}, \sigma(\tilde{Y}), \nabla\sigma \cdot \sigma(\tilde{Y}))$ , of RDE (3.17) must coincide near  $t = 0$ . Since they are  $\alpha$ -Hölder CPs and  $\beta < \alpha$ , they both belong to  $B_{[0, \tau']}^\xi$  for sufficiently small  $\tau' > 0$ . Recall that we already know that there is only one fixed point of  $\mathcal{M}_{[0, \tau']}^\xi$  in this ball. Hence, these two local solutions must be identically equal on  $[0, \tau']$ .  $\square$

**Remark 3.4.** (i) By examining the proof of Proposition 3.3, one naturally realize the following: Just to prove the existence of a unique global solution RDE (3.17) for any given  $\psi$ ,  $X$  and  $\xi$ , it suffices to assume that  $\sigma$  is of  $C_b^4$  and  $f$  satisfies that

$$\sup_{y \in \mathcal{W}, t \in [0, T]} |f(y, \psi_t)|_{\mathcal{W}} + \sup_{y, y' \in \mathcal{W}, y \neq y', t \in [0, T]} \frac{|f(y, \psi_t) - f(y', \psi_t)|_{\mathcal{W}}}{|y - y'|_{\mathcal{W}}} < \infty.$$

(ii) By a standard cut-off argument, it immediately follows from (i) above that if  $\sigma$  is of  $C^4$  and  $f$  is locally Lipschitz continuous in the following sense

$$\sup_{|y| \vee |y'| \leq N, y \neq y', t \in [0, T]} \frac{|f(y, \psi_t) - f(y', \psi_t)|_{\mathcal{W}}}{|y - y'|_{\mathcal{W}}} < \infty, \quad N \in \mathbb{N},$$

then RDE (3.17) has a unique local solution for any given  $\psi$ ,  $X$  and  $\xi$ . Hence, a unique solution exists up to either the explosion time or the time horizon  $T$ .

Together with RDE (3.17), we also consider the following RDE on  $[0, T]$  with the same  $X$ ,  $\sigma$  and  $\xi$ :

$$\tilde{Y}_t = \xi + \int_0^t \tilde{f}(\tilde{Y}_s, \tilde{\psi}_s) ds + \int_0^t \sigma(\tilde{Y}_s) dX_s, \quad \tilde{Y}_t^\dagger = \sigma(\tilde{Y}_t), \quad \tilde{Y}_t^{\dagger\dagger} = \nabla \sigma(\tilde{Y}_t) \tilde{Y}_t^\dagger. \quad (3.48)$$

We assume that  $\tilde{f}: \mathcal{W} \times \mathcal{S} \rightarrow \mathcal{W}$  is also continuous and satisfies Condition (3.16). Let  $\tilde{\psi}: [0, T] \rightarrow \mathcal{S}$  be another continuous path in  $\mathcal{S}$ .

**Proposition 3.5.** *Let  $\sigma, f, \tilde{f}, \xi$  be as above. For  $X \in G\Omega_\alpha(\mathcal{V})$ ,  $\xi \in \mathcal{W}$  and  $\psi, \tilde{\psi}$ , denote by  $(Y, \sigma(Y), \nabla \sigma \cdot \sigma(Y))$  and  $(\tilde{Y}, \sigma(\tilde{Y}), \nabla \sigma \cdot \sigma(\tilde{Y}))$  the unique solutions of RDEs (3.17) and (3.48) on  $[0, T]$ , respectively. For a bounded, globally Lipschitz map  $g: \mathcal{W} \rightarrow \mathcal{W}$ , set*

$$M_t := (Y_t - \tilde{Y}_t) - \int_0^t \{g(Y_s) - g(\tilde{Y}_s)\} ds - \int_0^t \{\sigma(Y_s) - \sigma(\tilde{Y}_s)\} dX_s, \quad t \in [0, T]. \quad (3.49)$$

Then,  $M \in \mathcal{C}^1(\mathcal{W})$  and the following estimate holds for every  $\beta \in (\frac{1}{4}, \alpha)$ : there exist positive constants  $c$  and  $\nu$  such that

$$\|Y - \tilde{Y}\|_\beta \leq c \exp[c(K' + 1)^\nu (\|X\|_\alpha + 1)^\nu] \|M\|_{3\beta}. \quad (3.50)$$

Here, we set  $K' = \max\{\|\sigma\|_{C_b^4}, \|f\|_\infty, L_f, \|\tilde{f}\|_\infty, L_{\tilde{f}}, \|g\|_\infty, L_g\}$  and the constants  $c$  and  $\nu$  are independent of  $X, \xi, \psi, \tilde{\psi}, \sigma, f, \tilde{f}, g, M$ .

*Proof.* Without loss of generality we assume  $T = 1$ . For simplicity we write  $(Y, Y^\dagger, Y^{\dagger\dagger})$  and  $(\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger})$  for  $(Y, \sigma(Y), \nabla \sigma \cdot \sigma(Y))$  and  $(\tilde{Y}, \sigma(\tilde{Y}), \nabla \sigma \cdot \sigma(\tilde{Y}))$ , respectively. It is easy to see that  $M \in \mathcal{C}^1(\mathcal{W})$ . Hence, (3.49) is in fact an equality in  $\mathcal{Q}_X^\beta(\mathcal{W})$  with the  $\dagger$ -parts and  $\dagger\dagger$ -parts being clearly equal.

We set  $\lambda' := [c_{14}(K' + 1)^{\nu_9} (\|X\|_\alpha + 1)^{\nu_9}]^{-1/(\alpha - \beta)}$  by just replacing  $K$  by  $K'$  in (3.46). Set  $s_j := j\lambda'$  for  $0 \leq j \leq \lfloor 1/\lambda' \rfloor$  and  $s_N := 1$  with  $N := \lfloor 1/\lambda' \rfloor + 1$ . Then, on each subinterval  $[s_{j-1}, s_j]$ ,  $(Y, Y^\dagger, Y^{\dagger\dagger}) \in B_{[s_{j-1}, s_j]}^{\xi_j}$ ,  $(\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger}) \in B_{[s_{j-1}, s_j]}^{\tilde{\xi}_j}$  and the estimates in the proof of Proposition 3.3 are available (with  $K$  being replaced by  $K'$ ). Here, we set  $\xi_j := Y_{s_j}$  and  $\tilde{\xi}_j := \tilde{Y}_{s_j}$ . From (3.47) we have for all  $j$  that

$$\begin{aligned} & \left\| \int_{s_{j-1}}^{\cdot} \{g(Y_s) - g(\tilde{Y}_s)\} ds - \int_{s_{j-1}}^{\cdot} \{\sigma(Y_s) - \sigma(\tilde{Y}_s)\} dX_s \right\|_{\mathcal{Q}_X^\beta, [s_{j-1}, s_j]} \\ & \leq \frac{1}{2} |\xi_{j-1} - \tilde{\xi}_{j-1}| + \frac{1}{2} \|(Y, Y^\dagger, Y^{\dagger\dagger}) - (\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta, [s_{j-1}, s_j]}. \end{aligned}$$

Taking the seminorms of both sides of (3.49) on each subinterval, we can easily see that

$$\|(Y, Y^\dagger, Y^{\dagger\dagger}) - (\tilde{Y}, \tilde{Y}^\dagger, \tilde{Y}^{\dagger\dagger})\|_{\mathcal{Q}_X^\beta, [s_{j-1}, s_j]} \leq 2\|M\|_{3\beta} + |\xi_{j-1} - \tilde{\xi}_{j-1}|. \quad (3.51)$$

Plugging (3.51) into (3.26), we obtain for all  $j$  that

$$\begin{aligned} |Y_{s,t}^1 - \tilde{Y}_{s,t}^1| &\leq 2\{(K' + 1)^2(\|X\|_\alpha^2 + 1)|\xi_{j-1} - \tilde{\xi}_{j-1}| \\ &\quad + (\|X\|_\alpha^2 + 1)(2\|M\|_{3\beta} + |\xi_{j-1} - \tilde{\xi}_{j-1}|)\}(t-s)^\alpha \\ &\leq |\xi_{j-1} - \tilde{\xi}_{j-1}| + \|M\|_{3\beta}, \quad (s,t) \in \Delta_{[s_{j-1}, s_j]} \end{aligned} \quad (3.52)$$

and in particular

$$|\xi_j - \tilde{\xi}_j| \leq 2|\xi_{j-1} - \tilde{\xi}_{j-1}| + \|M\|_{3\beta}.$$

We have also used that  $t - s \leq \lambda'$  in (3.52). By mathematical induction, we have

$$|\xi_j - \tilde{\xi}_j| \leq (1 + 2^1 + \cdots + 2^{j-1})\|M\|_{3\beta} = (2^j - 1)\|M\|_{3\beta}, \quad 1 \leq j \leq N.$$

Then, it follows from (3.52) that

$$\begin{aligned} \|Y^1 - \tilde{Y}^1\|_{\beta, [s_{j-1}, s_j]} &\leq 2\{(K' + 1)^2(\|X\|_\alpha^2 + 1)|\xi_{j-1} - \tilde{\xi}_{j-1}| \\ &\quad + (\|X\|_\alpha^2 + 1)(2\|M\|_{3\beta} + |\xi_{j-1} - \tilde{\xi}_{j-1}|)\}(\lambda')^{\alpha-\beta} \\ &\leq 2^N\|M\|_{3\beta}, \quad 1 \leq j \leq N. \end{aligned} \quad (3.53)$$

By Hölder's inequality for finite sums and the trivial inequality  $N^{1-\beta}2^N \leq 2^{2N}$ , we see that

$$\begin{aligned} \|Y^1 - \tilde{Y}^1\|_{\beta, [0,1]} &\leq N^{1-\beta}2^N\|M\|_{3\beta} \\ &\leq \exp[2N(\log 2)]\|M\|_{3\beta} \\ &\leq \exp[2(\lfloor 1/\lambda' \rfloor + 1)(\log 2)]\|M\|_{3\beta}. \end{aligned} \quad (3.54)$$

By adjusting the positive constants  $c$  and  $\nu$ , we can easily obtain (3.50) from (3.54).  $\square$

## 4 Driving rough path of rough slow-fast system

In this section we construct the driving RP of our slow-fast system of RDEs. Unlike the second level case in the previous works [13, 8], this part is not so easy. We use (a very special case of) the anisotropic RP theory developed in [6], in particular, an anisotropic version of Lyons' extension theorem.

### 4.1 Anisotropic rough paths

We write  $\mathcal{V}_1 = \mathbb{R}^d$  and  $\mathcal{V}_2 = \mathbb{R}^e$  and set  $\mathcal{V} := \mathcal{V}_1 \oplus \mathcal{V}_2 = \mathbb{R}^{d+e}$ . Denote by  $\pi_i: \mathcal{V} \rightarrow \mathcal{V}_i$  ( $i = 1, 2$ ) the natural projection onto the  $i$ th component. Then,  $\mathcal{V}^{\otimes 2} = \bigoplus_{i,j=1,2} \mathcal{V}_i \otimes \mathcal{V}_j$  and  $\mathcal{V}^{\otimes 3} = \bigoplus_{i,j,k=1,2} \mathcal{V}_i \otimes \mathcal{V}_j \otimes \mathcal{V}_k$ . Denote by  $\pi_{ij}: \mathcal{V}^{\otimes 2} \rightarrow \mathcal{V}_i \otimes \mathcal{V}_j$  ( $i, j = 1, 2$ ) the natural projection onto the  $(i, j)$ th component. Similarly,  $\pi_{ijk}: \mathcal{V}^{\otimes 3} \rightarrow \mathcal{V}_i \otimes \mathcal{V}_j \otimes \mathcal{V}_k$  ( $i, j, k = 1, 2$ ) is defined.

For a continuous map  $(\Xi^1, \Xi^2, B^3): \Delta_T \rightarrow \mathcal{V} \oplus \mathcal{V}^{\otimes 2} \oplus \mathcal{V}_1^{\otimes 3}$ , we write  $B^1 := \pi_1 \langle \Xi^1 \rangle$ ,  $W^1 := \pi_2 \langle \Xi^1 \rangle$ ,  $B^2 := \pi_{11} \langle \Xi^2 \rangle$ ,  $W^2 := \pi_{22} \langle \Xi^2 \rangle$ ,  $I[B, W] := \pi_{12} \langle \Xi^2 \rangle$  and  $I[W, B] := \pi_{21} \langle \Xi^2 \rangle$ .

Let  $\alpha \in (\frac{1}{4}, \frac{1}{3}]$  and  $\gamma \in (\frac{1}{3}, \frac{1}{2}]$  with  $2\alpha + \gamma > 1$ . We say that a continuous map

$$(\Xi^1, \Xi^2, B^3): \Delta_T \rightarrow \mathcal{V} \oplus \mathcal{V}^{\otimes 2} \oplus \mathcal{V}_1^{\otimes 3}$$

is an anisotropic rough path (ARP) of roughness  $(\alpha, \gamma)$  if it satisfies the following two conditions:

(1) (Hölder regularity)

$$\max\{\|B^1\|_\alpha, \|B^2\|_{2\alpha}, \|B^3\|_{3\alpha}, \|W^1\|_\gamma, \|W^2\|_{2\gamma}, \|I[B, W]\|_{\alpha+\gamma}, \|I[W, B]\|_{\alpha+\gamma}\} < \infty.$$

(2) (Chen's relation) For all  $0 \leq s \leq u \leq t \leq T$ , it holds that

$$\begin{aligned} \Xi_{s,t}^1 &= \Xi_{s,u}^1 + \Xi_{u,t}^1, & \Xi_{s,t}^2 &= \Xi_{s,u}^2 + \Xi_{u,t}^2 + \Xi_{s,u}^1 \otimes \Xi_{u,t}^1, \\ B_{s,t}^3 &= B_{s,u}^3 + B_{u,t}^3 + B_{s,u}^1 \otimes B_{u,t}^2 + B_{s,u}^2 \otimes B_{u,t}^1. \end{aligned}$$

We denote by  $\hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$  the set of all ARP of roughness  $(\alpha, \gamma)$ . This set naturally becomes a complete metric space equipped with the Hölder norms. For  $z = (b, w) \in \mathcal{C}_0^1(\mathcal{V})$ , we can define its natural lift  $\hat{S}(z) \in \hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$  by using the Riemann-Stieltjes integration in a similar way as in the usual RP theory (for instance,  $\pi_{21} \langle \hat{S}(z)_{s,t} \rangle := \int_s^t \int_s^{u_2} dw_{u_1} \otimes db_{u_2}$ ). We define  $G\hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$  to be the closure of  $\{\hat{S}(z) \mid z = (b, w) \in \mathcal{C}_0^1(\mathcal{V})\}$ , which is a complete separable metric space by definition. An element of  $G\hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$  is called a geometric ARP of roughness  $(\alpha, \gamma)$ .

In fact, there is a canonical continuous injection  $G\hat{\Omega}_{\alpha,\gamma}(\mathcal{V}) \hookrightarrow G\Omega_\alpha(\mathcal{V})$ . We will explain it in what follows. Let  $(\Xi^1, \Xi^2, B^3) \in \hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$ . For  $(i, j, k) \in \{1, 2\}^3$  with  $(i, j, k) \neq (1, 1, 1)$  and  $(s, t) \in \Delta_T$ , we define

$$\Xi_{s,t}^{3,[ijk]} := \lim_{|\mathcal{P}| \searrow 0} \sum_{l=1}^N \pi_{ijk} \langle \Xi_{s,t_{l-1}}^1 \otimes \Xi_{t_{l-1},t_l}^2 + \Xi_{s,t_{l-1}}^2 \otimes \Xi_{t_{l-1},t_l}^1 \rangle, \quad (4.1)$$

where  $\mathcal{P} := \{s = t_0 < t_1 < \dots < t_N = t\}$  is a partition of  $[s, t]$ . As we will see, the limit on the right hand side exists. We set  $\Xi_{s,t}^3$  by  $\pi_{ijk} \langle \Xi_{s,t}^3 \rangle = \Xi_{s,t}^{3,[ijk]}$  if  $(i, j, k) \neq (1, 1, 1)$  and  $\pi_{111} \langle \Xi_{s,t}^3 \rangle = B_{s,t}^3$ . We can easily see that if  $(\Xi^1, \Xi^2, B^3) = \hat{S}(z)$  for some  $z = (b, w) \in \mathcal{C}_0^1(\mathcal{V})$ , then  $(\Xi^1, \Xi^2, \Xi^3) = S_3(z)$ , i.e. the natural third-level lift of  $z$ .

The following two propositions are key results in constructing our random RP. They are actually a special case of an anisotropic version of Lyons' extension theorem (see [6, Theorem 2.6]). Since our situation is quite simple, we will provide direct proofs below.

**Proposition 4.1.** *For every  $(\Xi^1, \Xi^2, B^3) \in \hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$ ,  $\Xi_{s,t}^{3,[ijk]}$  in (4.1) is well-defined. If*

$$\max\{\|B^1\|_\alpha, \|B^2\|_{2\alpha}, \|B^3\|_{3\alpha}, \|W^1\|_\gamma, \|W^2\|_{2\gamma}, \|I[B, W]\|_{\alpha+\gamma}, \|I[W, B]\|_{\alpha+\gamma}\} \leq M \quad (4.2)$$

holds for a constant  $M > 0$ , then there is a constant  $C = C_{\alpha, \gamma} > 0$  such that

$$\|\Xi^{3,[ijk]}\|_{\delta} \leq CM^2 \quad \text{where } \delta := (6 - i - j - k)\alpha + (i + j + k - 3)\gamma. \quad (4.3)$$

Moreover,  $(s, t) \mapsto (\Xi_{s,t}^1, \Xi_{s,t}^2, \Xi_{s,t}^3)$  satisfies Chen's relation. In particular,  $(\Xi^1, \Xi^2, \Xi^3) \in \Omega_{\alpha}(\mathcal{V})$ .

*Proof.* Pick any  $(s, t) \in \Delta_T$  with  $s < t$ . For  $\mathcal{P} = \{s = t_0 < t_1 < \dots < t_N = t\}$ , we set

$$A_{s,t}(\mathcal{P}) := \sum_{l=1}^N (\Xi_{s,t_{l-1}}^1 \otimes \Xi_{t_{l-1},t_l}^2 + \Xi_{s,t_{l-1}}^2 \otimes \Xi_{t_{l-1},t_l}^1).$$

We throw away  $t_m$  ( $1 \leq m \leq N - 1$ ) from  $\mathcal{P}$ . By Chen's relation, we have

$$A_{s,t}(\mathcal{P}) - A_{s,t}(\mathcal{P} \setminus \{t_m\}) = \Xi_{t_{m-1},t_m}^1 \otimes \Xi_{t_m,t_{m+1}}^2 + \Xi_{t_{m-1},t_m}^2 \otimes \Xi_{t_m,t_{m+1}}^1. \quad (4.4)$$

We only prove the case  $(i, j, k) = (1, 2, 1)$  for brevity. (The other cases can be done in the same way.) For  $\mathcal{P}$  given as above, we can find  $m$  ( $1 \leq m \leq N - 1$ ) such that  $t_{m+1} - t_{m-1} \leq 2(t - s)/(N - 1)$ . Then, we see that

$$\begin{aligned} & |\pi_{121}\langle A_{s,t}(\mathcal{P}) \rangle - \pi_{121}\langle A_{s,t}(\mathcal{P} \setminus \{t_m\}) \rangle| \\ & \leq |B_{t_{m-1},t_m}^1 \otimes I[W, B]_{t_m,t_{m+1}}| + |I[B, W]_{t_{m-1},t_m} \otimes B_{t_m,t_{m+1}}^1| \\ & \leq M^2(t_{m+1} - t_{m-1})^{2\alpha+\gamma} \\ & \leq M^2 \left( \frac{2}{N-1} \right)^{2\alpha+\gamma} (t - s)^{2\alpha+\gamma}. \end{aligned} \quad (4.5)$$

Repeating the above estimate, we have

$$|\pi_{121}\langle A_{s,t}(\mathcal{P}) \rangle| \leq M^2 2^{2\alpha+\gamma} \zeta(2\alpha + \gamma) (t - s)^{2\alpha+\gamma}. \quad (4.6)$$

Here, we have used  $2\alpha + \gamma > 1$ . Hence, if  $\Xi_{s,t}^{3,[121]} := \lim_{|\mathcal{P}| \searrow 0} \pi_{121}\langle A_{s,t}(\mathcal{P}) \rangle$  exists, it has the desired Hölder regularity.

Now we show that  $\{\pi_{121}\langle A_{s,t}(\mathcal{P}) \rangle\}_{\mathcal{P}}$  is Cauchy in  $\mathcal{P}$  as  $|\mathcal{P}| \searrow 0$ . First, consider the case  $\mathcal{Q} \supset \mathcal{P}$ , that is,  $\mathcal{Q}$  is a refinement of  $\mathcal{P}$ . On each subinterval  $[t_{l-1}, t_l]$  ( $1 \leq l \leq N$ ), we throw away points of  $\mathcal{Q} \cap [t_{l-1}, t_l]$  one by one in the same way as above. Then, we see from (4.5) that

$$\begin{aligned} |\pi_{121}\langle A_{s,t}(\mathcal{Q}) \rangle - \pi_{121}\langle A_{s,t}(\mathcal{P}) \rangle| & \leq M^2 2^{2\alpha+\gamma} \zeta(2\alpha + \gamma) \sum_{l=1}^N (t_l - t_{l-1})^{2\alpha+\gamma} \\ & \leq M^2 2^{2\alpha+\gamma} \zeta(2\alpha + \gamma) T |\mathcal{P}|^{2\alpha+\gamma-1}. \end{aligned}$$

For general  $\mathcal{P}$  and  $\mathcal{Q}$ , note that  $\mathcal{P} \cup \mathcal{Q}$  is a refinement of both  $\mathcal{P}$  and  $\mathcal{Q}$ . Hence, we have

$$|\pi_{121}\langle A_{s,t}(\mathcal{Q}) \rangle - \pi_{121}\langle A_{s,t}(\mathcal{P}) \rangle|$$

$$\begin{aligned} &\leq |\pi_{121}\langle A_{s,t}(\mathcal{Q}) \rangle - \pi_{121}\langle A_{s,t}(\mathcal{P} \cup \mathcal{Q}) \rangle| + |\pi_{121}\langle A_{s,t}(\mathcal{P} \cup \mathcal{Q}) \rangle - \pi_{121}\langle A_{s,t}(\mathcal{P}) \rangle| \\ &\leq 2M^2 2^{2\alpha+\gamma} \zeta (2\alpha + \gamma) T (|\mathcal{P}| \vee |\mathcal{Q}|)^{2\alpha+\gamma-1}. \end{aligned}$$

This estimate implies the desired Cauchy condition.

Next, we show Chen's relation. Pick  $0 \leq s < u < t \leq T$  arbitrarily. Let  $\mathcal{P}$  and  $\mathcal{Q}$  be a partition of  $[s, u]$  and  $[u, t]$ , respectively. Then,  $\mathcal{P} \cup \mathcal{Q}$  is a partition of  $[s, t]$  and  $|\mathcal{P} \cup \mathcal{Q}| = |\mathcal{P}| \vee |\mathcal{Q}|$ . By Chen's relation for  $(\Xi^1, \Xi^2)$ , we have

$$A_{s,t}(\mathcal{P} \cup \mathcal{Q}) = A_{s,u}(\mathcal{P}) + A_{u,t}(\mathcal{Q}) + \Xi_{s,u}^1 \otimes \Xi_{u,t}^2 + \Xi_{s,u}^2 \otimes \Xi_{u,t}^1.$$

Applying  $\pi_{121}$  to both sides and letting  $|\mathcal{P} \cup \mathcal{Q}| \searrow 0$ , we have

$$\Xi_{s,t}^{3,[121]} = \Xi_{s,u}^{3,[121]} + \Xi_{u,t}^{3,[121]} + B_{s,u}^1 \otimes I[W, B]_{u,t} + I[B, W]_{s,u} \otimes B_{u,t}^1.$$

From this and the Hölder estimates, we can also show that  $\Delta \ni (s, t) \mapsto \Xi_{s,t}^{3,[121]} \in \mathcal{V}_1 \otimes \mathcal{V}_2 \otimes \mathcal{V}_1$  is continuous.  $\square$

We write  $(\Xi^1, \Xi^2, \Xi^3) = \mathbf{Ext}(\Xi^1, \Xi^2, B^3)$ . As we have just seen,  $\mathbf{Ext}$  is an injection from  $\hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$  to  $\Omega_\alpha(\mathcal{V})$ . Below we will check that this injection is locally Lipschitz continuous, i.e. Lipschitz continuous on every bounded subset of  $\hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$ .

**Proposition 4.2.** *Let  $(\Xi^1, \Xi^2, B^3), (\tilde{\Xi}^1, \tilde{\Xi}^2, \tilde{B}^3) \in \hat{\Omega}_{\alpha,\gamma}(\mathcal{V})$ . Suppose that (4.2) holds for both  $(\Xi^1, \Xi^2, B^3)$  and  $(\tilde{\Xi}^1, \tilde{\Xi}^2, \tilde{B}^3)$  with a common constant  $M > 0$  and*

$$\begin{aligned} &\max\{\|B^1 - \tilde{B}^1\|_\alpha, \|B^2 - \tilde{B}^2\|_{2\alpha}, \|B^3 - \tilde{B}^3\|_{3\alpha}, \|W^1 - \tilde{W}^1\|_\gamma, \|W^2 - \tilde{W}^2\|_{2\gamma}, \\ &\quad \|I[B, W] - I[\tilde{B}, \tilde{W}]\|_{\alpha+\gamma}, \|I[W, B] - I[\tilde{W}, \tilde{B}]\|_{\alpha+\gamma}\} \leq \varepsilon \end{aligned}$$

holds for a constant  $\varepsilon > 0$ , then there is a constant  $C' = C'(M, \alpha, \gamma) > 0$  (independent of  $\varepsilon$ ) such that

$$\|\Xi^{3,[ijk]} - \tilde{\Xi}^{3,[ijk]}\|_\delta \leq C' \varepsilon \quad \text{where } \delta := (6 - i - j - k)\alpha + (i + j + k - 3)\gamma. \quad (4.7)$$

In particular,  $\mathbf{Ext}: G\hat{\Omega}_{\alpha,\gamma}(\mathcal{V}) \hookrightarrow G\Omega_\alpha(\mathcal{V})$  is locally Lipschitz continuous.

*Proof.* We use the same notation as in the proof of Proposition 4.1. We see from (4.4) that

$$\begin{aligned} &\{A_{s,t}(\mathcal{P}) - \tilde{A}_{s,t}(\mathcal{P})\} - \{A_{s,t}(\mathcal{P} \setminus \{t_m\}) - \tilde{A}_{s,t}(\mathcal{P} \setminus \{t_m\})\} \\ &= \{\Xi_{t_{m-1}, t_m}^1 \otimes \Xi_{t_m, t_{m+1}}^2 - \tilde{\Xi}_{t_{m-1}, t_m}^1 \otimes \tilde{\Xi}_{t_m, t_{m+1}}^2\} \\ &\quad + \{\Xi_{t_{m-1}, t_m}^2 \otimes \Xi_{t_m, t_{m+1}}^1 - \tilde{\Xi}_{t_{m-1}, t_m}^2 \otimes \tilde{\Xi}_{t_m, t_{m+1}}^1\}. \end{aligned}$$

If  $m$  ( $1 \leq m \leq N - 1$ ) is such that  $t_{m+1} - t_{m-1} \leq 2(t - s)/(N - 1)$ , then

$$\begin{aligned} &|\pi_{121}\langle A_{s,t}(\mathcal{P}) - \tilde{A}_{s,t}(\mathcal{P}) \rangle - \pi_{121}\langle A_{s,t}(\mathcal{P} \setminus \{t_m\}) - \tilde{A}_{s,t}(\mathcal{P} \setminus \{t_m\}) \rangle| \\ &= |B_{t_{m-1}, t_m}^1 \otimes I[W, B]_{t_m, t_{m+1}} - \tilde{B}_{t_{m-1}, t_m}^1 \otimes I[\tilde{W}, \tilde{B}]_{t_m, t_{m+1}}| \end{aligned}$$

$$\begin{aligned}
& + |I[B, W]_{t_{m-1}, t_m} \otimes B_{t_m, t_{m+1}}^1 - I[\tilde{B}, \tilde{W}]_{t_{m-1}, t_m} \otimes \tilde{B}_{t_m, t_{m+1}}^1| \\
& \leq 4M\varepsilon(t_{m+1} - t_{m-1})^{2\alpha+\gamma} \\
& \leq 4M\varepsilon \left( \frac{2}{N-1} \right)^{2\alpha+\gamma} (t-s)^{2\alpha+\gamma}.
\end{aligned}$$

Using the same argument as in the proof of Proposition 4.1, we have

$$|\pi_{121}\langle A_{s,t}(\mathcal{P}) - \pi_{121}\langle \tilde{A}_{s,t}(\mathcal{P}) \rangle| \leq 4M\varepsilon 2^{2\alpha+\gamma} \zeta(2\alpha+\gamma) (t-s)^{2\alpha+\gamma}.$$

Letting  $|\mathcal{P}| \searrow 0$ , we obtain (4.7) for  $(i, j, k) = (1, 2, 1)$ . (The other cases can be shown in the same way.) Thus, we have proved the local Lipschitz continuity of  $\mathbf{Ext}: \hat{\Omega}_{\alpha,\gamma}(\mathcal{V}) \rightarrow \Omega_\alpha(\mathcal{V})$ . Combining this with  $\mathbf{Ext} \circ \hat{S} = S_3$ , we see that  $\mathbf{Ext}(G\hat{\Omega}_{\alpha,\gamma}(\mathcal{V})) \subset G\Omega_\alpha(\mathcal{V})$ .  $\square$

## 4.2 An anisotropic version of Kolmogorov's continuity criterion for random rough paths

In this subsection, we prove a Kolmogorov-type continuity criterion for random ARPs of second level. A standard version of this criterion in the RP setting is found in [3, Theorem 3.1]. We will slightly modify the proof of that theorem.

Let  $(Q, R, I) = \{(Q_{s,t}, R_{s,t}, I_{s,t})\}_{(s,t) \in \Delta_T}$  be an  $\mathbb{R}^3$ -valued (two-parameter) stochastic process defined on a certain probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with the following two properties:

- Chen's relation in a probabilistic sense, that is, for every  $0 \leq s \leq u \leq t \leq T$ ,

$$\mathbb{P}(Q_{s,t} = Q_{s,u} + Q_{u,t}, R_{s,t} = R_{s,u} + R_{u,t}, I_{s,t} = I_{s,u} + I_{u,t} + Q_{s,u}R_{u,t}) = 1. \quad (4.8)$$

- An  $L^q$ -version,  $q \in [2, \infty)$ , of the Hölder condition with exponent  $\beta_1, \beta_2 \in (0, 1]$ , that is, there exists  $C > 0$  such that for every  $(s, t) \in \Delta_T$

$$\|Q_{s,t}\|_{L^q} \leq C(t-s)^{\beta_1}, \|R_{s,t}\|_{L^q} \leq C(t-s)^{\beta_2}, \|I_{s,t}\|_{L^{q/2}} \leq C(t-s)^{\beta_1+\beta_2}. \quad (4.9)$$

**Proposition 4.3.** *Let  $(Q, R, I)$  be as above and let  $q \in [2, \infty)$  and  $\beta_1, \beta_2 \in (1/q, 1]$ . Assume that (4.8) and (4.9) hold. Then, there exists a continuous modification of  $(Q, R, I)$  (denoted by the same symbol again) with the following two properties:*

- *Chen's relation holds almost surely, that is,*

$$\begin{aligned}
\mathbb{P}(Q_{s,t} = Q_{s,u} + Q_{u,t}, R_{s,t} = R_{s,u} + R_{u,t}, I_{s,t} = I_{s,u} + I_{u,t} + Q_{s,u}R_{u,t} \\
\text{for all } 0 \leq s \leq u \leq t \leq T) = 1
\end{aligned} \quad (4.10)$$

- *For every  $\alpha_1 \in (0, \beta_1 - 1/q)$  and  $\alpha_2 \in (0, \beta_2 - 1/q)$ , there exist non-negative random variables  $M^Q, M^R \in L^q$  and  $M^I \in L^{q/2}$  such that*

$$\begin{aligned}
\mathbb{P}(|Q_{s,t}| \leq M^Q(t-s)^{\alpha_1}, |R_{s,t}| \leq M^R(t-s)^{\alpha_2}, \\
|I_{s,t}| \leq M^I(t-s)^{\alpha_1+\alpha_2} \text{ for all } (s, t) \in \Delta_T) = 1.
\end{aligned} \quad (4.11)$$

*Proof.* Without loss of generality, we may assume  $T = 1$ . Set  $D_n = \{k2^{-n} \mid 0 \leq k \leq 2^n\}$  for  $n \in \mathbb{N}_0$  and set  $D = \cup_{n=0}^{\infty} D_n$ . By (4.8) there exists  $\Omega_1 \in \mathcal{F}$  with  $\mathbb{P}(\Omega_1) = 1$  on which

$$Q_{s,t} = Q_{s,u} + Q_{u,t}, \quad R_{s,t} = R_{s,u} + R_{u,t}, \quad I_{s,t} = I_{s,u} + I_{u,t} + Q_{s,u}R_{u,t}$$

holds for all  $(s, u, t) \in D^3$  with  $s \leq u \leq t$ . In particular, it holds on  $\Omega_1$  that  $Q_{s,s} = R_{s,s} = I_{s,s} = 0$  for all  $s \in D$ . From here we will work on  $\Omega_1$ .

We set

$$K_n^Q = \sup_{1 \leq i \leq 2^n} |Q_{(i-1)2^{-n}, i2^{-n}}|, \quad n \in \mathbb{N}_0. \quad (4.12)$$

We also set  $K_n^R$  and  $K_n^I$  in the same way. Next, define  $M^Q$ ,  $M^R$  and  $M^I$  by

$$M^Q := 2 \sum_{n=0}^{\infty} 2^{n\alpha_1} K_n^Q, \quad M^R := 2 \sum_{n=0}^{\infty} 2^{n\alpha_2} K_n^R, \quad M^I := M^Q M^R + 2 \sum_{n=0}^{\infty} 2^{n(\alpha_1+\alpha_2)} K_n^I.$$

We can easily show

$$\mathbb{E}[|K_n^Q|^q] \leq \sum_{i=1}^{2^n} E[|Q_{(i-1)2^{-n}, i2^{-n}}|^q] \leq 2^n C^q (2^{-n})^{\beta_1 q} \leq C^q (2^{-n})^{\beta_1 q - 1}$$

and, similarly,  $\mathbb{E}[|K_n^R|^q] \leq C^q (2^{-n})^{\beta_2 q - 1}$  and  $\mathbb{E}[|K_n^I|^{q/2}] \leq C^{q/2} (2^{-n})^{(\beta_1+\beta_2)q/2 - 1}$ .

Then, we have

$$\|M^Q\|_{L^q} \leq 2 \sum_{n=0}^{\infty} 2^{n\alpha_1} \|K_n^Q\|_{L^q} \leq 2C \sum_{n=0}^{\infty} (2^{-n})^{-\alpha_1 + \beta_1 - 1/q} < \infty,$$

where we have used  $\beta_1 - 1/q > \alpha_1$ . In the same way, we have  $\|M^R\|_{L^q} < \infty$ . From Schwarz' inequality,  $\|M^Q M^R\|_{L^{q/2}} \leq \|M^Q\|_{L^q}^{1/2} \|M^R\|_{L^q}^{1/2} < \infty$  immediately follows. By a similar argument as above we also have

$$\|M^I - M^Q M^R\|_{L^{q/2}} \leq 2 \sum_{n=0}^{\infty} 2^{n(\alpha_1+\alpha_2)} \|K_n^I\|_{L^{q/2}} \leq 2C \sum_{n=0}^{\infty} (2^{-n})^{-\alpha_1 - \alpha_2 + \beta_1 + \beta_2 - 2/q} < \infty.$$

Thus, we have shown  $\|M^I\|_{L^{q/2}} < \infty$ .

Take any  $(s, t) \in D^2$  with  $s < t$ . Then, there uniquely exists  $m \in \mathbb{N}_0$  such that  $2^{-(m+1)} < t - s \leq 2^{-m}$ . Moreover, there exists a partition  $\{s = \tau_0 < \tau_1 \cdots < \tau_N = t\}$  with  $\tau_i \in D$  ( $0 \leq i \leq N$ ) such that (1)  $(\tau_{i-1}, \tau_i) \in D_n \times D_n$  for some  $n \geq m + 1$  ( $1 \leq i \leq N$ ) and (2) at most two of  $[\tau_{i-1}, \tau_i]$ 's have length  $2^{-n}$  for every  $n \geq m + 1$ .

Then we have

$$|Q_{s,t}| \leq \max_{1 \leq i \leq N} |Q_{s,\tau_i}| \leq \sum_{i=1}^N |Q_{\tau_{i-1}, \tau_i}| \leq 2 \sum_{n=m+1}^{\infty} K_n^Q$$

and

$$\frac{|Q_{s,t}|}{(t-s)^{\alpha_1}} \leq 2^{\alpha_1(m+1)} 2 \sum_{n=m+1}^{\infty} K_n^Q \leq 2 \sum_{n=m+1}^{\infty} 2^{\alpha_1 n} K_n^Q \leq M^Q.$$

In the same way, we have  $|R_{s,t}|/(t-s)^{\alpha_2} \leq M^R$ , too.

Next we estimate  $I_{s,t}$ . By Chen's relation, we have

$$I_{s,t} = \sum_{i=1}^N (I_{\tau_{i-1}, \tau_i} + Q_{s, \tau_{i-1}} R_{\tau_{i-1}, \tau_i}) \quad (4.13)$$

and therefore

$$\begin{aligned} |I_{s,t}| &\leq \sum_{i=1}^N |I_{\tau_{i-1}, \tau_i}| + \max_{1 \leq i \leq N} |Q_{s, \tau_i}| \sum_{i=1}^N |R_{\tau_{i-1}, \tau_i}| \\ &\leq 2 \sum_{n=m+1}^{\infty} K_n^I + \left( 2 \sum_{n=m+1}^{\infty} K_n^Q \right) \left( 2 \sum_{n=m+1}^{\infty} K_n^R \right). \end{aligned}$$

Repeating a similar argument as above, we can easily see that

$$\frac{|I_{s,t}|}{(t-s)^{\alpha_1+\alpha_2}} \leq 2 \sum_{n=m+1}^{\infty} 2^{(\alpha_1+\alpha_2)n} K_n^I + \left( 2 \sum_{n=m+1}^{\infty} 2^{\alpha_1 n} K_n^Q \right) \left( 2 \sum_{n=m+1}^{\infty} 2^{\alpha_2 n} K_n^R \right) \leq M^I.$$

From here, we construct a continuous modification. Set

$$\Omega_2 = \{\omega \in \Omega_1 \mid M^Q(\omega) \vee M^R(\omega) \vee M^I(\omega) < \infty\}.$$

Clearly,  $\mathbb{P}(\Omega_2) = 1$ . For  $\omega \in \Omega_2$ ,  $(Q(\omega), R(\omega), I(\omega))$  satisfies Chen's relation on  $\Delta_1 \cap D^2$  and the following estimates:

$$\begin{aligned} |Q_{s,t}(\omega)| &\leq M^Q(\omega)(t-s)^{\alpha_1}, \quad |R_{s,t}(\omega)| \leq M^R(\omega)(t-s)^{\alpha_2}, \\ |I_{s,t}(\omega)| &\leq M^I(\omega)(t-s)^{\alpha_1+\alpha_2} \quad \text{for every } (s, t) \in \Delta_1 \cap D^2. \end{aligned} \quad (4.14)$$

This implies that the map

$$\Delta_1 \cap D^2 \ni (s, t) \mapsto (Q_{s,t}(\omega), R_{s,t}(\omega), I_{s,t}(\omega)) \in \mathbb{R}^3$$

is uniformly continuous. Since  $\Delta_1 \cap D^2$  is dense in  $\Delta_1$ , this map extends to a uniformly continuous map from  $\Delta_1$  to  $\mathbb{R}^3$ , which is denoted by  $(\tilde{Q}(\omega), \tilde{R}(\omega), \tilde{I}(\omega))$ . More concretely, we take a sequence  $\{(s_n, t_n)\}_{n \in \mathbb{N}} \subset \Delta_1 \cap D^2$  which converges to  $(s, t) \in \Delta_1$  and define

$$(\tilde{Q}_{s,t}(\omega), \tilde{R}_{s,t}(\omega), \tilde{I}_{s,t}(\omega)) = \lim_{n \rightarrow \infty} (Q_{s_n, t_n}(\omega), R_{s_n, t_n}(\omega), I_{s_n, t_n}(\omega)).$$

Of course, the limit does not depend on the choice of  $\{(s_n, t_n)\}$ . Obviously, this extended map satisfies Chen's relation and the inequalities in (4.14) for every  $(s, t) \in \Delta_1$ .

Finally, we prove that  $(\tilde{Q}, \tilde{R}, \tilde{I})$  is a modification of  $(Q, R, I)$ . Pick any  $(s, t) \in \Delta$  and take a sequence  $\{(s_n, t_n)\}_{n \in \mathbb{N}} \subset \Delta_1 \cap D^2$  which converges to  $(s, t) \in \Delta_1$ . We have just seen that  $(Q_{s_n, t_n}, R_{s_n, t_n}, I_{s_n, t_n})$  converges to  $(\tilde{Q}_{s, t}, \tilde{R}_{s, t}, \tilde{I}_{s, t})$  a.s. as  $n \rightarrow \infty$ . On the other hand, it follows from (4.8) and (4.9) that  $(Q_{s_n, t_n}, R_{s_n, t_n}, I_{s_n, t_n})$  converges to  $(Q_{s, t}, R_{s, t}, I_{s, t})$  in  $L^1$  as  $n \rightarrow \infty$ . Thus, we have shown that  $(\tilde{Q}_{s, t}, \tilde{R}_{s, t}, \tilde{I}_{s, t}) = (Q_{s, t}, R_{s, t}, I_{s, t})$  holds a.s. for every fixed  $(s, t) \in \Delta_1$ .  $\square$

Next we estimate the difference of two  $\mathbb{R}^3$ -valued (two-parameter) stochastic processes  $(Q, R, I) = \{(Q_{s, t}, R_{s, t}, I_{s, t})\}_{(s, t) \in \Delta_T}$  and  $(\hat{Q}, \hat{R}, \hat{I}) = \{(\hat{Q}_{s, t}, \hat{R}_{s, t}, \hat{I}_{s, t})\}_{(s, t) \in \Delta_T}$ .

**Proposition 4.4.** *Let  $q \in [2, \infty)$  and  $\beta_1, \beta_2 \in (1/q, 1]$ . Assume that both  $(Q, R, I)$  and  $(\hat{Q}, \hat{R}, \hat{I})$  satisfy (4.8)–(4.9) with a common constant  $C > 0$ . (Their continuous modifications as in Proposition 4.3 are denoted by the same symbols again.) Assume further that*

$$\begin{aligned} \|Q_{s, t} - \hat{Q}_{s, t}\|_{L^q} &\leq \varepsilon(t - s)^{\beta_1}, & \|R_{s, t} - \hat{R}_{s, t}\|_{L^q} &\leq \varepsilon(t - s)^{\beta_2}, \\ \|I_{s, t} - \hat{I}_{s, t}\|_{L^{q/2}} &\leq \varepsilon(t - s)^{\beta_1 + \beta_2} & \text{for all } (s, t) \in \Delta_T. \end{aligned} \quad (4.15)$$

holds for a constant  $\varepsilon > 0$ . Then, for every  $\alpha_1 \in (0, \beta_1 - 1/q)$  and  $\alpha_2 \in (0, \beta_2 - 1/q)$ , there exists a constant  $C' > 0$  such that

$$\|\|Q - \hat{Q}\|_{\alpha_1}\|_{L^q} + \|\|R - \hat{R}\|_{\alpha_2}\|_{L^q} + \|\|I - \hat{I}\|_{\alpha_1 + \alpha_2}\|_{L^{q/2}} \leq C' \varepsilon. \quad (4.16)$$

Here,  $C'$  depends on  $q, \beta_1, \beta_2, \alpha_1, \alpha_2, C, T$ , but not on  $\varepsilon$ .

*Proof.* We assume  $T = 1$  again. We argue in the same way and use the same symbols as in the proof of Proposition 4.3.

By replacing  $Q$  by  $Q - \hat{Q}$  in (4.12), we set  $K_n^{Q - \hat{Q}}$ . We set  $K_n^{R - \hat{R}}$  and  $K_n^{I - \hat{I}}$  in a similar way. Then, we have

$$\mathbb{E}[|K_n^{Q - \hat{Q}}|^q] \leq \sum_{i=1}^{2^n} E[|Q_{(i-1)2^{-n}, i2^{-n}} - \hat{Q}_{(i-1)2^{-n}, i2^{-n}}|^q] \leq \varepsilon^q (2^{-n})^{\beta_1 q - 1}. \quad (4.17)$$

We also have  $\mathbb{E}[|K_n^{R - \hat{R}}|^q] \leq \varepsilon^q (2^{-n})^{\beta_2 q - 1}$  and  $\mathbb{E}[|K_n^{I - \hat{I}}|^{q/2}] \leq \varepsilon^{q/2} (2^{-n})^{(\beta_1 + \beta_2)q/2 - 1}$ .

Take any  $(s, t) \in D^2$  with  $s < t$  and let  $\{s = \tau_0 < \tau_1 \dots < \tau_N = t\}$  be the partition as in the proof of Proposition 4.3. From (4.17) we can easily see that

$$|Q_{s, t} - \hat{Q}_{s, t}| \leq \max_{1 \leq i \leq N} |Q_{s, \tau_i} - \hat{Q}_{s, \tau_i}| \leq \sum_{i=1}^N |Q_{\tau_{i-1}, \tau_i} - \hat{Q}_{\tau_{i-1}, \tau_i}| \leq 2 \sum_{n=m+1}^{\infty} K_n^{Q - \hat{Q}}$$

and

$$\frac{|Q_{s, t} - \hat{Q}_{s, t}|}{(t - s)^{\alpha_1}} \leq 2^{\alpha_1(m+1)} 2 \sum_{n=m+1}^{\infty} K_n^{Q - \hat{Q}} \leq 2 \sum_{n=m+1}^{\infty} 2^{\alpha_1 n} K_n^{Q - \hat{Q}}$$

and therefore

$$\|\|Q - \hat{Q}\|_{\alpha_1}\|_{L^q} \leq 2 \sum_{n=1}^{\infty} 2^{\alpha_1 n} \|K_n^{Q-\hat{Q}}\|_{L^q} \leq 2\varepsilon \sum_{n=m+1}^{\infty} (2^{-n})^{-\alpha_1 + \beta_1 - 1/q} = C'_1 \varepsilon,$$

where we set  $C'_1 := 2 \sum_{n=0}^{\infty} (2^{-n})^{-\alpha_1 + \beta_1 - 1/q} < \infty$ . In essentially the same way, we can also show  $\|\|R - \hat{R}\|_{\alpha_2}\|_{L^q} \leq C'_2 \varepsilon$  for some constant  $C'_2 > 0$ .

Next we turn to the  $I$ -component. Since both  $I$  and  $\hat{I}$  satisfy (4.13), we see that

$$\begin{aligned} |I_{s,t} - \hat{I}_{s,t}| &\leq \sum_{i=1}^N |I_{\tau_{i-1}, \tau_i} - \hat{I}_{\tau_{i-1}, \tau_i}| \\ &\quad + \max_{1 \leq i \leq N} |Q_{s, \tau_i} - \hat{Q}_{s, \tau_i}| \sum_{i=1}^N |R_{\tau_{i-1}, \tau_i}| \\ &\quad + \max_{1 \leq i \leq N} |\hat{Q}_{s, \tau_i}| \sum_{i=1}^N |R_{\tau_{i-1}, \tau_i} - \hat{R}_{\tau_{i-1}, \tau_i}|. \\ &\leq 2 \sum_{n=m+1}^{\infty} K_n^{I-\hat{I}} + \left(2 \sum_{n=m+1}^{\infty} K_n^{Q-\hat{Q}}\right) \left(2 \sum_{n=m+1}^{\infty} K_n^R\right) \\ &\quad + \left(2 \sum_{n=m+1}^{\infty} K_n^Q\right) \left(2 \sum_{n=m+1}^{\infty} K_n^{R-\hat{R}}\right). \end{aligned}$$

By using a similar argument as above, we can easily see from the above estimate that  $\|\|I - \hat{I}\|_{\alpha_1 + \alpha_2}\|_{L^{q/2}} \leq C'_3 \varepsilon$  holds for some constant  $C'_3 > 0$ . This completes the proof.  $\square$

### 4.3 Lemmas for the geometric property of random anisotropic rough paths

The Kolmogorov-type continuity criterion alone is not sufficient for proving the geometric property of random ARPs which drives our slow-fast system. In this subsection we will provide a few lemmas for that purpose.

Let  $w = (w_t)_{t \in [0, T]}$  be a one-dimensional Brownian motion defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . For  $m \in \mathbb{N}$ ,  $w(m) = (w(m)_t)_{t \in [0, T]}$  be the  $m$ th dyadic piecewise linear approximation of  $w$ , that is the piecewise linear approximation associated with the partition  $\{kT2^{-m} \mid 0 \leq k \leq 2^m\}$  of  $[0, T]$ . Let  $g \in \mathcal{C}_0^\alpha(\mathbb{R})$  with  $\alpha \in (0, 1)$ . (Note that  $g$  is not random). We set, for  $(s, t) \in \Delta_T$ ,

$$I[g, w]_{s,t} = \int_s^t (g_u - g_s) d^I w_u, \tag{4.18}$$

$$I[g, w(m)]_{s,t} = \int_s^t (g_u - g_s) dw(m)_u. \tag{4.19}$$

Here, the integral in (4.18) is an Itô integral, while that in (4.19) is a (random) Riemann-Stieltjes integral.

**Lemma 4.5.** *For every  $q \in [1, \infty)$  and  $\alpha \in (0, 1)$ , there exists a constant  $C = C_{q,\alpha} > 0$  such that*

$$\|I[g, w]_{s,t}\|_{L^q} \vee \sup_{m \in \mathbb{N}} \|I[g, w(m)]_{s,t}\|_{L^q} \leq C(t-s)^{\alpha+\frac{1}{2}} \|g\|_\alpha, \quad (s, t) \in \Delta_T, \quad g \in \mathcal{C}_0^\alpha(\mathbb{R}).$$

Here,  $C$  does not depend on  $(s, t)$  or  $g$ .

*Proof.* In this proof,  $C_j$  ( $0 \leq j \leq 4$ ) are certain positive constants which depend on  $q$  and  $\alpha$  only. Without loss of generality we may assume  $T = 1$ .

By Burkholder's inequality we have

$$\mathbb{E} [|I[g, w]_{s,t}|^q] \leq C_0 \left( \int_s^t |g_u - g_s|^2 du \right)^{\frac{q}{2}} \leq C_0 \|g\|_\alpha^q (2\alpha + 1)^{-q/2} (t-s)^{(2\alpha+1)q/2},$$

which is the desired estimate for  $I[g, w]$ .

We calculate  $I[g, w(m)]$  for given  $m \in \mathbb{N}$ . We write  $t_k^m := k2^{-m}$  ( $0 \leq k \leq 2^m$ ) and  $\Delta_k^m w := w_{t_k^m} - w_{t_{k-1}^m}$  for simplicity. Note that the law of  $\Delta_k^m w / 2^{m/2}$  is the standard normal distribution. We will write  $g_{s,t}^1 = g_t - g_s$  as usual. Set

$$[g]_k^m := 2^m \int_{t_{k-1}^m}^{t_k^m} g_u du, \quad m \in \mathbb{N}, \quad 1 \leq k \leq 2^m.$$

First, consider the case that  $s$  and  $t$  belong to the same subinterval, that is, there exists  $k$  such that  $s, t \in [t_{k-1}^m, t_k^m]$ . Then, we can easily see that

$$\begin{aligned} |I[g, w(m)]_{s,t}| &= \left| \int_s^t g_{s,u}^1 \frac{\Delta_k^m w}{2^{-m}} du \right| \\ &\leq \|g\|_\alpha \frac{(t-s)^{\alpha+1}}{\alpha+1} \left| \frac{\Delta_k^m w}{2^{-m}} \right| \leq \|g\|_\alpha \frac{(t-s)^{\alpha+\frac{1}{2}}}{\alpha+1} \left| \frac{\Delta_k^m w}{2^{-m/2}} \right| \end{aligned}$$

and therefore

$$\|I[g, w(m)]_{s,t}\|_{L^q} \leq C_1 \|g\|_\alpha (t-s)^{\alpha+\frac{1}{2}} \quad (4.20)$$

for some constant  $C_1 > 0$ .

Next, consider the case  $s \in [t_{k-1}^m, t_k^m]$  and  $t \in [t_l^m, t_{l+1}^m]$  for some  $k, l$  with  $k \leq l$ . By Chen's relation we have

$$\begin{aligned} I[g, w(m)]_{s,t} &= I[g, w(m)]_{s,t_k^m} + I[g, w(m)]_{t_k^m, t_l^m} + I[g, w(m)]_{t_l^m, t} \\ &\quad + g_{s,t_k^m}^1 w(m)_{t_k^m, t_l^m}^1 + g_{s,t_k^m}^1 w(m)_{t_l^m, t}^1 + g_{t_k^m, t_l^m}^1 w(m)_{t_l^m, t}^1 \\ &=: A_1 + \cdots + A_6. \end{aligned} \quad (4.21)$$

Note that  $A_1$  and  $A_3$  were already estimated in (4.20). Since  $w(m)_{t_k^m, t_l^m}^1 = w_{t_k^m, t_l^m}^1$ , we see that  $\|A_4\|_{L^q} \leq C_2 \|g\|_\alpha (t - s)^{\alpha+1/2}$ . Since  $w(m)_{t_l^m, t}^1 = [\Delta_{l+1}^m w / 2^{m/2}] \cdot [(t - t_l^m) / 2^{m/2}]$ , the variance of  $w(m)_{t_l^m, t}^1$  is dominated by  $t - t_l^m$ . Then, in the same way as above, we see that  $\|A_5\|_{L^q} + \|A_6\|_{L^q} \leq C_3 \|g\|_\alpha (t - s)^{\alpha+1/2}$ .

Finally, we estimate  $A_2$ . We may assume  $k < l$ , here. We can easily see that

$$A_2 = \sum_{j=k+1}^l [g \cdot - g_{t_k^m}]_j^m (\Delta_j^m w) = \int_{t_k^m}^{t_l^m} \sum_{j=k+1}^l [g \cdot - g_{t_k^m}]_j^m \mathbf{1}_{[t_{j-1}^m, t_j^m]}(u) dw_u.$$

Noting that  $|[g \cdot - g_{t_k^m}]_j^m| \leq \|g\|_\alpha (t_l^m - t_k^m)^\alpha$ , we see from Burkholder's inequality that

$$\mathbb{E}[|A_2|^q]^{\frac{1}{q}} \leq C_4 \left( \int_{t_k^m}^{t_l^m} \left| \sum_{j=k+1}^l [g \cdot - g_{t_k^m}]_j^m \mathbf{1}_{[t_{j-1}^m, t_j^m]}(u) \right|^2 du \right)^{\frac{1}{2}} \leq C_4 \|g\|_\alpha (t_l^m - t_k^m)^{\alpha+1/2},$$

which completes the proof of the lemma.  $\square$

**Lemma 4.6.** *For every  $q \in [1, \infty)$ ,  $\alpha \in (0, 1)$  and  $\kappa \in (0, \alpha)$ , there exists a sequence of positive numbers  $\{\varepsilon_m\}_{m \in \mathbb{N}}$  converging to 0 as  $m \rightarrow \infty$  such that*

$$\|I[g, w(m)]_{s,t} - I[g, w]_{s,t}\|_{L^q} \leq \varepsilon_m \|g\|_\alpha (t - s)^{\alpha+\frac{1}{2}-\kappa}$$

for all  $m \in \mathbb{N}$ ,  $(s, t) \in \Delta_T$  and  $g \in \mathcal{C}_0^\alpha(\mathbb{R})$ . Here,  $\{\varepsilon_m\}$  does not depend on  $(s, t)$  or  $g$ .

*Proof.* We assume  $T = 1$  again and use the same notation as in the proof of Lemma 4.5. We will estimate  $I[g, w(m)] - I[g, w]$  for given  $m \in \mathbb{N}$ . Below,  $C_i$  ( $1 \leq i \leq 4$ ) are certain positive constants which depends only on  $q, \alpha, \kappa$ .

First, consider the case that  $s$  and  $t$  belong to the same subinterval, that is, there exists  $k$  such that  $s, t \in [t_{k-1}^m, t_k^m]$ . Then, we see from (4.20) that

$$\begin{aligned} \|I[g, w(m)]_{s,t} - I[g, w]_{s,t}\|_{L^q} &\leq \|I[g, w(m)]_{s,t}\|_{L^q} + \|I[g, w]_{s,t}\|_{L^q} \\ &\leq C_1 \|g\|_\alpha (t - s)^{\alpha+\frac{1}{2}} \\ &\leq C_1 (2^{-m})^\kappa \|g\|_\alpha (t - s)^{\alpha+\frac{1}{2}-\kappa}. \end{aligned} \quad (4.22)$$

Next, consider the case  $s \in [t_{k-1}^m, t_k^m]$  and  $t \in [t_l^m, t_{l+1}^m]$  with  $k \leq l$ . By Chen's relation we have

$$\begin{aligned} I[g, w(m)]_{s,t} - I[g, w]_{s,t} &= \{I[g, w(m)]_{s, t_k^m} - I[g, w]_{s, t_k^m}\} \\ &\quad + \{I[g, w(m)]_{t_k^m, t_l^m} - I[g, w]_{t_k^m, t_l^m}\} \\ &\quad + \{I[g, w(m)]_{t_l^m, t} - I[g, w]_{t_l^m, t}\} \\ &\quad + g_{s, t_k^m}^1 \{w(m)_{t_k^m, t_l^m}^1 - w_{t_k^m, t_l^m}^1\} \end{aligned}$$

$$\begin{aligned}
& + g_{s,t_k^m}^1 \{w(m)_{t_l^m,t}^1 - w_{t_l^m,t}^1\} \\
& + g_{t_k^m,t_l^m}^1 \{w(m)_{t_l^m,t}^1 - w_{t_l^m,t}^1\} =: B_1 + \cdots + B_6. \tag{4.23}
\end{aligned}$$

Note that  $B_4 = 0$  and  $B_1$  and  $B_3$  were already estimated in (4.22). We can easily check that

$$\begin{aligned}
\|B_6\|_{L^q} & \leq \|g\|_\alpha (t_l^m - t_k^m)^\alpha \{\|w(m)_{t_l^m,t}^1\|_{L^q} + \|w_{t_l^m,t}^1\|_{L^q}\} \\
& \leq C_2 \|g\|_\alpha (t_l^m - t_k^m)^\alpha (t - t_l^m)^{\frac{1}{2}} \\
& \leq C_2 (2^{-m})^\kappa \|g\|_\alpha (t - s)^{\alpha + \frac{1}{2} - \kappa}.
\end{aligned}$$

Obviously,  $B_5$  satisfies the same estimate, too.

It remains to estimate  $B_2$  when  $k < l$ . First, we should note that

$$B_2 = \int_{t_k^m}^{t_l^m} \left[ \sum_{j=k+1}^l [g - g_{t_k^m}]_j^m \mathbf{1}_{[t_{j-1}^m, t_j^m]}(u) - (g_u - g_{t_k^m}) \right] dw_u.$$

The absolute value of the integrand is dominated by  $\|g\|_\alpha (2^{-m})^\alpha$ . By Burkholder's inequality, we have

$$\|B_2\|_{L^q} \leq C_3 \|g\|_\alpha (2^{-m})^\alpha (t_l^m - t_k^m)^{\frac{1}{2}} \leq C_3 (2^{-m})^\kappa \|g\|_\alpha (t - s)^{\alpha + \frac{1}{2} - \kappa}$$

since  $0 < \kappa < \alpha$  and  $t - s \geq 2^{-m}$ .

Hence, taking  $\varepsilon_m := C_4 (2^{-m})^\kappa$  for a suitable constant  $C_4 > 0$ , we finish the proof of the lemma.  $\square$

**Lemma 4.7.** *Let  $\alpha \in (0, 1)$  and  $\{g_m\}_{m \in \mathbb{N}} \subset \mathcal{C}_0^\alpha(\mathbb{R})$  be a sequence which converges to  $g \in \mathcal{C}_0^\alpha(\mathbb{R})$  as  $m \rightarrow \infty$  in the  $\alpha$ -Hölder norm. Then, for every  $q \in [1, \infty)$  and  $\kappa \in (0, \alpha)$ , there exists a sequence of positive numbers  $\{\tilde{\varepsilon}_m\}_{m \in \mathbb{N}}$  converging to 0 as  $m \rightarrow \infty$  such that*

$$\|I[g_m, w(m)]_{s,t} - I[g, w]_{s,t}\|_{L^q} \leq C' \tilde{\varepsilon}_m (t - s)^{\alpha + \frac{1}{2} - \kappa}, \quad (s, t) \in \Delta_T, \quad m \in \mathbb{N}.$$

Here, we set  $C' := \|g\|_\alpha \vee \sup_{m \geq 1} \|g_m\|_\alpha$  and  $\{\tilde{\varepsilon}_m\}$  does not depend on  $(s, t)$ ,  $g$ ,  $\{g_m\}$ .

*Proof.* We may assume  $T = 1$  as before. The left hand side of the desired estimate is dominated by

$$\|I[g_m - g, w(m)]_{s,t}\|_{L^q} + \|I[g, w(m)]_{s,t} - I[g, w]_{s,t}\|_{L^q}.$$

By Lemma 4.5, the first term is dominated by  $C \|g_m - g\|_\alpha (t - s)^{\alpha + 1/2}$ . By Lemma 4.6, the second term is dominated by  $\varepsilon_m \|g\|_\alpha (t - s)^{\alpha + \frac{1}{2} - \kappa}$ . By setting  $\tilde{\varepsilon}_m := 2C + \varepsilon_m$ , we complete the proof.  $\square$

## 4.4 Construction of driving rough path

Let  $\frac{1}{4} < \alpha_0 \leq \frac{1}{3}$  and  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space. Let  $w = (w_t)_{0 \leq t \leq T}$  be a standard  $e$ -dimensional Brownian motion and let  $B = \{(B_{s,t}^1, B_{s,t}^2, B_{s,t}^3)\}_{0 \leq s \leq t \leq T}$  be an  $G\Omega_\alpha(\mathbb{R}^d)$ -valued random variable (i.e., random RP) defined on  $(\Omega, \mathcal{F}, \mathbb{P})$  for every  $\alpha \in (1/4, \alpha_0)$ . We assume that  $w$  and  $B$  are independent. As for the integrability of  $B$ , Assumption **(A)** is assumed, that is,  $\|B\|_\alpha$  has moments of all orders. Let  $\{\mathcal{F}_t\}_{0 \leq t \leq T}$  be a filtration satisfying the usual condition as well as the following two conditions: (i)  $w$  is an  $\{\mathcal{F}_t\}$ -BM and (ii)  $t \mapsto (B_{0,t}^1, B_{0,t}^2, B_{0,t}^3)$  is  $\{\mathcal{F}_t\}$ -adapted. We set  $W = (W^1, W^2)$  as the Stratonovich-type Brownian RP, that is,

$$W_{s,t}^1 = w_t - w_s, \quad W_{s,t}^2 = \int_s^t (w_u - w_s) \otimes \circ dw_u, \quad (s, t) \in \Delta_T,$$

where  $\circ dw$  stands for the Stratonovich integral. It is well-known that  $W \in G\Omega_\gamma(\mathbb{R}^e)$  a.s. and that  $\|W\|_\gamma$  has moments of all orders for every  $\gamma \in (1/3, 1/2)$ .

We set

$$\begin{aligned} I[B, W]_{s,t} &:= \int_s^t B_{s,u}^1 \otimes d^I w_u, \\ I[W, B]_{s,t} &:= W_{s,t}^1 \otimes B_{s,t}^1 - \int_s^t (d^I w_u) \otimes B_{s,u}^1. \end{aligned}$$

for every  $(s, t) \in \Delta_T$ . Here,  $d^I w_u$  stands for the Itô integration. We can easily see that

$$\begin{aligned} I[B, W]_{s,t} &= I[B, W]_{s,u} + I[B, W]_{u,t} + B_{s,u}^1 \otimes W_{u,t}^1, \\ I[W, B]_{s,t} &= I[W, B]_{s,u} + I[W, B]_{u,t} + W_{s,u}^1 \otimes B_{u,t}^1 \end{aligned}$$

hold a.s. for each fixed  $0 \leq s \leq u \leq t \leq T$ .

**Lemma 4.8.** *Let the situation be as above.*

(1) *For every  $(s, t) \in \Delta_T$ , we have*

$$I[B, W]_{s,t} = \lim_{|\mathcal{P}| \searrow 0} \sum_{i=1}^N B_{s,t_{i-1}}^1 \otimes W_{t_{i-1},t_i}^1, \quad I[W, B]_{s,t} = \lim_{|\mathcal{P}| \searrow 0} \sum_{i=1}^N W_{s,t_{i-1}}^1 \otimes B_{t_{i-1},t_i}^1.$$

Here, the limits are in  $L^2(\mathbb{P})$  and  $\mathcal{P} = \{s = t_0 < t_1 < \dots < t_N = t\}$  is a partition of  $[s, t]$ .

(2) *For every  $\alpha \in (\frac{1}{4}, \alpha_0)$  and  $q \in [1, \infty)$ , there exists a positive constant  $C > 0$  independent of  $(s, t)$  such that*

$$\|I[B, W]_{s,t}\|_{L^q} \vee \|I[W, B]_{s,t}\|_{L^q} \leq C(t-s)^{\alpha+\frac{1}{2}}, \quad (s, t) \in \Delta_T.$$

*Proof.* We can take any  $k \in \llbracket 1, d \rrbracket$  and  $l \in \llbracket 1, e \rrbracket$  and compute each  $(k, l)$ -component. Hence, it is enough to prove the lemma for  $d = e = 1$ .

We prove the first assertion. The left one is almost obvious from the definition of Itô integral. For the right one, note that

$$\sum_{i=1}^N B_{s,t_{i-1}}^1 W_{t_{i-1},t_i}^1 + \sum_{i=1}^N W_{s,t_{i-1}}^1 B_{t_{i-1},t_i}^1 = W_{s,t}^1 B_{s,t}^1 - \sum_{i=1}^N W_{t_{i-1},t_i}^1 B_{t_{i-1},t_i}^1.$$

By the independence, it holds for  $i < j$  that

$$\mathbb{E}[W_{t_{i-1},t_i}^1 B_{t_{i-1},t_i}^1 W_{t_{j-1},t_j}^1 B_{t_{j-1},t_j}^1] = \mathbb{E}[B_{t_{i-1},t_i}^1 B_{t_{j-1},t_j}^1] \mathbb{E}[W_{t_{i-1},t_i}^1] \mathbb{E}[W_{t_{j-1},t_j}^1] = 0.$$

This implies that

$$\begin{aligned} \mathbb{E} \left[ \left| \sum_{i=1}^N W_{t_{i-1},t_i}^1 B_{t_{i-1},t_i}^1 \right|^2 \right] &= \sum_{i=1}^N \mathbb{E}[(W_{t_{i-1},t_i}^1 B_{t_{i-1},t_i}^1)^2] \\ &\leq \sum_{i=1}^N \mathbb{E}[(W_{t_{i-1},t_i}^1)^4]^{1/2} \mathbb{E}[(B_{t_{i-1},t_i}^1)^4]^{1/2} \\ &\leq c_1 \sum_{i=1}^N (t_i - t_{i-1})^{1+2\alpha} \leq c_2 |\mathcal{P}|^{2\alpha}. \end{aligned}$$

Here, we used Assumption **(A)** and  $c_1$  and  $c_2$  are certain positive constants. The right hand side tends to zero as  $|\mathcal{P}| \searrow 0$ . Thus, we have shown (1).

From Burkholder's inequality and Assumption **(A)**, the second assertion immediately follows.  $\square$

In what follows, we assume  $\alpha \in (\frac{1}{4}, \alpha_0)$  and  $\gamma \in (\frac{1}{3}, \frac{1}{2})$  with  $2\alpha + \gamma > 1$ . By Proposition 4.3 and Lemma 4.8 (2), a continuous modification of  $I[B, W]$  and hence that of  $I[W, B]$  exist, which will be denoted by the same symbols. Moreover,  $\|I[B, W]\|_{\alpha+\gamma}$  and  $\|I[W, B]\|_{\alpha+\gamma}$  have moments of all orders. Thus, we have seen that  $(\Xi^1, \Xi^2, B^3) \in \hat{\Omega}_{\alpha, \gamma}(\mathcal{V})$ , a.s.

**Definition 4.9.** We write

$$\Xi_{s,t}^1 := \begin{pmatrix} B_{s,t}^1 \\ W_{s,t}^1 \end{pmatrix}, \quad \Xi_{s,t}^2 := \begin{pmatrix} B_{s,t}^2 & I[B, W]_{s,t} \\ I[W, B]_{s,t} & W_{s,t}^2 \end{pmatrix}$$

and set  $(\Xi^1, \Xi^2, \Xi^3) := \mathbf{Ext}(\Xi^1, \Xi^2, B^3)$ . (We also write  $\Xi = (\Xi^1, \Xi^2, \Xi^3)$  for simplicity.) As we will see below,  $(\Xi^1, \Xi^2, B^3) \in G\hat{\Omega}_{\alpha, \gamma}(\mathcal{V})$  a.s. Therefore, due to Proposition 4.2,  $\Xi = (\Xi^1, \Xi^2, \Xi^3) \in G\Omega_\alpha(\mathcal{V})$  a.s.

This random RP  $\Xi$  is the driving RP of our slow-fast system of RDEs. As we have seen in Proposition 4.1,  $\|\Xi^{3,[ijk]}\|_\delta$ , where  $\delta := (6 - i - j - k)\alpha + (i + j + k - 3)\gamma$ , have moments of all orders.

We write the components of  $\Xi^3$  as follows:  $\Xi^{3,[111]} = B^3$  and  $\Xi^{3,[222]} = W^3$ . For  $(i, j, k) \neq (1, 1, 1), (2, 2, 2)$ , we write  $\Xi^{3,[121]} = I[B, W, B]$ ,  $\Xi^{3,[221]} = I[W, W, B]$ , etc. Except when  $(i, j, k) = (1, 1, 1)$ , the Hölder regularity of  $\Xi^{3,[ijk]}$  is larger than 1. Hence, although these seven components are involved in the Riemann-type sum for a RP integral along  $\Xi$ , they actually make no contribution to the RP integral.

It remains to show that the random anisotropic RP  $(\Xi^1, \Xi^2, B^3)$  is geometric.

**Lemma 4.10.** *Let the notation be as above. Then,  $(\Xi^1, \Xi^2, B^3) \in G\hat{\Omega}_{\alpha, \gamma}(\mathcal{V})$ , a.s.*

*Proof.* Clearly,  $(\Xi^1, \Xi^2, B^3)$  is a functional of  $B = (B^1, B^2, B^3)$  and  $w$ . Since  $B$  and  $w$  are independent, we may write  $\mathbb{P} = \mathbb{P}^B \times \mathbb{P}^W$  and  $\mathbb{E} = \mathbb{E}^B \times \mathbb{E}^W$ . By Fubini's theorem, it is enough to show that, for every fixed realization of  $B$ ,  $(\Xi^1, \Xi^2, B^3) \in G\hat{\Omega}_{\alpha, \gamma}(\mathcal{V})$ ,  $\mathbb{P}^W$ -a.s. Hence, we will assume below that  $B$  is an arbitrary (non-random) element of  $G\Omega_\alpha(\mathcal{V}_1)$ . For simplicity, we assume  $T = 1$  again.

Let  $\{b(m)\}_{m \in \mathbb{N}} \subset \mathcal{C}_0^1(\mathcal{V}_1)$  be a sequence such that  $\lim_{m \rightarrow \infty} S_3(b(m)) = B$  in  $G\Omega_\alpha(\mathcal{V}_1)$ . For the  $m$ th dyadic approximation  $w(m)$ , we write  $(W(m)^1, W(m)^2)$  for  $S_3(w(m))$ . It is well-known that for every  $\gamma \in (1/3, 1/2)$  and  $q \in [1, \infty)$ , there are constants  $c_1, c_2 > 0$  and  $r \in (0, 1)$  such that

$$\mathbb{E}^W[\|W^i\|_{i\gamma}^{q/i}] \vee \mathbb{E}^W[\|W(m)^i\|_{i\gamma}^{q/i}] \leq c_1, \quad \mathbb{E}^W[\|W^i - W(m)^i\|_{i\gamma}^{q/i}] \leq c_2 r^m$$

holds for all  $m \in \mathbb{N}$  and  $i = 1, 2$ . Here,  $c_1, c_2, r$  are independent of  $m$ .

By Lemma 4.5, 4.7 and Proposition 4.4, we have

$$\lim_{m \rightarrow \infty} \mathbb{E}^W[\|I[b(m), w(m)] - I[b, w]\|_{\alpha+\gamma-2\kappa}] = 0$$

for every sufficiently small  $\kappa > 0$ . From this we can easily see that

$$\lim_{m \rightarrow \infty} \mathbb{E}^W[\|I[w(m), b(m)] - I[w, b]\|_{\alpha+\gamma-2\kappa}] = 0.$$

Hence, a subsequence of  $\{\hat{S}(b(m), w(m))\}_{m \in \mathbb{N}}$  converges to  $(\Xi^1, \Xi^2, B^3)$  in  $\hat{\Omega}_{\alpha-\kappa, \gamma-\kappa}(\mathcal{V})$  and hence  $(\Xi^1, \Xi^2, B^3) \in G\hat{\Omega}_{\alpha-\kappa, \gamma-\kappa}(\mathcal{V})$ ,  $\mathbb{P}^W$ -a.s. Noting that  $\alpha - \kappa$  and  $\gamma - \kappa$  can be arbitrarily close to  $\alpha_0$  and  $1/2$ , respectively, we complete the proof.  $\square$

In our construction of the random RP  $\Xi$ , Itô integration is used. So, it is not a priori obvious whether  $\Xi$  can be obtained via the piecewise linear approximation. At least, in the case of fractional Brownian motion with Hurst parameter  $H \in (1/4, 1/3]$ , we can prove it.

**Remark 4.11.** Let  $b^H = (b_t^H)_{t \in [0, T]}$  and  $w = (w_t)_{t \in [0, T]}$  be a  $d$ -dimensional fractional Brownian motion with Hurst parameter  $H \in (1/4, 1/3]$  and an  $e$ -dimensional Brownian motion, respectively, which are assumed to be independent. Their  $m$ th dyadic piecewise linear approximation ( $m \in \mathbb{N}$ ) are denoted by  $b^H(m)$  and  $w(m)$ , respectively.

According to [4, Theorem 15.42 and Proposition 15.5], the following limits of three sequences of random RPs exist a.s and in  $L^p$  ( $1 \leq p < \infty$ ):

- (i)  $\lim_{m \rightarrow \infty} S_3((b^H(m), w(m)))$  in  $G\Omega_\alpha(\mathcal{V}_1 \oplus \mathcal{V}_2)$  with  $\alpha \in (1/4, H)$ .
- (ii)  $\lim_{m \rightarrow \infty} S_3(b^H(m))$  in  $G\Omega_\alpha(\mathcal{V}_1)$  with  $\alpha \in (1/4, H)$ .
- (iii)  $\lim_{m \rightarrow \infty} S_2(w(m))$  in  $G\Omega_\gamma(\mathcal{V}_2)$  with  $\gamma \in (1/3, 1/2)$ .

(Concerning the lift of Gaussian processes of this kind, a recent work [5] should also be referred to.)

We call  $B^H = \lim_{m \rightarrow \infty} b^H(m)$  the fractional Brownian RP, i.e. a canonical lift of  $b^H$ . In this case we may take  $\alpha_0 = H$ . We can show that the mixed RP  $\Xi$  as in Definition 4.9 (with  $B$  being replaced by  $B^H$ ) coincides with  $\lim_{m \rightarrow \infty} S_3((b^H(m), w(m)))$  as expected. We will check this in the next paragraph.

Obviously from (ii) and (iii) above, 5 (out of 7) components of  $\hat{S}((b^H(m), w(m))) \in G\hat{\Omega}_{\alpha, \gamma}(\mathcal{V})$  converge a.s. In essentially the same way as in the proof of Lemma 4.10,

$$\lim_{m \rightarrow \infty} \mathbb{E}^W [\|I[b^{H,j}(m), w^k(m)] - I[b^{H,j}, w^k]\|_{\alpha+\gamma}] = 0$$

and therefore

$$\lim_{m \rightarrow \infty} \mathbb{E}^W [\|I[w^k(m), b^{H,j}(m)] - I[w^k, b^{H,j}]\|_{\alpha+\gamma}] = 0$$

hold for almost all fixed  $b^H$ , where the superscripts  $j$  ( $1 \leq j \leq d$ ) and  $k$  ( $1 \leq k \leq e$ ) stands for the coordinates of  $\mathcal{V}_1 = \mathbb{R}^d$  and  $\mathcal{V}_2 = \mathbb{R}^e$ , respectively. If we take a subsequence which may depend on  $b^H$ , we have  $\lim_{m \rightarrow \infty} \hat{S}((b^H(m), w(m))) = (\Xi^1, \Xi^2, (B^H)^3)$  for almost all  $w$ . Since  $\mathbf{Ext} \circ \hat{S} = S_3$  and  $\mathbf{Ext}$  is continuous, we have

$$(\Xi^1, \Xi^2, \Xi^3) = \mathbf{Ext}((\Xi^1, \Xi^2, (B^H)^3)) = \lim_{m \rightarrow \infty} S_3((b^H(m), w(m))) \quad \text{for a.a. } w.$$

Note that the right hand side above converges even if we do not take subsequence. Hence, we have  $(\Xi^1, \Xi^2, \Xi^3) = \lim_{m \rightarrow \infty} S_3((b^H(m), w(m)))$  for almost all  $(b^H, w)$  by Fubini's theorem.

## 5 Slow-fast system of rough differential equations

In this section we define the slow-fast system (2.1) of RDEs precisely and study its deterministic and probabilistic aspects. In this and the next sections, we always assume that  $\sigma$  and  $h$  are of  $C_b^4$  and  $f$  and  $g$  are locally Lipschitz continuous. The regularity parameters satisfy that  $\frac{1}{4} < \beta < \alpha < \alpha_0 \leq \frac{1}{3}$  and  $\frac{1}{3} < \gamma < \frac{1}{2}$  with  $2\alpha + \gamma > 1$ . These assumptions guarantee that our slow-fast system of RDEs has a unique time-local solution up to the explosion time. As before, we write  $\mathcal{V}_1 = \mathbb{R}^d$ ,  $\mathcal{V}_2 = \mathbb{R}^e$  and  $\mathcal{V} = \mathcal{V}_1 \oplus \mathcal{V}_2$ .

### 5.1 Deterministic aspects

First, we discuss deterministic aspects of our slow-fast system of RDEs. As in Definition 4.9, let  $(\Xi^1, \Xi^2, B^3) \in G\hat{\Omega}_{\alpha, \gamma}(\mathcal{V})$  and write  $\Xi = \mathbf{Ext}(\Xi^1, \Xi^2, B^3)$ . We will often write

$\Xi := (\Xi^1, \Xi^2, \Xi^3) \in G\Omega_\alpha(\mathcal{V})$ . This is the driving RP of our slow-fast system. For a CP  $(Z, Z^\dagger, Z^{\dagger\dagger}) \in \mathcal{Q}_\Xi^\beta(\mathbb{R}^{m+n})$  with respect to  $\Xi \in G\Omega_\alpha(\mathcal{V})$ , we often write  $Z = (X, Y)$ .

Respecting the direct sum decomposition  $\mathbb{R}^{m+n} = \mathbb{R}^m \oplus \mathbb{R}^n$ , a generic element of  $\mathbb{R}^{m+n}$  is denoted by  $z = (x, y)$ . The (partial) gradient operators with respect to  $z$ ,  $x$  and  $y$  are denoted by  $\nabla_z$ ,  $\nabla_x$  and  $\nabla_y$ , respectively. Hence,  $\nabla_z = (\nabla_x, \nabla_y)$  at least formally when it acts on nice functions on  $\mathbb{R}^{m+n}$ . We will write the canonical projections as  $\rho_1\langle z \rangle = x$  and  $\rho_2\langle z \rangle = y$ . We set

$$F_\varepsilon(x, y) = \begin{pmatrix} f(x, y) \\ \varepsilon^{-1}g(x, y) \end{pmatrix}, \quad \Sigma_\varepsilon(x, y) = \begin{pmatrix} \sigma(x) & O \\ O & \varepsilon^{-1/2}h(x, y) \end{pmatrix}.$$

Then,  $F_\varepsilon: \mathbb{R}^{m+n} \rightarrow \mathbb{R}^{m+n}$  and  $\Sigma_\varepsilon: \mathbb{R}^{m+n} \rightarrow L(\mathcal{V}_1 \oplus \mathcal{V}_2, \mathbb{R}^{m+n})$ . In this subsection,  $\varepsilon \in (0, 1]$  is arbitrary but fixed.

The precise definition of the slow-fast system (2.1) of RDEs (in the deterministic sense) is given as follows:

$$\begin{aligned} Z_t^\varepsilon &= z_0 + \int_0^t F_\varepsilon(Z_s^\varepsilon) ds + \int_0^t \Sigma_\varepsilon(Z_s^\varepsilon) d\Xi_s, \\ (Z^\varepsilon)_t^\dagger &= \Sigma_\varepsilon(Z_t^\varepsilon), \quad (Z^\varepsilon)_t^{\dagger\dagger} = (\nabla_z \Sigma_\varepsilon \cdot \Sigma_\varepsilon)(Z_t^\varepsilon), \quad t \in [0, T]. \end{aligned} \quad (5.1)$$

(Note that  $(Z^\varepsilon)^\dagger$  and  $(Z^\varepsilon)^{\dagger\dagger}$  take value in  $L(\mathcal{V}, \mathbb{R}^{m+n})$  and  $L(\mathcal{V} \otimes \mathcal{V}, \mathbb{R}^{m+n})$ , respectively.) We consider this RDE in the  $\beta$ -Hölder topology.

Let  $(Z^\varepsilon, \Sigma_\varepsilon(Z^\varepsilon), (\nabla_z \Sigma_\varepsilon \cdot \Sigma_\varepsilon)(Z^\varepsilon))$  be a (necessarily unique) solution of RDE (5.1) on a certain time interval  $[0, \tau]$ , where  $\tau \in (0, T]$ . Then, the summand of the modifies Riemann sum that approximates the RP integral on the right hand side of (5.1) is given by

$$K_{s,t} := \Sigma_\varepsilon(Z_s^\varepsilon) \Xi_{s,t}^1 + \{\Sigma_\varepsilon(Z^\varepsilon)\}_s^\dagger \Xi_{s,t}^2 + \{\Sigma_\varepsilon(Z^\varepsilon)\}_s^{\dagger\dagger} \Xi_{s,t}^3, \quad (s, t) \in \Delta_{[0, \tau]}. \quad (5.2)$$

Here,

$$\begin{aligned} \{\Sigma_\varepsilon(Z^\varepsilon)\}_s^\dagger &= (\nabla_z \Sigma_\varepsilon \cdot \Sigma_\varepsilon)(Z_s^\varepsilon) = \nabla_z \Sigma_\varepsilon(Z_s^\varepsilon) \langle \Sigma_\varepsilon(Z_s^\varepsilon) \bullet, \star \rangle \in L(\mathcal{V}^{\otimes 2}, \mathbb{R}^{m+n}), \\ \{\Sigma_\varepsilon(Z^\varepsilon)\}_s^{\dagger\dagger} &= \nabla_z \Sigma_\varepsilon(Z_s^\varepsilon) \langle (\nabla_z \Sigma_\varepsilon \cdot \Sigma_\varepsilon)(Z_s^\varepsilon) \langle \bullet, \star \rangle, * \rangle \\ &\quad + \nabla_z^2 \Sigma_\varepsilon(Z_s^\varepsilon) \langle \Sigma_\varepsilon(Z_s^\varepsilon) \bullet, \Sigma_\varepsilon(Z_s^\varepsilon) \star, * \rangle \in L(\mathcal{V}^{\otimes 3}, \mathbb{R}^{m+n}). \end{aligned}$$

Here, we used the third item of Example 3.1, again.

**Remark 5.1.** For the rest of this section, we use the following notation. Let  $\tau \in (0, T]$  and let  $\mathcal{X}$  be a Euclidean space. If a continuous map  $\eta: \Delta_{[0, \tau]} \rightarrow \mathcal{X}$  belongs to  $\mathcal{C}_{(2)}^\delta([0, \tau], \mathcal{X})$  for  $\delta > 0$ , we simply write  $O((t-s)^\delta)$  for  $\eta_{s,t}$ . It should be noted:

- When we use this “big  $O$ ” symbol, we do not assume that  $t-s$  is small.
- When  $\eta$  depends on a driving RP  $\Xi$  or  $(\Xi^1, \Xi^2, B^3)$ ,  $\|\eta\|_{\delta, [0, \tau]}$  may depend on the RP (and the parameter  $\varepsilon$ ). In other words, this symbol only means RP-wise estimates.

**Remark 5.2.** An element of a direct sum space is denoted by both a “column vector” and a “row vector.” These are not precisely distinguished.

**Lemma 5.3.** *Let  $0 < \tau \leq T$  and let the situation be as above. Suppose that*

$$(Z^\varepsilon, \Sigma_\varepsilon(Z^\varepsilon), (\nabla_z \Sigma_\varepsilon \cdot \Sigma_\varepsilon)(Z^\varepsilon)) \in \mathcal{Q}_\Xi^\beta([0, \tau], \mathbb{R}^{m+n}),$$

*is a unique solution of RDE (5.1) on  $[0, \tau]$  and write  $Z^\varepsilon = (X^\varepsilon, Y^\varepsilon)$ .*

*Then,  $(X^\varepsilon, \sigma(X^\varepsilon), (\nabla_x \sigma_\varepsilon \cdot \sigma_\varepsilon)(X^\varepsilon))$  belongs to  $\mathcal{Q}_B^\beta([0, \tau], \mathbb{R}^m)$  and is a unique local solution of the following RDE driven by  $B = (B^1, B^2, B^3)$  on  $[0, \tau]$ :*

$$\begin{aligned} X_t^\varepsilon &= x_0 + \int_0^t f(X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t \sigma(X_s^\varepsilon) dB_s, \\ (X^\varepsilon)_t^\dagger &= \sigma(X_t^\varepsilon), \quad (X^\varepsilon)_t^{\dagger\dagger} = (\nabla_x \sigma \cdot \sigma)(X_t^\varepsilon), \quad t \in [0, \tau]. \end{aligned} \tag{5.3}$$

Recall that an RDE of this type was introduced in (3.17) and discussed in Subsection 3.3.

*Proof.* By Proposition 3.2, we have

$$X_t^\varepsilon - X_s^\varepsilon = O((t-s)^1) + \rho_1 \langle K_{s,t} \rangle, \quad (s, t) \in \Delta_{[0, \tau]} \tag{5.4}$$

since  $4\beta > 1$ .

It is clear that

$$\rho_1 \langle \Sigma_\varepsilon(Z_s^\varepsilon) \Xi_{s,t}^1 \rangle = \sigma(X_s^\varepsilon) B_{s,t}^1. \tag{5.5}$$

Note that the  $\mathbb{R}^m$ -component of  $\Sigma_\varepsilon(z)$  equals  $\sigma(\rho_1 \langle z \rangle) \circ \pi_1 = \sigma(x) \circ \pi_1$ . In particular,  $\nabla_y \rho_1 \langle \Sigma_\varepsilon(z) \rangle$  vanishes. By standard calculation for block matrices, we see that

$$\begin{aligned} \rho_1 \langle \nabla_z \Sigma_\varepsilon(Z_s^\varepsilon) \langle \Sigma_\varepsilon(Z_s^\varepsilon) \bullet, \star \rangle \rangle &= \nabla_z [\rho_1 \langle \Sigma_\varepsilon(Z_s^\varepsilon) \rangle] \langle \Sigma_\varepsilon(Z_s^\varepsilon) \bullet, \star \rangle \\ &= (\nabla_x \sigma)(X_s^\varepsilon) \langle \rho_1 \langle \Sigma_\varepsilon(Z_s^\varepsilon) \bullet \rangle, \pi_1 \star \rangle \\ &= (\nabla_x \sigma)(X_s^\varepsilon) \langle \sigma(X_s^\varepsilon) \pi_1 \bullet, \pi_1 \star \rangle \end{aligned}$$

and therefore

$$\rho_1 \langle \{\Sigma_\varepsilon(Z^\varepsilon)\}_s^\dagger \Xi_{s,t}^2 \rangle = (\nabla_x \sigma)(X_s^\varepsilon) \langle \sigma(X_s^\varepsilon) \bullet', \star' \rangle|_{(\bullet', \star')=B_{s,t}^2} = (\nabla_x \sigma \cdot \sigma)(X_s^\varepsilon) \langle B_{s,t}^2 \rangle. \tag{5.6}$$

Set  $(X^\varepsilon)^\dagger = \sigma(X^\varepsilon)$  and  $(X^\varepsilon)^{\dagger\dagger} = (\nabla_x \sigma \cdot \sigma)(X^\varepsilon)$ . Then, we can easily see from (5.4)–(5.6) that  $X_t^\varepsilon - X_s^\varepsilon = (X_s^\varepsilon)_s^\dagger B_{s,t}^1 + (X_s^\varepsilon)_s^{\dagger\dagger} B_{s,t}^2 + O((t-s)^{3\beta})$ . We can also easily check that  $(X^\varepsilon)_t^\dagger - (X^\varepsilon)_s^\dagger = (X_s^\varepsilon)_s^{\dagger\dagger} \langle B_{s,t}^1, \star \rangle + O((t-s)^{2\beta})$ . Hence,  $(X^\varepsilon, (X^\varepsilon)^\dagger, (X^\varepsilon)^{\dagger\dagger})$  is a CP with respect to  $B = (B^1, B^2, B^3)$ . It should be noted that  $(X^\varepsilon)^{\dagger\dagger} = \{\sigma(X^\varepsilon)\}^\dagger$ .

By cumbersome but similar calculations for block matrices as above, we also have

$$\begin{aligned} \rho_1 \langle \{\Sigma_\varepsilon(Z^\varepsilon)\}_s^{\dagger\dagger} \Xi_{s,t}^3 \rangle &= \nabla_x \sigma(X_s^\varepsilon) \langle (\nabla_x \sigma \cdot \sigma)(X_s^\varepsilon) \langle \bullet', \star' \rangle, \star' \rangle|_{(\bullet', \star', \star')=B_{s,t}^3} \\ &\quad + \nabla_x^2 \sigma(X_s^\varepsilon) \langle \sigma(X_s^\varepsilon) \bullet', \sigma(X_s^\varepsilon) \star', \star' \rangle|_{(\bullet', \star', \star')=B_{s,t}^3} \end{aligned}$$

$$= \{\sigma(X^\varepsilon)\}_s^{\dagger\dagger} \langle B_{s,t}^3 \rangle. \quad (5.7)$$

Hence, we have

$$\rho_1 \langle K_{s,t} \rangle = \sigma(X_s^\varepsilon) B_{s,t}^1 + \{\sigma(X^\varepsilon)\}_s^\dagger \langle B_{s,t}^2 \rangle + \{\sigma(X^\varepsilon)\}_s^{\dagger\dagger} \langle B_{s,t}^3 \rangle$$

and therefore

$$\rho_1 \left\langle \int_s^t \Sigma_\varepsilon(Z_u^\varepsilon) d\Xi_u \right\rangle = \int_s^t \sigma(X_u^\varepsilon) dB_u.$$

This completes the proof of the lemma.  $\square$

As for the slow component of  $K_{s,t}$ , we can easily see the following lemma, in which  $\nabla_x h(x, y) \cdot \sigma(x)$  is a shorthand for the linear map from  $\mathcal{V}_1 \otimes \mathcal{V}_2$  to  $\mathbb{R}^n$  defined by  $\xi \otimes \eta \mapsto \nabla_x h(x, y) \langle \sigma(x) \xi, \eta \rangle$ . Note that  $2\alpha + \gamma > 1$  and that the third term on the right hand side of (5.8) is (formally) the same as the Itô-Stratonovich correction term. (In the probabilistic part of this paper,  $W^2$  and  $\bar{W}^2$  will be the second level of the Stratonovich-type and the Itô-type Brownian RP, respectively.)

**Lemma 5.4.** *Let the assumptions be the same as in Lemma 5.3 above. Then, we have*

$$\begin{aligned} \rho_2 \langle K_{s,t} \rangle &= \varepsilon^{-1/2} h(X_s^\varepsilon, Y_s^\varepsilon) W_{s,t}^1 + \varepsilon^{-1/2} \nabla_x h(X_s^\varepsilon, Y_s^\varepsilon) \cdot \sigma(X_s^\varepsilon) \langle I[B, W]_{s,t} \rangle \\ &\quad + \varepsilon^{-1} (\nabla_y h \cdot h)(X_s^\varepsilon, Y_s^\varepsilon) \langle W_{s,t}^2 \rangle + O((t-s)^{2\alpha+\gamma}) \\ &= \varepsilon^{-1/2} h(X_s^\varepsilon, Y_s^\varepsilon) W_{s,t}^1 + \varepsilon^{-1/2} \nabla_x h(X_s^\varepsilon, Y_s^\varepsilon) \cdot \sigma(X_s^\varepsilon) \langle I[B, W]_{s,t} \rangle \\ &\quad + \varepsilon^{-1} (\nabla_y h \cdot h)(X_s^\varepsilon, Y_s^\varepsilon) \langle \bar{W}_{s,t}^2 \rangle \\ &\quad + \frac{1}{2} \varepsilon^{-1} \text{Trace}[(\nabla_y h \cdot h)(X_s^\varepsilon, Y_s^\varepsilon) \langle \bullet, \star \rangle] (t-s) + O((t-s)^{2\alpha+\gamma}) \end{aligned} \quad (5.8)$$

for all  $(s, t) \in \Delta_{[0, \tau]}$ . Here, we set  $\bar{W}_{s,t}^2 := W_{s,t}^2 - \frac{t-s}{2} \sum_{k=1}^e \mathbf{e}_k \otimes \mathbf{e}_k$  for the canonical orthonormal basis  $\{\mathbf{e}_k\}_{k=1}^e$  of  $\mathcal{V}_2 = \mathbb{R}^e$ .

*Proof.* We only check the first equality since the second one is obvious. The components of  $\Xi^3$  involved in  $\rho_2 \langle K_{s,t} \rangle$  are  $I[B, B, W]_{s,t}$ ,  $I[B, W, W]_{s,t}$ ,  $I[W, B, W]_{s,t}$  and  $W_{s,t}^3$ , all of which are  $O((t-s)^{2\alpha+\gamma})$ . The rest is trivial.  $\square$

## 5.2 Probabilistic aspects

In what follows we work under **(A)** and assume that  $1/4 < \beta < \alpha < \alpha_0 (\leq 1/3)$ . For the rest of this section,  $\Xi$  is as in Definition 4.9. The precise meaning of the random RDE in our main theorem is RDE (5.1) driven by this  $\Xi$ .

We extend the time interval of the filtration  $\{\mathcal{F}_t\}$  by setting  $\mathcal{F}_t = \mathcal{F}_{t \wedge T}$  for  $t \geq 0$ . Denote by  $\hat{\mathbb{R}}^{m+n} := \mathbb{R}^{m+n} \cup \{\infty\}$  the one-point compactification of  $\mathbb{R}^{m+n}$ . If a global solution  $(Z_t^\varepsilon)_{t \in [0, T]}$  exists, then we set  $Z_t^\varepsilon = Z_{t \wedge T}^\varepsilon$  for  $t \geq 0$ . Otherwise, denote by  $(Z_t^\varepsilon)_{t \in [0, u^\varepsilon]}$ ,  $0 < u^\varepsilon \leq T$ , be a maximal local solution and set  $Z_t^\varepsilon = \infty$  for  $t \in [u^\varepsilon, \infty)$ . Either way,  $(Z_t^\varepsilon)$  is constant in  $t$  on  $[T, \infty)$ , a.s.

Define  $\tau_N^\varepsilon = \inf\{t \geq 0 \mid |Z_t^\varepsilon| \geq N\}$  for each  $N \in \mathbb{N}$  and  $\tau_\infty^\varepsilon = \lim_{N \rightarrow \infty} \tau_N^\varepsilon$ . (As usual  $\inf \emptyset := \infty$ .) These are  $\{\mathcal{F}_t\}$ -stopping times. Then, the following are equivalent:

- A global solution  $(Z_t^\varepsilon)_{t \in [0, T]}$  of RDE (5.1) exists.
- $(Z_t^\varepsilon)_{t \in [0, T]}$  defined as above is bounded in  $\mathbb{R}^{m+n}$ .
- $\tau_N^\varepsilon = \infty$  for some  $N$ .
- $\tau_\infty^\varepsilon > T$ .

It should be noted that while a solution of RDE (5.1) moves in a bounded set, its trajectory is uniformly continuous in  $t$  (because its Hölder norm is bounded). Hence, if  $(Z_t^\varepsilon)_{t \in [0, s]}$ ,  $0 < s \leq T$ , is bounded, then  $(Z_t^\varepsilon)_{t \in [0, s]}$  solves RDE (5.1).

On the other hand, if no global solution exists, then we have  $u^\varepsilon = \tau_\infty^\varepsilon \in (0, T]$  and  $\limsup_{t \nearrow \tau_\infty^\varepsilon} |Z_t| = \infty$ . Moreover,  $\lim_{t \nearrow \tau_\infty^\varepsilon} |Z_t| = \infty$  because of the uniform continuity mentioned above. Therefore,  $(Z_t^\varepsilon)_{t \geq 0}$  is a continuous process that takes values in  $\hat{\mathbb{R}}^{m+n}$ .

**Proposition 5.5.** *Let the notation be as above and assume (A). Then, for every  $\varepsilon \in (0, 1]$ ,  $Y^\varepsilon$  satisfies the following Itô SDE up to the explosion time of  $Z^\varepsilon = (X^\varepsilon, Y^\varepsilon)$ :*

$$Y_t^\varepsilon = y_0 + \frac{1}{\varepsilon} \int_0^{t \wedge T} \tilde{g}(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^{t \wedge T} h(X_s^\varepsilon, Y_s^\varepsilon) d^I w_s, \quad 0 \leq t < \tau_\infty^\varepsilon.$$

Recall that  $\tilde{g}(x, y)$  was defined by (2.2), that is,

$$\tilde{g}(x, y) = g(x, y) + \frac{1}{2} \text{Trace}[(\nabla_y h \cdot h)(x, y) \langle \bullet, \star \rangle].$$

*Proof.* The proof is very similar to the corresponding one in the case  $\alpha_0 \in (1/3, 1/2]$  in [8, Proposition 4.7]. Hence, our argument here is a little bit sketchy.

Let  $\mathcal{P} = \{0 = t_0 < t_1 < \dots < t_K = t\}$  be a partition of  $[0, t]$  for  $0 < t \leq T$ . Recall that the summand of the Riemann sum for the RP integral in RDE (5.1) was calculated in the previous subsection.

First, we prove the lemma when  $h, \sigma$  are of  $C_b^4$  and  $f, g$  are bounded and globally Lipschitz continuous. In this case the solution never explodes, i.e.  $\tau_\infty^\varepsilon = \infty$ , a.s. It is easy to see that

$$\lim_{|\mathcal{P}| \searrow 0} \sum_{i=1}^K h(X_{t_{i-1}}^\varepsilon, Y_{t_{i-1}}^\varepsilon) W_{t_{i-1}, t_i}^1 = \int_0^t h(X_s^\varepsilon, Y_s^\varepsilon) d^I w_s \quad \text{in } L^2(\mathbb{P})$$

and

$$\begin{aligned} & \lim_{|\mathcal{P}| \searrow 0} \sum_{i=1}^K \text{Trace}[(\nabla_y h \cdot h)(X_{t_{i-1}}^\varepsilon, Y_{t_{i-1}}^\varepsilon) \langle \bullet, \star \rangle] (t_i - t_{i-1}) \\ &= \int_0^t \text{Trace}[(\nabla_y h \cdot h)(X_s^\varepsilon, Y_s^\varepsilon) \langle \bullet, \star \rangle] ds, \quad \text{a.s.} \end{aligned}$$

Note that  $\frac{1}{2} \text{Trace}[(\nabla_y h \cdot h)(x, y) \langle \bullet, \star \rangle] = \tilde{g}(x, y) - g(x, y)$ .

If  $A_{s,t} = O((t-s)^{2\alpha+\gamma})$ , one can easily see that  $\lim_{|\mathcal{P}| \searrow 0} \sum_{i=1}^K A_{t_{i-1}, t_i} = 0$ , a.s. since  $2\alpha + \gamma > 1$ . By the same argument as in [8], we can prove

$$\begin{aligned} \lim_{|\mathcal{P}| \searrow 0} \sum_{i=1}^K (\nabla_y h \cdot h)(X_{t_{i-1}}^\varepsilon, Y_{t_{i-1}}^\varepsilon) \langle \bar{W}_{t_{i-1}, t_i}^2 \rangle &= 0 \quad \text{in } L^2(\mathbb{P}), \\ \lim_{|\mathcal{P}| \searrow 0} \sum_{i=1}^K \nabla_y h(X_{t_{i-1}}^\varepsilon, Y_{t_{i-1}}^\varepsilon) \cdot \sigma(X_{t_{i-1}}^\varepsilon, Y_{t_{i-1}}^\varepsilon) \langle I[B, W]_{t_{i-1}, t_i} \rangle &= 0 \quad \text{in } L^2(\mathbb{P}). \end{aligned}$$

Note that  $\bar{W}^2$  is the second level of Brownian RP of Itô type. Combining these all, we have shown that

$$Y_t^\varepsilon = y_0 + \frac{1}{\varepsilon} \int_0^{t \wedge T} \tilde{g}(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^{t \wedge T} h(X_s^\varepsilon, Y_s^\varepsilon) d^I w_s, \quad t \geq 0 \quad (5.9)$$

holds a.s. in this case.

From here only the standing assumption is assumed on the coefficients  $h, \sigma, f, g$ . Take any sufficiently large  $N$ . Let  $\phi_N: \mathbb{R}^{m+n} \rightarrow [0, 1]$  be a smooth function with compact support such that  $\phi_N \equiv 1$  on the ball  $\{z \in \mathbb{R}^{m+n} \mid |z| \leq N\}$  and set  $\hat{h} := h\phi_N$ . Also,  $\hat{\sigma}, \hat{f}, \hat{g}$  are defined in the same way. We replace the coefficients of RDE (5.1) by these corresponding data with “hat” and denote a unique solution by  $\hat{Z}^\varepsilon = (\hat{X}^\varepsilon, \hat{Y}^\varepsilon)$ . Then, (5.9) holds with  $\hat{X}^\varepsilon, \hat{Y}^\varepsilon, \hat{h}, \hat{g}$  in place of  $X^\varepsilon, Y^\varepsilon, h, g$ . By the uniqueness of the RDE, it holds that  $\hat{Z}_{t \wedge \tau_N^\varepsilon}^\varepsilon = Z_{t \wedge \tau_N^\varepsilon}^\varepsilon$  for all  $0 \leq t \leq T$ . Therefore, we almost surely have

$$\begin{aligned} Y_{t \wedge \tau_N^\varepsilon \wedge T}^\varepsilon &= \hat{Y}_{t \wedge \tau_N^\varepsilon \wedge T}^\varepsilon \\ &= y_0 + \frac{1}{\varepsilon} \int_0^{t \wedge \tau_N^\varepsilon \wedge T} (\hat{g})^\sim(\hat{X}_s^\varepsilon, \hat{Y}_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^{t \wedge \tau_N^\varepsilon \wedge T} \hat{h}(\hat{X}_s^\varepsilon, \hat{Y}_s^\varepsilon) d^I w_s, \\ &= y_0 + \frac{1}{\varepsilon} \int_0^{t \wedge \tau_N^\varepsilon \wedge T} \tilde{g}(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^{t \wedge \tau_N^\varepsilon \wedge T} h(X_s^\varepsilon, Y_s^\varepsilon) d^I w_s, \quad t \geq 0. \end{aligned} \quad (5.10)$$

Since  $\tau_N^\varepsilon \nearrow \tau_\infty^\varepsilon$  as  $N \rightarrow \infty$  a.s. on the set  $\{\tau_\infty^\varepsilon \leq T\}$ , we finish the proof by letting  $N \rightarrow \infty$ .  $\square$

In the same way as in the author’s previous work [8], we can show non-explosion of the solution under the assumptions on the coefficients. Note that our assumptions on the coefficients are stronger than the corresponding ones in [8].

**Proposition 5.6.** *Assume (A) and (H1)–(H5). Then, the probability that  $Z^\varepsilon = (X^\varepsilon, Y^\varepsilon)$  explodes on  $[0, T]$  is zero. Moreover, we have*

$$\sup_{0 < \varepsilon \leq 1} \mathbb{E}[\|X^\varepsilon\|_{\beta, [0, T]}^p] < \infty, \quad 1 \leq p < \infty, \quad (5.11)$$

$$\sup_{0 < \varepsilon \leq 1} \sup_{0 \leq t \leq T} \mathbb{E}[|Y_t^\varepsilon|^2] < \infty. \quad (5.12)$$

*Proof.* Thanks to Proposition 3.3, Lemma 5.3 and Proposition 5.5, the same proof as in [8, Proposition 4.8] still works.  $\square$

Now that Proposition 5.6 has been obtained, our arguments in what follows are very similar to the level 2 case in the corresponding part of [8]. Therefore, in order to avoid repetition, our proofs will be sketchy from here.

Now, we introduce a new parameter  $\delta$  with  $0 < \varepsilon < \delta \leq 1$ . (In spirit,  $0 < \varepsilon \ll \delta \ll 1$ . Later, we will set  $\delta := \varepsilon^{1/(6\beta)} \log \varepsilon^{-1}$ .) We divide  $[0, T]$  into subintervals of equal length  $\delta$  (except perhaps the last subinterval). For  $s \geq 0$ , set  $s(\delta) := \lfloor s/\delta \rfloor \delta$ , which is the nearest breaking point preceding or equal to  $s$ .

We set

$$\hat{Y}_t^\varepsilon = y_0 + \frac{1}{\varepsilon} \int_0^t \tilde{g}(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) ds + \frac{1}{\sqrt{\varepsilon}} \int_0^t h(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) d^I w_s, \quad t \in [0, T]. \quad (5.13)$$

Note that  $\hat{Y}^\varepsilon$ 's dependence on  $\delta$  is suppressed in the notation. This approximation process satisfies the following two estimates. (The next two lemmas are basically the same as or easier than [8, Lemma 5.1 and Lemma 5.2].)

**Lemma 5.7.** *Under the same assumptions as in Proposition 5.6, we have the following: For every  $\delta$  and  $\varepsilon$  with  $0 < \varepsilon < \delta \leq 1$ , the above process  $\hat{Y}^\varepsilon$  does not explode and satisfies*

$$\sup_{0 < \varepsilon < \delta \leq 1} \sup_{0 \leq t \leq T} \mathbb{E}[|\hat{Y}_t^\varepsilon|^2] < \infty. \quad (5.14)$$

*Proof.* The proof is essentially the same as that of Proposition 5.6. (In fact, this one is easier because we already know  $X_{s(\delta)}^\varepsilon$  exists and satisfies the estimate (5.11).)  $\square$

**Lemma 5.8.** *Assume (A) and (H1)–(H6). Then, there exists a positive constant  $C$  independent of  $\delta$  such that*

$$\sup_{\varepsilon \in (0, \delta)} \sup_{0 \leq t \leq T} \mathbb{E}[|Y_t^\varepsilon - \hat{Y}_t^\varepsilon|^2] \leq C\delta^{2\beta}.$$

*Proof.* We omit the proof since it is essentially the same as in [8, Lemma 5.2].  $\square$

It is easy to see that, if we define

$$\begin{aligned} M_t &= \int_0^t \{f(X_s^\varepsilon, Y_s^\varepsilon) - f(X_{s(\delta)}^\varepsilon, Y_s^\varepsilon)\} ds + \int_0^t \{f(X_{s(\delta)}^\varepsilon, Y_s^\varepsilon) - f(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon)\} ds \\ &\quad + \int_0^t \{f(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{f}(X_{s(\delta)}^\varepsilon)\} ds + \int_0^t \{\bar{f}(X_{s(\delta)}^\varepsilon) - \bar{f}(X_s^\varepsilon)\} ds, \end{aligned} \quad (5.15)$$

then

$$X_t^\varepsilon - \bar{X}_t = \int_0^t f(X_s^\varepsilon, Y_s^\varepsilon) ds + \int_0^t \sigma(X_s^\varepsilon) dB_s - \int_0^t \bar{f}(\bar{X}_s) ds - \int_0^t \sigma(\bar{X}_s) dB_s \quad (5.16)$$

$$= M_t + \left( \int_0^t \{ \bar{f}(X_s^\varepsilon) - \bar{f}(\bar{X}_s) \} ds + \int_0^t \{ \sigma(X_s^\varepsilon) - \sigma(\bar{X}_s) \} dB_s \right)$$

holds as an equality of CPs with respect to  $B$ . We will later apply Proposition 3.5 to (5.16) after estimating  $\|M\|_{3\beta}$ .

**Lemma 5.9.** *Assume the same condition as in Lemma 5.8. Then, there exists a positive constant  $C$  independent of  $\varepsilon, \delta$  such that*

$$\begin{aligned} \mathbb{E} \left[ \left\| \int_0^\cdot \{ f(X_s^\varepsilon, Y_s^\varepsilon) - f(X_{s(\delta)}^\varepsilon, Y_s^\varepsilon) \} ds \right\|_1^2 + \left\| \int_0^\cdot \{ \bar{f}(X_{s(\delta)}^\varepsilon) - \bar{f}(X_s^\varepsilon) \} ds \right\|_1^2 \right. \\ \left. + \left\| \int_0^\cdot \{ f(X_{s(\delta)}^\varepsilon, Y_s^\varepsilon) - f(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) \} ds \right\|_1^2 \right] \leq C\delta^{2\beta} \end{aligned}$$

for all  $0 < \varepsilon < \delta \leq 1$ . Here,  $\|\cdot\|_1$  stands for the 1-Hölder (i.e. Lipschitz) norm.

*Proof.* Using the globally Lipschitz property of  $f$ , we can show this lemma in a straightforward way. The proof is easy and essentially the same as in [8, Lemma 5.3].  $\square$

We can estimate the most difficult one among the four terms in the definition of  $M$  as follows.

**Lemma 5.10.** *Assume the same condition as in Lemma 5.8 and let  $0 < \gamma < 1$ . Then, there exists a positive constant  $C$  independent of  $\varepsilon, \delta$  such that*

$$\mathbb{E} \left[ \left\| \int_0^\cdot \{ f(X_{s(\delta)}^\varepsilon, \hat{Y}_s^\varepsilon) - \bar{f}(X_{s(\delta)}^\varepsilon) \} ds \right\|_\gamma^2 \right] \leq C(\delta^{2(1-\gamma)} + \delta^{-2\gamma}\varepsilon)$$

for all  $0 < \varepsilon < \delta \leq 1$ .

*Proof.* Essentially the same proof as in [8, Lemma 5.4] works in this case, too. However, it should be noted that the proof is not easy. Both a careful approximation on each subinterval and the Markov property of the frozen SDE must be used.  $\square$

Now we prove our main theorem.

*Proof of Theorem 2.1.* Applying Proposition 3.5 to (5.15) and (5.16), we obtain that

$$\|X^\varepsilon - \bar{X}\|_\beta \leq C \exp(C\|B\|_\alpha^\nu) \|M\|_{3\beta}$$

for certain positive constant  $C$  and  $\nu$  which are independent of  $\varepsilon, \delta$ . (Below,  $C$  and  $\nu$  may vary from line to line.) By Lemmas 5.9 and 5.10, we have

$$\mathbb{E}[\|M\|_{3\beta}^2] \leq C(\delta^{2\beta} + \delta^{2(1-3\beta)} + \delta^{-6\beta}\varepsilon).$$

Therefore, if we set  $\delta := \varepsilon^{1/(6\beta)} \log \varepsilon^{-1}$  for example, then  $\|M\|_{3\beta}$  converges to 0 in  $L^2$ -sense as  $\varepsilon \searrow 0$ . It immediately follows that  $\|X^\varepsilon - \bar{X}\|_\beta^p$  converges to 0 in probability as  $\varepsilon \searrow 0$  for every  $p \in [1, \infty)$ .

On the other hand, we see from Proposition 3.3 that

$$\sup_{0 < \varepsilon \leq 1} \mathbb{E}[\|X^\varepsilon - \bar{X}\|_\beta^p] \leq 2^{p-1} \sup_{0 < \varepsilon \leq 1} \mathbb{E}[\|X^\varepsilon\|_\beta^p] + 2^{p-1} \mathbb{E}[\|\bar{X}\|_\beta^p] \leq C(\mathbb{E}[\|B\|_\alpha^{\nu p}] + 1) < \infty$$

for every  $p \in [1, \infty)$ . This implies that  $\{\|X^\varepsilon - \bar{X}\|_\beta^p\}_{0 < \varepsilon \leq 1}$  are uniformly integrable for each fixed  $p$ . Hence, we have  $\mathbb{E}[\|X^\varepsilon - \bar{X}\|_\beta^p] \rightarrow 0$  as  $\varepsilon \searrow 0$ . This completes the proof of the main theorem.  $\square$

**Acknowledgement:** The author is supported by JSPS KAKENHI Grant No. 23K20216.

## References

- [1] H. Boedihardjo: X. Geng: *Lipschitz-stability of controlled rough paths and rough differential equations*, Osaka J. Math. **59** (2022), no. 3, 653–682.
- [2] M. Freidlin: A. D. Wentzell: Random perturbations of dynamical systems, Translated from the 1979 Russian original by Joseph Szücs, Third edition, Grundlehren der mathematischen Wissenschaften 260, Springer, Heidelberg, 2012.
- [3] P. Friz: M. Hairer: A course on rough paths, Springer, Cham, 2014.
- [4] P. Friz: N. Victoir: Multidimensional stochastic processes as rough paths, Cambridge University Press, Cambridge, 2010.
- [5] P. Gassiat: T. Klose: *Gaussian rough paths lifts via complementary Young regularity*, Electron. Commun. Probab. **29** (2024), Paper No. 75, 13 pp.
- [6] L. Gyurko: *Differential equations driven by  $\Pi$ -rough paths*, Proc. Edinb. Math. Soc. (2) **59** (2016), no. 3, 741–758.
- [7] F. A. Harang: *On the theory of rough paths, fractional and multifractional Brownian motion—with application to finance*, Master’s thesis (2015) University of Oslo. (Unpublished).  
Available at: <https://www.duo.uio.no/handle/10852/50068>
- [8] Y. Inahama: *Averaging principle for slow-fast systems of rough differential equations via controlled paths*, Tohoku Math. J. (2) **77** (2025), no. 2, 189–227.
- [9] R. Z. Khas’minskii: *On the principle of averaging the Itô’s stochastic differential equations*, Kybernetika **4** (1968), 260–279.
- [10] M. Li: Y. Li: B. Pei: Y. Xu: *Averaging principle for semilinear slow-fast rough partial differential equations*, Stochastic Process. Appl. **188** (2025), Paper No. 104683, 27 pp.

- [11] B. Pei: R. Hesse: B. Schmalfuss: Y. Xu: *Almost sure averaging for fast-slow stochastic differential equations via controlled rough path*, preprint (2023). arXiv:2307.13191.
- [12] B. Pei: Y. Inahama: Y. Xu: *Averaging principles for mixed fast-slow systems driven by fractional Brownian motion*, Kyoto J. Math. **63** (2023), no. 4, 721–748.
- [13] B. Pei: Y. Inahama: Y. Xu: *Averaging principle for fast-slow system driven by mixed fractional Brownian rough path*, J. Differential Equations **301** (2021), 202–235.
- [14] X. Yan: M. Ji: X.-G. Yang: A. Miranville: *Averaging principle of slow-fast systems driven by fractional Brownian motion*, preprint (2024).  
Available at: <https://www.researchgate.net>
- [15] X. Yang: Y. Xu: *Large deviation principle for slow-fast rough differential equations via controlled rough paths*, to appear in Proceedings of the Royal Society of Edinburgh Section A-Mathematics.

Yuzuru INAHAMA  
 Faculty of Mathematics,  
 Kyushu University,  
 744 Motooka, Nishi-ku, Fukuoka, 819-0395, JAPAN.  
 Email: [inahama@math.kyushu-u.ac.jp](mailto:inahama@math.kyushu-u.ac.jp)