PROVING AN ENHANCED SANOV-TYPE LARGE DEVIATION PRINCIPLE ON MULTIDIMENSIONAL (α, β) ROUGH PATHS

BASILE TERVER

Contents

Date: April-August 2023.

Abstract. The (α, β) rough paths of [17] were introduced to get a direct rough path version for the stochastic integral appearing in rough volatility models. It is then possible to prove a Large Deviation Principles (LDP) and many other properties on the process $\log(S_t)$ representing the log-price of an asset under rough volatility in a simpler way than using regularity structures (as in [3]). Usually only Schilder type LDPs or small time asymptotics are proven for rough volatility models, Stochastic Differential Equations (SDEs) and rough paths, as in [11], [3],[17] and [15] (Theorem 13.38, Proposition 19.14). On the other hand, Deuschel et al. [7] prove a Sanov type LDP for an ensemble of interacting Brownian rough paths. The main objective of this research project was to prove an enhanced Sanov theorem analogous to Theorem 3.6 of [7] for (α, β) rough paths as defined in [17], instead of Brownian rough paths. The first step was to define a N-layer driving noise generalising $(\hat{X}, X)$ of the rough volatility model and the lifting to an (α, β) rough path of such a N-layer noise. This imposed to define a N-dimensional Chen's relation and the Hölder regularity of an N-dimensional (α, β) rough path. The generalisation we define is conceptually non-trivial because it allows to generalise the SDEs considered in rough volatility to i.i.d. sequences of driving noises. That being done, our objective has been to prove a strong approximation estimate similar to [15](Corollary 13.21, Theorem 15.42) in the case of N-dimensional (α, β) rough paths. The reason for that is that it would allow to show a lemma analogous to Lemma 3.8 of [7] but for (α, β) rough paths. If we are able to prove such an analog, we expect the subsequent arguments of [7] allowing to prove the enhanced Sanov LDP to still hold for (α, β) rough paths. We were able to prove the strong approximation estimate in one of the cases. The remaining open cases are part of ongoing work.

Acknowledgements

I want to warmly thank my supervisor Antoine Jacquier for his constant support, his releavant advice and guidance, and his kindness throughout this research project. I would really like to thank Ioannis Gasteratos for his important role in providing mathematical insights, ideas and his pedagogy. I was honoured to work with such researchers that allowed to make this research project such an enriching period, scientifically but also in terms of personal reflection on my professionnal and academic project.

1. Introduction

1.1. Rough volatility.

A new paradigm in modelling the volatility of an asset's price, rough volatility models emerged after Gatheral et al. [18] observed for the first time remarkable consistence of their rough fractional stochastic volatility model with financial (high frequency) time series data. Fukasawa shows [16] that rough volatility models are the only class of continuous price models being consistent to a power law of the implied volatility term structure observed in equity options markets.

When considering derivatives, the log-price, denoted Y_t was classically modelled as a continuous semi-martingale:

$$dY_t = \mu_t dt + \sigma_t dX_t, \tag{1}$$

where μ_t is a drift term, σ_t is the volatility process and X_t is a one-dimensional Brownian motion.

In the Black-Scholes model, the volatility σ is constant or deterministic.

In local volatility models, like Dupire's [9], the local volatility $\sigma(Y_t, t)$ is a deterministic function of the underlying price and time, chosen to match observed European option prices exactly.

In stochastic volatility models, like the Hull and White model [28], the Heston model [27], and the SABR model [25], the volatility σ_t is modeled as a continuous Brownian semi-martingale. Stochastic volatility dynamics are more realistic than local volatility dynamics but generated option prices are not consistent with observed European option prices.

Let us focus, as in [3], on stochastic volatility, which models the asset price S_t by the following SDE, given in Itô differential form by

$$dS_t = S_t \sigma_t dX_t = S_t \sqrt{v_t(\omega)} dX_t,$$

where X is a standard Brownian motion and σ_t (resp. v_t) are known as stochastic volatility (resp. variance) process. Many classical Markovian asset price models fall in this framework, including Dupire's local volatility model, the SABR and Heston model. In all stochastic volatility model, one has Markovian dynamics for the variance process, of the form

$$dv_t = g(v_t) dW_t + h(v_t) dt,$$

with f, g deterministic functions and constant correlation $\rho := \frac{d\langle X, W\rangle_t}{dt}$ is modelled by defining a 2-dimensional (2D) standard Brownian motion $(W, \bar{W})$ and setting $X := \rho W + \bar{\rho}\bar{W}$ with $\bar{\rho} := \sqrt{1-\rho^2}$.

Framework. *In all this report, we work on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)_{t\geq 0}$ throughout, supporting two independent standard Brownian motions $(W, \bar{W})$ and satisfying the usual conditions of Definition 1.2.25 of* [32].

In the rough volatility framework, one is concerned with non-Markovian (fractional) stochastic volatility models, where σ_t is modelled via a fractional Brownian motion (fBM) in the regime $H \in (0, 1/2)$. The volatility is named rough because the Hölder regularity of the volatility process's sample paths is H^-, which is "rougher" than Brownian paths. Evidence for this rough regime has now been established ([2],[1]) and [18] suggest $H \simeq 0.14$. Most of the literature on this subject has focused on "simple" rough volatility models, defined as:

$$\begin{aligned} \sigma_t &:= f(\hat{X}_t), \text{ "simple" rough volatility} \\ \hat{X}_t &:= \int_0^t k_H(t-s)dW_s, \end{aligned} \tag{2}$$

with $k_H(r) := \frac{1}{\Gamma(H+1/2)} r^{H-\frac{1}{2}}$ the power law kernel on which is based the Riemann-Liouville definition of the fractional Brownian motion (see Table IV.1 of [10], Theorem 4.1 of [38] and [35]). In such a model, volatility is a function of a fractional Brownian motion $\hat{X}$, built from W, that has constant correlation ρ to the standard Brownian motion X.

Notation: In what follows, we will use both notations $\sigma_t = f(\hat{X}_t)$ and $\sigma_t = \sigma(\hat{X}_t)$ for the stochastic volatility process.

Remark 1.1. *Consider the SDE*

$$dS_t = S_t\sigma(W_t)dX_t,\ S_0 = x \in (0,\infty).$$

Using the 2D Itô's formula, writing $dX_t = \rho dW_t + \bar{\rho} d\bar{W}_t$ *with the* log *function, the above Stochastic Differential Equation (SDE) is equivalent to*

$$d(\log(S_t)) = \sigma(W_t)dX_t - \frac{1}{2}\sigma^2(W_t)dt,\ S_0 = x \in (0,\infty).$$

Denoting $Y_t := \log(S_t)$ *and* $\mu_t := -\frac{1}{2}\sigma^2(W_t)$, *the above equation is equivalent to* (1). *By the exact same arguments using adaptedness of* $\hat{X}$ *to* W *and independence between* $\hat{X}$ *and* $\bar{W}$, *the two following equations are equivalent:*

$$\begin{aligned} dS_t &= S_t\sigma(\hat{X}_t)dX_t,\ S_0 = x \in (0,\infty) \\ d(\log(S_t)) &= \sigma(\hat{X}_t)dX_t - \frac{1}{2}\sigma^2(\hat{X}_t)dt,\ S_0 = x \in (0,\infty). \end{aligned}$$

1.2. **Motivation and overview of the contribution of rough path theory.** Let us now give an overview of why rough path theory was invented and what is has brought in the realm of SDEs.

1.2.1. *SDEs are analytically ill-posed.* In Section 1.1 of [12], which explains the motivation of the theory of rough paths, we consider classical ordinary differential equations (ODEs) of the form $\dot{Y}_t = f(Y_t, t)$, of which an important subclass is given by controlled ODEs of the form

$$\dot{Y}_t = f_0(Y_t) + \sum_{i=1}^d f_i(Y_t)\dot{X}^i_t, \qquad Y_0 = y_0, \tag{3}$$

where X models the input (taking values in $\mathbb{R}^d$) and Y the output (in $\mathbb{R}^e$) of a system modelled by non-linear functions f_0 and f an by the initial state Y_0. If we rewrite $\dot{Y}_t = \frac{dY_t}{dt}$, (3) becomes $\dot{Y}_t = \frac{dY_t}{dt} = f_0(Y_t) + \sum_{i=1}^d f_i(Y_t)\frac{dX^i_t}{dt}$ which is equivalent to

$$dY_t = f_0(Y_t)dt + \sum_{i=1}^d f_i(Y_t)dX^i_t, \qquad Y_0 = y_0. \tag{4}$$

A non-smooth theory is needed when the system is subject to white noise, which can be interpreted as the scaling limit as $h \to 0$ of the discrete evolution equation

$$Y_{i+1} = Y_i + hf_0(Y_i) + \sqrt{h}f(Y_i)\xi_{i+1}, \tag{5}$$

where the (ξ_i) are i.i.d. standard Gaussian random variables. Itô's stochastic differential equations (SDEs) give a rigorous mathematical framework to treat this case with martingale arguments. However, passage to continous time makes us lose stability: although it is trivial to solve (5) for a fixed realisation of $\xi_i(\omega)$ (because $(\xi_1,\dots,\xi_T;Y_0) \mapsto Y_i$ is surely a continuous map), the continuity of the solution Y as a function of the driving noise X is lost in the limit. Indeed, taking $\dot{X}_t = \xi_t$ to be a white noise in time (which amounts to say that X is a Brownian motion, say B), the solution map $S : B \mapsto Y$ to (3), known as Itô map, is a measurable map which in general lacks continuity, whatever norm one uses to equip the space of realisations of B. Actually, one (cf. [31]) can show the negative result 1.2 (Proposition 1.1 of [12]), of purely deterministic sense.

Proposition 1.2. *There exists no separable Banach space* $\mathcal{B} \subseteq C([0,1])$ *with the following properties:*
(i) Sample paths of Brownian motions lie in $\mathcal{B}$ *almost surely.*
(ii) The map $(f,g) \mapsto \int_0^\cdot f(t)dg_t$ *defined on smooth functions extends to a continuous map from* $\mathcal{B} \times \mathcal{B}$

For example, consider B to be a d-dimensional Brownian motion (which means it has independent components $B^i, i \in \{1,\dots,d\}$), then for any two distinct indices i and j (otherwise, we are looking at an integral of B^i against itself), the map

$$B \mapsto \int_0^\cdot B^i_t dB^j_t, \tag{6}$$

is itself the solution of one of the simplest possible differential equations driven by B (take $Y \in \mathbb{R}^2$ solving $\dot{Y}_1 = \dot{B}^i$ and $\dot{Y}_2 = Y_1\dot{B}^j$ or $dY^2_t = Y^1_t dB^j_t$, we indeed have, provided $Y^1_0 = B^i_0$, that $Y^2_t =$

$\int_0^t B_s^i dB_s^j$). Therefore, we are in the framework of 1.2 and this shows that, even for a very simple SDE, S can lack continuity. In this sense, solving SDEs is an analytically ill-posed task.

On the other hand, before the emergence of rough path theory, well-known probabilistic well-posedness results had already been proven for SDEs like (4), but with integration interpreted in the Stratonovitch sense, that is

$$dY_t = f_0(Y_t)dt + \sum_{i=1}^d f_i(Y_t) \circ dB_t^i, \qquad Y_0 = y_0, \tag{7}$$

(see e.g. [37] Theorem 4.1) which states

Theorem 1.3. *Let $\xi_\epsilon = \delta_\epsilon * \xi$ denote the convolution of white noise in time with a compactly supported smooth mollifier δ_ϵ. Denote by Y^ϵ the solutions to* (3) *driven by $\dot{X} = \xi_\epsilon$. Then Y^ϵ converges in probability (uniformly on compact sets). The limiting process process does not depend on the choice of mollifier δ_ϵ, and in fact is the Stratonovitch solution to* (7)*.*

There are many variations on such "Wong-Zakai" results, another popular choice being $\xi_\epsilon = \dot{B}^{(\epsilon)}$ where $B^{(\epsilon)}$ is a piecewise linear approximation (of mesh size $\sim \epsilon$) to Brownian motion. However, as consequence of the aforementioned lack of continuity of the Itô-map, there are also reasonable approximations to white noise for which the above convergence fails.

1.2.2. *The Itô-Lyons map.* It turns out that well-posedness is restored via the iterated integrals (6) which are in fact the only data that is missing to turn the solution map $S : B \mapsto Y$ of (3) (or equivalently (4)) into a continuous map. Rough path analysis introduced by Terry Lyons in the pioneering article [34] and by now exposed in several monographs ([33],[39],[15]), provides the following essential result: Itô's solution map can be factorised into a measurable "universal" map ψ and a continuous solution map $\hat{S}$ as

$$B(\omega) \overset{\psi}{\mapsto} (B, \mathbb{B})(\omega) \overset{\hat{S}}{\mapsto} Y(\omega).$$

ψ is universal because it depends neither on the initial condition Y_0 nor on the vector fields f_0, f driving the stochastic differential equation (3) (or equivalently (4)) but only consists of enhancing the driving Brownian motion B with iterated integrals of the form

$$\mathbb{B}_{st}^{i,j} := \int_s^t B_{sr}^i dB_r^j, \tag{8}$$

where we denote $B_{sr}^i := B_r^i - B_s^i$. At this stage, the choice of stochastic integration in (8) (e.g. Itô or Stratonovich) does matter and probabilistic techniques are required for the construction of ψ. Indeed, the map ψ is only measurable and usually requires the use of some sort of stochastic integration theory or some equivalent construction, see for example Section 10 of [12] or Section 15 of [15] for a general construction in a Gaussian, non-semimartingale context. Indeed, we focused on the example of Brownian motion in the above, but the map $\psi : X \mapsto \mathbb{X}$, called the canonical lifting of the driving stochastic process X, is based on a stochastic integration technique.

On the other hand, the map $\hat{S}$, the solution map to a *rough differential equation* (RDE), also known as *Itô-Lyons map* (discussed in Section 8.1 of [12] and Sections 10 and 11 of [15]) is purely deterministic and only makes use of analytical constructions. More precisely, it allows input signals to be arbitrary *rough paths* which, as discussed in Chapter 2 of [12], are objects (thought of as enhanced paths) of the form $(X, \mathbb{X})$, defined via certain algebraic properties (those of the *free nilpotent groups*, which mimic the interplay between a path and its iterated integrals and are described in Chapter 7 of [15]) and certain analytical, Hölder-type regularity conditions, as one can see in Chapters 8 and 9 of [15]. These conditions will be seen to hold true a.s. for $(B, \mathbb{B})$ in Chapter 3 of [12] and 13 of [15], a typical realisation is thus called Brownian rough path.

Two of the most important results of the theory of rough paths are the continuity and definition of the Itô-Lyons map and the lifting of a class of Gaussian processes to the space of rough paths (Section 10 of [12] or Section 15 of [15]). We have just introduced right above the Itô-Lyons map. Let us now depict briefly how the second main contribution of rough path theory is to allow for a theory of stochastic differential equations with new types of driving signals, signals that are non-semi-martingales. In the body of the report, I will focus more on the lifting of Gaussian processes, as my goal has been, for most of my research project, to prove a strong approximation (by piecewise linear approximations) lemma of the (α, β) lifting defined by Fukasawa [17].

1.2.3. *Allowing for stochastic differential equations with more general driving signals.* It is remarkable that almost every Brownian sample path $(B_t(\omega) : t \in [0, T])$ has infinite variation and there is no help from the classical Riemann-Stieltjes integration theory to build integrals of the form $\int \ldots dB$. Instead, Itô's theory of stochastic integration relies crucially on the fact that B is a martingale and stochastic integrals themselves are constructed as martingales. If one recalls the elementary interpretation of martingales as fair games one sees that Itô integration is some sort of martingale transform in which the integrand has the meaning of a gambling strategy. Clearly then, the integrand must not anticipate the random movements of the driving Brownian motion and one is led to the class of so-called previsible processes which can be integrated against Brownian motion. When such integration is possible, it allows for a theory of SDEs of the form of (4) with X being a d-dimensional Brownian motion (that is, with d independent components).

It is natural to ask whether the meaning of (4) can be extended to driving processes X other than Brownian motion. For instance, there is motivation from mathematical finance to generalize the driving process to general (semi-)martingales and luckily Itô's approach can be carried out naturally in this context.

We can also ask for a Gaussian generalization, for instance by considering a differential equation of the form (4) in which the driving signal may be taken from a reasonably general class of Gaussian processes. Such equations have been proposed, often in the setting of fractional Brownian motion of Hurst parameter $H > 1/2$, as toy models to study the ergodic behaviour of non-Markovian systems or to provide new examples of arbitrage-free markets under transactions costs. Or we can ask for a Markovian generalization. Indeed, it is not hard to think of motivating physical examples (such as heat flow in rough media) in which the Brownian motion B may be replaced by a Markov process with uniformly elliptic generator in divergence form.

The Gaussian and Markovian examples have in common that the sample path behaviour can be arbitrarily close to Brownian motion (e.g. by taking $H = 1/2 \pm \epsilon$ for the Hurst parameter of the fractional Brownian motion). And yet, Itô's theory has a complete breakdown! This is due to the absence of the martingale property for such processes.

It has emerged over recent years, with the seminal work of Lyons [34], that differential equations driven by such non-semi-martingales can be solved in the rough path sense. Moreover, the so-obtained solutions have firm probabilistic justification. For instance, if the driving signal converges to Brownian motion (in some reasonable sense which covers $\epsilon \to 0$ in the aforementioned example) the corresponding rough path solutions converge to the classical Stratonovich solution of (4), as one would hope.

1.2.4. *More on the Itô-Lyons map.* While this alone seems to allow for flexible and robust stochastic modelling, the contribution of rough path theory is not all about dealing with new types of driving signals. Even in the classical case of Brownian motion, we get some remarkable insights. Namely, the (Stratonovich) solution to (7) can be represented as a deterministic and continuous image of Brownian motion and Lévy's stochastic area

$$A_t^{jk}(\omega) = \frac{1}{2}\left(\int_0^t B^j dB^k - \int_0^t B^k dB^j\right) \tag{9}$$

alone. In fact, the Itô-Lyons map yields, upon setting $\mathbf{x} = (B^i, A^{j,k} : i, j, k \in \{1, ..., d\})$ a very pleasing version of the solution of (7). Indeed, subject to sufficient regularity of the coefficients, we see that (7) can be solved simultaneously for all starting points y_0, and even all coefficients. Clearly then, one can allow the starting point and coefficients to be random (even dependent on the entire future of the Brownian driving signals) without problems; in stark contrast to Itô's theory which struggles with the integration of non-previsible integrands. Also, construction of stochastic flows becomes a trivial corollary of purely deterministic regularity properties of the Itô-Lyons map. This brings us to the (deterministic) main result of the theory: continuity of the Itô-Lyons map $\hat{S}$ in "rough path" topology. When applied in a standard SDE context, it quickly gives an entire catalogue of limit theorems. It also allows to reduce (highly non-trivial) results, such as the Stroock-Varadhan support theorem or the Freidlin-Wentzell estimates (Theorem 3.28), to relatively simple statements about Brownian motion and Lévy's area.

The Itô-Lyons map turns out (Section 8.6 of [12]) to be "nice" in the sense that it is a continuous map of both its initial condition and the driving noise $(X, \mathbb{X})$, provided that the dependency on the latter is measured in a suitable "rough path" metric. In other words, rough path analysis allows for a pathwise solution theory for SDEs, i.e. for a fixed realisation of the Brownian rough path. The solution map $\hat{S}$ is however a much richer object than the original Itô map, since its construction is completely independent of the choice of stochastic integral and even of the knowledge that the driving path is Brownian. For example, if we denote by ψ^I (resp. ψ^S) the maps $B \mapsto (B, \mathbb{B})$ obtained by Itô (resp. Stratonovitch) integration, then we have the almost sure identities

$$S^I = \hat{S} \circ \psi^I, \quad S^S = \hat{S} \circ \psi^S$$

where S^I (resp. S^S) denotes the solution to (4) interpreted in the Itô (resp. Stratonovitch) sense. Returning to Theorem 1.3, we see, by continuity of $\hat{S}$ that the convergence in this theorem is really a deterministic consequence of the probabilistic question whether or not $\psi^S(B^\epsilon) \to \psi^S(B)$ in probability and rough path topology, with $\dot{B}^\epsilon = \xi^\epsilon$. This can be shown to hold in the case of mollifier, piecewise linear, and many other approximations.

Part 1. Background on rough volatility, rough path theory and large deviations

2. Elements of rough path theory

In what follows, unless he is perfectly familiar with the theory of rough paths, the reader is advised to have the following two manuals at hand [15], [12]. As we will skip the details for the sake of brevity, we will indicate any references to each of these books so that the curious reader can refer to them. More, the proofs we write in this report will often be there because they are not in one of the two books.

Although a Banach formulation of the theory of rough paths is possible, we shall remain in finite dimensions (with $\mathbb{R}^d$) in what follows.

Although it will not always be stated, most of the statements made for finite p-variation continuous paths in what follows will hold true for $1/p$-Hölder paths. This is due to Remark 2.3.

2.1. Variation and Hölder spaces.

Definition 2.1 (Supremum/infinity distance, Definition 1.1 of [15]). *Let (E, d) be a metric space and $[0, T] \subseteq \mathbb{R}$. Then $C([0, T], E)$ denotes the set of all continuous paths $x : [0, T] \to E$. The supremum or infinity distance of $x, y \in C([0, T], E)$ is defined by*

$$d_{\infty;[0,T]}(x, y) := \sup_{t \in [0,T]} d(x_t, y_t).$$

For a single path $x \in C([0,T],E)$, we set

$$|x|_{0;[0,T]} := \sup_{u,v\in[0,T]} d(x_u, x_v),$$

and, given a fixed element $o \in E$, identified with the constant path $\equiv o$,

$$|x|_{\infty;[0,T]} := d_{\infty;[0,T]}(o,x) = \sup_{u\in[0,T]} d(o, x_u).$$

If no confusion is possible we shall omit [0, T] and simply write $d_\infty, |\cdot|_0$ and $|\cdot|_\infty$. If E has a group structure such as $(\mathbb{R}^d,+)$ the neutral element is the usual choice for o. In the present generality, however, the definition of $|\cdot|_\infty$ depends on the choice of o.

Notation 2.1. *Let us also agree that $C_o([s,t],E)$ denotes those paths in $C([s,t],E)$ which start at o, i.e.*

$$C_o([s,t],E) = \{x \in C([s,t],E) : x(s) = o\}$$

Definition 2.2 (α-Hölder and p-variation distances, Definition 5.1 of [15])**.** *Let (E,d) be a Polish space with a compatible structure of Lie group (it will be $\mathbb{R}^e$ or $G^N(\mathbb{R}^e)$). A path $x:[0,T]\to E$ is said to be*

(i) Hölder continuous with exponent $\alpha \geq 0$, or simply α-Hölder, if

$$|x|_{\alpha-Hol;[0,T]} := \sup_{0\leq s<t\leq T} \frac{d(x_s,x_t)}{(t-s)^\alpha} < \infty; \tag{10}$$

(ii) of finite p-variation for some $p>0$ if

$$|x|_{p-var;[0,T]} := \left(\sup_{(t_i)\in\mathcal{D}([0,T])} \sum_i d(x_{t_i}, x_{t_{i+1}})^p \right)^{1/p} < \infty. \tag{11}$$

We will use the notations $C^{\alpha-Hol}([0,T],E)$ for the set of α-Hölder paths x and $C^{p-var}([0,T],E)$ for the set of continuous paths $x:[0,T]\to E$ of finite p-variation.

$C^{\alpha-Hol}([0,T],E)$ is a complete metric space endowed with the distance

$$d_{\alpha-Hol}(x,y) = \sup_{t\in[0,T]} d(x_t,y_t) + \sup_{0\leq s<t\leq T} \frac{d(x_s^{-1}x_t, y_s^{-1}y_t)}{(t-s)^\alpha}.$$

This space is not separable in general. However, the subspace $C^{0,\alpha-Hol}([0,T];E)$ given by the closure, with respect of $d_{\alpha-Hol}$, of the smooth (C^∞) paths is separable, hence Polish.

It is obvious from these definitions that a path $x:[0,T]\to E$ is constant, i.e. $x_t \equiv o$ for some $o\in E$, if and only if $|x|_{\alpha-Hol;[0,T]} = 0$ and if and only if $|x|_{p-var;[0,T]} = 0$. (In particular, if $E=\mathbb{R}^d$ our quantities (10), (11) are only semi-norms.) Observe that $C^{0-Hol}([0,T],E)$ is nothing but the set of continuous paths from $[0,T]$ into E and $|x|_{0-Hol;[0,T]} = |x|_{0;[0,T]}$, where the latter was defined in 2.1.

Remark 2.3. *Any $\alpha>0$ can be written as $\alpha=1/p$. If x is a $(1/p)$-Hölder path, then is a continuous path of finite p-variation and we have $|x|_{p-var;[0,T]} \leq T^{1/p}\|x\|_{\frac{1}{p}-Hol}$.*

Proof. Remark, by 19.1, that, for any x of bounded p-variation

$$|x|_{p-var;[0,T]} := \sup_{(t_i)\in\mathcal{D}([0,T])} \left(\sum_i d(x_{t_i}, x_{t_{i+1}})^p \right)^{1/p}.$$

Now consider x to be a $1/p$-Hölder-continuous path with $p>0$. Fix a dissection $(t_i)\in\mathcal{D}([0,T])$ and denote $\|\cdot\|_p$ the l^p norm on vectors of $\#(t_i)$ components. Then,

$$\begin{aligned}\left(\sum_i d(x_{t_i}, x_{t_{i+1}})^p \right)^{1/p} &= \|d(x_{t_i}, x_{t_{i+1}})\|_p = \|\frac{d(x_{t_i}, x_{t_{i+1}})}{(t_{i+1}-t_i)^{1/p}}(t_{i+1}-t_i)^{1/p}\|_p \\ &\leq T^{1/p}\|\frac{d(x_{t_i}, x_{t_{i+1}})}{(t_{i+1}-t_i)^{1/p}}\|_p \leq T^{1/p}\|x\|_{\frac{1}{p}-Hol}\end{aligned}$$

□

Conversely, although a path of finite p-variation need not be continuous (e.g. a step-function), our focus is on continuous paths, hence the definition of C^{p-var} as a space of continuous paths of finite p-variation. It is a well-know fact in analysis that, if x is continuous and has finite p-variation, there exists a reparameterisation, τ , such that $f\circ\tau$ is $1/p-$Hölder continuous. Moreover, the p-variation of a path is independent of its parametrisation.

The following simple proposition then explains why our main interest lies in $\alpha\in[0,1]$ and $p\geq 1$.

Proposition 2.4 (Proposition 5.2 of [15])**.** *Assume $x:[0,T]\to E$ is α-Hölder continuous, with $\alpha\in(1,\infty)$, or continuous of finite p-variation with $p\in(0,1)$. Then x is constant, i.e. $x(\cdot)\equiv x_0$.*

By the proposition below, we have a compact embedding of variation and Hölder spaces.

Proposition 2.5 (Proposition 5.3 of [15])**.** *Let $x\in C([0,T],E)$ and $1\leq p\leq p'<\infty$. Then,*

$$|x|_{p'-var;[0,T]} \leq |x|_{p-var;[0,T]}.$$

In particular, $C^{p-var}([0,T],E)\subseteq C^{p'-var}([0,T],E)$.

Similarly, for Hölder norms, for $1\geq\alpha\geq\alpha'\geq 0$, we have

$$|x|_{\alpha'-Hol;[0,T]} \leq T^{\alpha'-\alpha}|x|_{\alpha-Hol;[0,T]}.$$

Proof. The p-variation statement follows from the elementary inequality of 19.5 like in 2.3.

For the Hölder case, use

$$\frac{d(x_{t_i},x_{t_{i+1}})}{(t-s)^{\alpha}}=\frac{d(x_{t_i},x_{t_{i+1}})}{(t-s)^{\alpha'}(t-s)^{\alpha-\alpha'}}\geq\frac{d(x_{t_i},x_{t_{i+1}})}{(t-s)^{\alpha'}T^{\alpha-\alpha'}}.$$

□

Proposition 2.6 (interpolation, Proposition 5.5 of [15]). *Let $x\in c([0,T],E)$. (i) For $1\leq p<p'<\infty$, we have*

$$|x|_{p'-var;[0,T]}\leq|x|_{p-var;[0,T]}^{p/p'}|x|_{0;[0,T]}^{1-p/p'}.$$

(ii) For $1\geq\alpha\geq\alpha'\geq0$, we have

$$|x|_{\alpha'-Hol;[0,T]}\leq|x|_{\alpha-Hol;[0,T]}^{\alpha'/\alpha}|x|_{0;[0,T]}^{1-\alpha'/\alpha}.$$

2.2. **Riemann-Stieltjes and Young integration.** Define $L(\mathbb{R}^d,\mathbb{R}^e)$ the set of all linear maps from $\mathbb{R}^d$ to $\mathbb{R}^e$. In what follows, we denote the increment of a path $x:[0,T]\to E$ with image in a normed vector space E between $0\leq s<t\leq T$ as $x_{s,t}:=x_t-x_s$.

Definition 2.7 (Riemann-Stieltjes integral, Definition 2.1 of [15]). *Let x and y be two functions from $[0,T]$ into $\mathbb{R}^d$ and $L(\mathbb{R}^d,\mathbb{R}^e)$. Let $D_n=(t_i^n:i)$ be a sequence of dissections of $[0,T]$ with $|D_n|\to0$, and ξ_i^n some points in $[t_i^n,t_{i+1}^n]$. Assume $\sum_{i=0}^{\#D_n-1}y(\xi_i^n)x_{t_i^n,t_{i+1}^n}$ converges when n tends to ∞ to a limit I independent of the choice of ξ_i^n and the sequence (D_n). Then we say that the Riemann–Stieltjes integral of y against x (on $[0,T]$) exists and write*

$$\int_0^T ydx:=\int_0^T y_udx_u:=I$$

We call y the integrand and x the integrator. Of course, $[0,T]$ may be replaced by any other interval $[s,t]$.

Proposition 2.8 (Proposition 2.2 of [15]). *Let $x\in C^{1-var}([0,T],\mathbb{R}^d)$ and $y:[0,T]\to L(\mathbb{R}^d,\mathbb{R}^e)$ piecewise continuous*[1]*. Then, the Riemann-Stieltjes integral $\int_0^T ydx$ exists, is linear in y and x, and we have the estimate*

$$\left|\int_0^T ydx\right|\leq|y|_{\infty;[0,T]}|x|_{1-var;[0,T]}.$$

In particular, $t\in[0,T]\mapsto\int_0^t ydx$ is in $C^{1-var}([0,T],\mathbb{R}^e)$ (by Propositions 1.11 and 1.12 of [15]).

Moreover[2]

$$\int_0^t y_udx_u-\int_0^s y_udx_u=\int_s^t y_udx_u\text{ for all }0\leq s<t\leq T.\tag{12}$$

Proposition 2.9 (Proposition 2.4 (integration by parts) of [15]). *Let $x\in C^{1-var}([0,T],\mathbb{R}^d)$ and $x\in C^{1-var}([0,T],L(\mathbb{R}^d,\mathbb{R}^e))$. Then*

$$\int_0^T y_udx_u+\int_0^T(dy_u)dx_u=y_Tx_T-y_0x_0$$

Proposition 2.10 (Young-Lóeve estimate, Proposition 6.4 of [15]). *Assume*

$$x\in C^{1-var}([0,T],\mathbb{R}^d),\ y\in C^{1-var}([0,T],L(\mathbb{R}^d,\mathbb{R}^e))$$

so that we can define the Riemann-Stieltjes integral $\int ydx$ (according to proposition 2.8) and let $p,q\geq1$ with $\theta:=1/p+1/q>1$. Define

$$\Gamma_{s,t}:=\int_s^t y_udx_u-y_sx_{s,t}=\int_s^t y_{s,u}dx_u$$

We have

$$|\Gamma_{s,t}|\leq\frac{1}{1-2^{1-\theta}}|x|_{p-var;[s,t]}|y|_{q-var;[s,t]}.$$

2.3. **Free nilpotent groups and their variation and Hölder spaces.** Let us denote the symmetric and antisymmetric parts of any real matrix $A\in R^{d\times d}\simeq\mathbb{R}^d\otimes\mathbb{R}^d$:

$$\mathrm{Sym}(A)=\frac{1}{2}(A+A^T),\qquad\mathrm{Anti}(A)=\frac{1}{2}(A-A^T)$$

where ^{T} denotes the usual transposition on an element of a vector space.

Let $(e_i)_{i=1,\dots,d}$ denote the canonical basis of $\mathbb{R}^d$. Note that, as vector spaces $(\mathbb{R}^d)^{\otimes k}\simeq\mathbb{R}^{d^k}$ and set the convention $(\mathbb{R}^d)^{\otimes0}:=\mathbb{R}$.

In what follows, E is a normed vector space.

Definition 2.11 (Definition 7.2 of [15]). *The step-N signature of $x\in C^{1-var}([s,t],\mathbb{R}^d)$ is given by*

$$S_N(x)_{s,t}:=\left(1,\int_{s<u<t}dxu,\dots,\int_{s<u_1<\dots<u_k<t}dx_{u_1}\otimes\dots\otimes dx_{u_k}\right)\in\bigoplus_{k=0}^N(\mathbb{R}^d)^{\otimes k}.$$

Each of the iterated integrals above are well-defined according to 2.8 with $\int_{s<u_1<\dots<u_{l-1}<t}dx_{u_1}\otimes\dots\otimes dx_{u_{l-1}}\in C^{1-var}([s,u_l],(\mathbb{R}^d)^{l-1})$ instead of y integrated with respect to $x\in C^{1-var}([s,u_l],\mathbb{R}^d)$.

The path $u\mapsto S_N(x)_{s,u}$ is called the (step-N) lift of x.

---

[1]this covers all our applications, in particular $y\in C^{1-var}$.

[2]All integrals in (12) are understood in the sense of Definition 2.7 with $[0,T]$ replaced by $[0,t],[0,s],[s,t]$ respectively.

Given a vector $a \in (\mathbb{R}^d)^{\otimes k}$ one can write

$$a = \sum_{0 \le i_1, \dots, i_k \le d} a^{i_1, \dots, i_k} e_{i_1} \otimes \cdots \otimes e_{i_k} \in (\mathbb{R}^d)^{\otimes k} \tag{13}$$

then, with similar notation $b \in (\mathbb{R}^d)^{\otimes l}$, we agree that $a \otimes b$ is defined by

$$\begin{aligned} a \otimes b = & \sum_{0 \le i_1, \dots, i_k, j_1, \dots, j_l \le d} a^{i_1, \dots, i_k} b^{j_1, \dots, j_l} e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e_{j_1} \otimes \cdots \otimes e_{j_l} \\ & \in (\mathbb{R}^d)^{\otimes k} \otimes (\mathbb{R}^d)^{\otimes l} \simeq (\mathbb{R}^d)^{\otimes (k+l)}. \end{aligned} \tag{14}$$

In fact, this is the linear extension of the map $\otimes : (\mathbb{R}^d)^{\otimes k} \times (\mathbb{R}^d)^{\otimes l} \to (\mathbb{R}^d)^{\otimes k} \otimes (\mathbb{R}^d)^{\otimes l}$ which maps the tuple $(e_{i_1}, \dots, e_{i_k}, e_{j_1}, \dots, e_{j_l})$ to the $(i_1, \dots, i_k, j_1, \dots, j_k)$th basis element of $(\mathbb{R}^d)^{\otimes k} \otimes (\mathbb{R}^d)^{\otimes l}$, for which we already used the suggestive notation $e_{i_1} \otimes \cdots \otimes e_{i_k} \otimes e_{j_1} \otimes \cdots \otimes e_{j_l}$.

We now define

$$T^N(\mathbb{R}^d) := \bigoplus_{k=0}^{N} (\mathbb{R}^d)^{\otimes k}$$

and write $\pi_k : T^N(\mathbb{R}^d) \to (\mathbb{R}^d)^{\otimes k}$ for the projection to the kth tensor level. We shall also use the projection

$$\pi_{0,k} : T^N(\mathbb{R}^d) \to T^k(\mathbb{R}^d), \quad \text{for} \quad k \le N,$$

which maps $g = (g^0, \dots, g^N) \in T^N(\mathbb{R}^d)$ into $(g^0, \dots, g^k) \in T^k(\mathbb{R}^d)$. Here, $g^k \in (\mathbb{R}^d)^{\otimes k}$ is sometimes referred to as the kth level of g.

Definition 2.12. *Given $g, h \in T^N(\mathbb{R}^d)$, one extends* (14)*, which was the tensor product between elements of $(\mathbb{R}^d)^{\otimes k}$ and $(\mathbb{R}^d)^{\otimes l}$ to $T^N(\mathbb{R}^d)$ by setting*

$$g \otimes h = \sum_{i,j \ge 0; i+j \le N} g^i \otimes h^j \Leftrightarrow \forall k \in \{0, \dots, N\} : \pi_k(g \otimes h) = \sum_{i=0}^{N} g^i \otimes h^{k-i}.$$

The vector space $T^N(\mathbb{R}^d)$ becomes an (associative) algebra under $\otimes$. More precisely, we have

Proposition 2.13 (Proposition 7.4 of [15])**.** *The space $(T^N(\mathbb{R}^d), +, \cdot; \otimes)$ is an associative algebra with neutral element*

$$1 := (1, 0, ..., 0) = 1 + 0 + \ + 0 \in T^N(\mathbb{R}^d).$$

(The unit element for $+$ is $0 = (0, 0, ..., 0)$, of course.) We will call $T^N(\mathbb{R}^d)$ the truncated tensor algebra of level N.

Remark 2.14. *Similar to the algebra of square matrices, the algebra product is not commutative (unless $N = 1$ or $d = 1$).*

Let us now define a norm on $T^N(\mathbb{R}^d)$. To this end, we equip each tensor level $(\mathbb{R}^d)^{\otimes k}$ with Euclidean structure, which amounts to declaring the canonical basis $\{e_{i_1} \otimes \cdots \otimes e_{i_k} : i_1, \dots, i_k \in \{1, \dots, d\}\}$ to be orthonormal so that for any $a \in (\mathbb{R}^d)^{\otimes k}$ of form (13)

$$|a|_{(\mathbb{R}^d)^{\otimes k}} = \sqrt{\sum_{i_1, \dots, i_k} |a^{i_1, \dots, i_k}|^2}$$

and when no confusion is possible we shall simply write $|a|$. Let us also observe that for $0 \le i \le k \le N$,

$$(a, b) \in (\mathbb{R}^d)^{\otimes i} \times (\mathbb{R}^d)^{\otimes (k-i)}, \ |a \otimes b|_{(\mathbb{R}^d)^{\otimes k}} = |a|_{(\mathbb{R}^d)^{\otimes i}} |b|_{(\mathbb{R}^d)^{\otimes (k-i)}}$$

which is a compatibility relation between the tensor norms on the respective tensor levels. Then, for any $g \in T^N(\mathbb{R}^d)$ we set

$$|g|_{T^N(\mathbb{R}^d)} := \sum_{k=0}^{N} |\pi_k(g)|_{(\mathbb{R}^d)^{\otimes k}}$$

which makes $T^N(\mathbb{R}^d)$ a Banach space (of finite dimension $1 + d + d^2 + \cdots + d^N$); again we shall write $|g|$ if no confusion is possible. We remark that there are other choices of norms on $T^N(\mathbb{R}^d)$, like the max norm $|g|_{T^N(\mathbb{R}^d), \infty} := \max_{k=0}^{N} |\pi_k(g)|_{(\mathbb{R}^d)^{\otimes k}}$, of course all equivalent.

Given $x \in C^{1-var}([0, T], \mathbb{R}^d)$, the path $S_N(x)_{s,\cdot}$ takes values in $T^N(\mathbb{R}^d) \simeq \mathbb{R}^{1+d+\cdots+d^N}$, as vector spaces.

The following theorem says the signature of the concatenation of $(x_v)_{s \le v \le t}$ and $(x_v)_{t \le v \le u}$ is precisely the tensor product of the respective signatures of $(x_v)_{s \le v \le t}$ and $(x_v)_{t \le v \le u}$.

Theorem 2.15 (Theorem 7.11 of [15])**.** *Given $x \in C^{1-var}([0, T], \mathbb{R}^d)$ and $0 \le s < t < u \le T$ we have*

$$S_N(x)_{s,u} = S_N(x)_{s,t} \otimes S_N(x)_{t,u}$$

In the case $N = 2$, let $X \in C^{1-var}([0, T], \mathbb{R}^d)$ and denote $(1, X, \mathbb{X}) := S_2(X) =: \mathbf{X}$, meaning $\mathbb{X} := \pi_2(\mathbf{X})$. Then, by 2.12 we have $S_N(X)_{s,t} \otimes S_N(X)_{t,u} = (1, X_{s,t}, \mathbb{X}_{s,t}) \otimes (1, X_{t,u}, \mathbb{X}_{t,u}) = 1 + X_{s,t} + X_{t,u} +$

$\mathbb{X}_{s,t}+\mathbb{X}_{t,u}+X_{s,t}\otimes X_{t,u} = 1+X_{s,u}+\mathbb{X}_{s,t}+\mathbb{X}_{t,u}+X_{s,t}\otimes X_{t,u}$. On the other hand, $S_N(X)_{s,u} = (1, X_{s,u}, \int_{v=s}^{u}\int_{w=s}^{v} dX_w \otimes dX_v)$ and

$$\int_{v=s}^{u}\int_{w=s}^{v} dX_w\otimes dX_v = \int_{v=s}^{t}\int_{w=s}^{v} dX_w\otimes dX_v + \int_{v=t}^{u}\int_{w=s}^{v} dX_w\otimes dX_v$$
$$= \mathbb{X}_{s,t} + \int_{v=t}^{u}\int_{w=s}^{t} dX_w\otimes dX_v + \int_{v=t}^{u}\int_{w=t}^{v} dX_w\otimes dX_v = \mathbb{X}_{s,t}+\mathbb{X}_{t,u} + \int_{v=t}^{u}\int_{w=s}^{t} dX_w\otimes dX_v$$

which is consistent with 2.15. The relations

$$X_{s,u} = X_{s,t}+X_{t,u}$$
$$\mathbb{X}_{s,u} = \mathbb{X}_{s,t}+\mathbb{X}_{t,u}+X_{s,t}\otimes X_{t,u}$$

are called the step-2 *Chen's relation.*

Definition 2.16 (Definition 7.13 of [15])**.** *For $\lambda\in\mathbb{R}$, we define the dilation map*

$$\delta_\lambda : T^N(\mathbb{R}^d)\to T^N(\mathbb{R}^d)$$

such that $\pi_k(\delta_\lambda(g)) = \lambda^k\pi_k(g)$.

One can easily check that if λ is a real, $x:[0,T]\to\mathbb{R}^d$ is a continuous path of bounded variation, λx the path x scaled by λ, then $S_N(\lambda x)_{s,t} = \delta_\lambda(S_N(x)_{s,t})$.

We introduce two simple subspaces (linear and affine-linear, respectively) which will guide us towards the (crucial) free nilpotent Lie algebra and group and their Lie algebra.

Definition 2.17. *Let us set*

$$t^N(\mathbb{R}^d) := \{g\in T^N(\mathbb{R}^d) : \pi_0(g) = 0\}$$

so that

$$1+t^N(\mathbb{R}^d) := \{g\in T^N(\mathbb{R}^d) : \pi_0(g) = 1\}$$

By Lemma 7.16 of [15], we have that any element of $1+t^N(\mathbb{R}^d)$ is invertible with respect to the tensor product $\otimes$.

It is also obvious that if g and h are in $1+t^N(\mathbb{R}^d)$, then $g\otimes h\in 1+t^N(\mathbb{R}^d)$. Finally, $1+t^N(\mathbb{R}^d)$ is an affine-linear subspace of $T^N(\mathbb{R}^d)$ hence a smooth manifold, trivially diffeomorphic to $t^N(\mathbb{R}^d)\simeq R^{d+\dots+d^N}$. Noting that the group operations $\otimes$, $^{-1}$ are smooth maps (in fact, polynomial when written out in coordinates) we have

Proposition 2.18. *The space $1+t^N(\mathbb{R}^d)$ is a Lie group with respect to tensor multiplication $\otimes$.*

The vector space $(t^N(\mathbb{R}^d),+,\cdot)$ becomes itself an algebra under $\otimes$. As in every algebra, the *commutator*, in our case

$$(g,h)\mapsto [g,h] := g\otimes h - h\otimes g\in t^N(\mathbb{R}^d)$$

for $g,h\in t^N(\mathbb{R}^d)$, defines a bilinear map which is easily seen to be *anti-commutative*, i.e. $[g,h] = -[h,g]$, for all $g,h\in t^N(\mathbb{R}^d)$ and to satisfy the *Jacobi identity* for all $g,h,k\in t^N(\mathbb{R}^d)$; that is,

$$[g,[h,k]]+[h,[k,g]]+[k,[g,h]] = 0.$$

Recalling that a vector space $V=(V,+,\cdot)$ equipped with a bilinear, anti-commutative map $[\cdot,\cdot]:V\times V\to V$ which satisfies the Jacobi identity is called a *Lie algebra* (the map $[\cdot,\cdot]$ is called the Lie bracket), this can be summarized as

Proposition 2.19 (Proposition 7.19 of [15])**.** *$(t^N(\mathbb{R}^d),+,\cdot,[\cdot,\cdot])$ is a Lie algebra.*

We now define the exponential and logarithm maps via their power series:

Definition 2.20 (Definition 7.20 of [15])**.** *The exponential map is defined by*

$$\exp : t^N(\mathbb{R}^d)\to 1+t^N(\mathbb{R}^d)$$
$$a\mapsto 1+\sum_{k=1}^{N}\frac{a^{\otimes k}}{k!}$$

while the logarithm map is defined by

$$\log : 1+t^N(\mathbb{R}^d)\to t^N(\mathbb{R}^d)$$
$$1+a\mapsto \sum_{k=1}^{N}(-1)^{k+1}\frac{a^{\otimes k}}{k}.$$

A direct calculation shows that $\exp(\log(1+a)) = a, \log(\exp(a)) = a$, for all $a\in t^N(\mathbb{R}^d)$.

Definition 2.21 (Definition 7.25 of [15])**.** *Define $g^N(\mathbb{R}^d)\subseteq t^N(\mathbb{R}^d)$ as the smallest sub-Lie algebra of $t^N(\mathbb{R}^d)$ which contains $\pi_1(t^N(\mathbb{R}^d))\simeq\mathbb{R}^d$. That is,*

$$g^N(\mathbb{R}^d) = \mathbb{R}^d\oplus[\mathbb{R}^d,\mathbb{R}^d]\oplus\cdots\oplus[\mathbb{R}^d,[\dots,[\mathbb{R}^d,\mathbb{R}^d]]]$$

where the last $[\mathbb{R}^d,[\dots,[\mathbb{R}^d,\mathbb{R}^d]]]$ is made of $N-1$ iterated brackets. We call it the free step-N nilpotent Lie algebra.

For example, $[\mathbb{R}^d, \mathbb{R}^d]$ equals the set of real $d \times d$ antisymmetric matrices.

Let us consider

(i) the set of all step-N signatures of continuous paths of finite length,

$$G^N(\mathbb{R}^d) := \{S_N(x)_{0,1} : x \in C^{1-var}([0,T], \mathbb{R}^d)\};$$

(ii) the image of the sub-Lie algebra $g^N(\mathbb{R}^d) \subseteq t^N(\mathbb{R}^d)$, cf. Definition (2.21), under the exponential map

$$\exp(g^N(\mathbb{R}^d)) \subseteq 1 + t^N(\mathbb{R}^d).$$

(iii) the subgroup $\langle \exp(\mathbb{R}^d) \rangle$ of $1 + t^N(\mathbb{R}^d)$ generated by elements in $\exp(\mathbb{R}^d)$, i.e.

$$\langle \exp(\mathbb{R}^d) \rangle := \{\bigotimes_{i=1}^{m} \exp(v_i) : m \geq 1, v_1, \ldots, v_m \in \mathbb{R}^d\}.$$

Theorem 2.22 (Theorem 7.30 of [15])**.**

$$G^N(\mathbb{R}^d) = \exp(g^N(\mathbb{R}^d)) = \langle \exp(\mathbb{R}^d) \rangle$$

and $G^N(\mathbb{R}^d)$ is a closed sub-Lie group of $(1 + t^N(\mathbb{R}^d), \otimes)$, called the free nilpotent group of step N over $\mathbb{R}^d$.

Let us now define the Carnot-Caratheodory norm.

Definition 2.23 (Theorem 7.32 of [15])**.** *For every $g \in G^N(\mathbb{R}^d)$, the so-called "Carnot-Caratheodory norm"*

$$\|g\| := \inf \left\{ \int_0^1 |d\gamma| : \gamma \in C^{1-var}([0,1], \mathbb{R}^d) \quad \text{and } S_N(\gamma)_{0,1} = g \right\}$$

is finite and achieved at some minimizing path γ^*, i.e.

$$\|g\| = \int_0^1 |d\gamma^*| \quad \text{and } S_N(\gamma^*)_{0,1} = g$$

Moreover, this minimizer can (and will) be parametrized to be Lipschitz (i.e. 1-Hölder) continuous and of constant speed, i.e. $|\dot{\gamma}^*(r)| \equiv (const)$ for a.e. $r \in [0,1]$.

Let us now define the important concept of a homogenous norm on $G^N(\mathbb{R}^d)$.

Definition 2.24 (Definition 7.34 of [15])**.** *A homogenous norm is a continuous map $|||\cdot||| : G^N(\mathbb{R}^d) \to \mathbb{R}^+$ which satisfies*

(i) $|||g||| = 0$ if and only if g equals the unit element $1 \in G^N(\mathbb{R}^d)$,

(ii) homogeneity with respect to the dilation operator δ_λ,

$$|||\delta_\lambda(g)||| = |\lambda| \cdot |||g||| \quad \text{for all } \lambda \in \mathbb{R}.$$

A homogenous norm is said to be symmetric if $|||g||| = |||g^{-1}|||$, and sub-additive if $|||g \otimes h||| \leq |||g||| + |||h|||$.

Example 2.1 (Example 7.37 of [15])**.** *The simplest example of a homogenous norm are the maps*

$$g \in G^N(\mathbb{R}^d) \mapsto |||g|||_{\max} := \max_{i=1,\ldots,N} |\pi_i(g)|^{1/i}, \quad \text{and } g \in G^N(\mathbb{R}^d) \mapsto |||g|||_{sum} := \sum_{i=1}^{N} |\pi_i(g)|^{1/i}$$

In general, they are neither symmetric nor sub-additive.

The two maps

$$|||\cdot|||_{\log,\max} : g \in G^N(\mathbb{R}^d) \mapsto \max_{i=1,\ldots,N} |\pi_i(\log(g))|^{1/i}, \quad \text{and } |||\cdot|||_{\log,sum} : g \in G^N(\mathbb{R}^d) \mapsto \sum_{i=1}^{N} |\pi_i(\log(g))|^{1/i}$$

are symmetric homogenous norms.

Remark 2.25. *It is essential to remark that one can write any $g \in G^N(\mathbb{R}^d)$ as $g = \exp(\sum_{i=1,\ldots,N} g^i)$ with g^i made of $i-1$ iterated brackets, according to Definition 2.21. Thus, $\pi_i(\log(g)) = g^i$ for all $i \in \{1, \ldots, N\}$. Therefore*

$$|||g|||_{\log,\max} = \max_{i=1,\ldots,N} |g^i|^{1/i}, \quad \text{and } |||g|||_{\log,sum} = \sum_{i=1}^{N} |g^i|^{1/i}$$

In particular, as mentioned in (9), since $g^2(\mathbb{R}^d) \simeq \mathbb{R}^d \oplus so(d)$ ($so(d)$ being the set of antisymmetric $d \times d$ matrices) we have, for $\boldsymbol{B}_{s,t} := \exp(B_{s,t} + A_{s,t})$

$$|||\boldsymbol{B}_{s,t}|||_{\log,\max} = |B_{s,t}| \vee |A_{s,t}|^{1/2}.$$

Proof. The homogeneity comes from the $^{1/i}$ power on each level's norm. We have

$$|||\delta_\lambda(g)|||_{sum} = \sum_{i=1}^{N} |\lambda^i \pi_i(g)|^{1/i} = |\lambda| \sum_{i=1}^{N} |\pi_i(g)|^{1/i} = |\lambda| |||g|||_{sum},$$

and the proof is the same for the three other norms. □

Proposition 2.26 (Proposition 7.40 of [15])**.** *The Carnot-Caratheodory norm, defined in 2.23 is a homogenous, symmetric, subadditive, continuous norm on $G^N(\mathbb{R}^d)$.*

Definition 2.27 (Definition 7.41 of [15])**.** *The Carnot-Caratheodory norm on $G^N(\mathbb{R}^d)$ induces a genuine (left-invariant, continuous) metric d on $G^N(\mathbb{R}^d)$, called the Carnot–Caratheodory metric.*

Theorem 2.28 (Theorem 7.44 of [15]). *All homogenous norms on $G^N(\mathbb{R}^d)$ are equivalent. More precisely, if $\|\cdot\|_1$ and $\|\cdot\|_2$ are two homogenous norms, there exists $C \geq 1$ such that for all $g \in G^N(\mathbb{R}^d)$, we have*

$$\frac{1}{C}\|g\|_1 \leq \|g\|_2 \leq C\|g\|_1.$$

Theorem 2.29 (Theorem 7.62 of [15]). *Let $N \geq 1$ and $x \in C_o^{1-var}([0,T],\mathbb{R}^d)$. Then $\mathbf{x} = S_N(x)$ is the unique "lift" of x in the sense that $\pi_1(x) = \mathbf{x}$ and such that $\mathbf{x} \in C_o^{1-var}([0,T], G^N(\mathbb{R}^d))$. Moreover,*

$$S_N : C_o^{1-var}([0,T],\mathbb{R}^d) \to C_o^{1-var}([0,T], G^N(\mathbb{R}^d))$$

is a bijection with inverse π_1 and, for all $0 \leq s < t \leq T$,

$$\|\mathbf{x}\|_{1-var;[s,t]} = |x|_{1-var;[s,t]},$$

where $\|\cdot\|_{1-var;[s,t]}$ is the 1-variation with respect to Carnot-Caratheodory metric (induced by the Carnot-Caratheodory norm defined in 2.23).

Definition 2.30 (Homogenous variation and Hölder distances and norms on path spaces). *As a constant reminder that the (Carnot–Caratheodory) metric d on $G^N(\mathbb{R}^d)$ was derived from $\|\cdot\|$, the Carnot-Caratheodory norm, we shall then use the notation*

$$\|\mathbf{x}\|_{p-var;[0,T]} = \sup_{(t_i)\subseteq[0,T]} \left(\sum_i d(\mathbf{x}_{t_i}, \mathbf{x}_{t_{i+1}})^p\right)^{1/p} = \sup_{(t_i)\subseteq[0,T]} \left(\sum_i \|\mathbf{x}_{t_i,t_{i+1}}\|^p\right)^{1/p}$$

and, thanks to homogeneity of $\|\cdot\|$ with respect to dilation, speak of a homogenous p-variation norm.

As a special case of Definition 2.2,

$$C^{p-var}([0,T], G^N(\mathbb{R}^d)) = \{\mathbf{x} \in C([0,T], G^N(\mathbb{R}^d)) : \|\mathbf{x}\|_{p-var;[0,T]} < \infty\}$$

and we shall assume $p \geq 1$ unless otherwise stated.

Similarly, the homogenous 1/p-Hölder norm is given by

$$\|\mathbf{x}\|_{1/p-Hol;[0,T]} = \sup_{0\leq s<t\leq T} \frac{d(\mathbf{x}_s,\mathbf{x}_t)}{|t-s|^{1/p}} = \sup_{0\leq s<t\leq T} \frac{\|\mathbf{x}_{s,t}\|}{|t-s|^{1/p}}$$

and

$$C^{1/p-Hol}([0,T], G^N(\mathbb{R}^d)) = \{\mathbf{x} \in C([0,T], G^N(\mathbb{R}^d)) : \|\mathbf{x}\|_{1/p-Hol;[0,T]} < \infty\}.$$

Given $\mathbf{x},\mathbf{y} \in C([0,T], G^N(\mathbb{R}^d))$ we define

$$d_{p-var;[0,T]}(\mathbf{x},\mathbf{y}) := \left(\sup_D \sum_{t_i\in D} d(\mathbf{x}_{t_i,t_{i+1}}, \mathbf{y}_{t_i,t_{i+1}})^p\right)^{1/p}$$

and

$$d_{1/p-Hol;[0,T]}(\mathbf{x},\mathbf{y}) := \sup_{0\leq s<t\leq T} \frac{d(\mathbf{x}_{s,t},\mathbf{y}_{s,t})}{(t-s)^{1/p}},$$

With $o = 1$, the unit element in $G^N(\mathbb{R}^d)$, we note that $d_{p-var;[0,T]}(\mathbf{x}, o) = \|\mathbf{x}\|_{p-var;[0,T]}$, the homogenous p-variation norm of $\mathbf{x}$ and similarly in the Hölder case.

One can generalise homogenous distances and norms using control functions, see in the Appendix, the subsection 13.

Remark 2.31. *Although the above definition mentions the Carnot-Caratheodory norm and metric, we obviously have similar definitions for the distance norms mentioned in Example 2.1. The induced variation and Hölder norms are equivalent by equivalence of homogenous norms.*

In practice, for usual applications like $G^2(\mathbb{R}^d)$ and the driving continuous path x being a Brownian sample path, like in 2.25 we will use $|||\cdot|||_{\log,sum}$. This yields, for $\boldsymbol{B}_{s,t} := \exp(B_{s,t} + A_{s,t})$

$$\begin{aligned}|||\boldsymbol{B}|||_{1/p-Hol,\log,sum} &= \sup_{0\leq s<t\leq T} \frac{|||\boldsymbol{B}_s^{-1}\otimes\boldsymbol{B}_t|||_{\log,sum}}{(t-s)^{1/p}} = \sup_{0\leq s<t\leq T}\left(\frac{|B_{s,t}|}{(t-s)^{1/p}} + \frac{|A_{s,t}|^{1/2}}{(t-s)^{1/p}}\right)\\ &\sim |B|_{1/p-Hol} + |A|_{2/p-Hol},\end{aligned}$$

where the last line comes from 19.3.

Our definition of homogenous (variation, Hölder, ω-modulus) distance was based on measuring the distance of increments $\mathbf{x}_{s,t}, \mathbf{y}_{s,t} \in G^N(\mathbb{R}^d)$ using the Carnot-Caratheodory distance. Alternatively, recalling that $G^N(\mathbb{R}^d) \subseteq T^N(\mathbb{R}^d)$ we can use the (vector space) norm defined on the latter which leads to the distance of increments given by $|\mathbf{x}_{s,t} - \mathbf{y}_{s,t}|_{T^N(\mathbb{R}^d)} = \max_{k=1}^N |\pi_k(\mathbf{x}_{s,t} - \mathbf{y}_{s,t})|$. Observe that, for $N > 1$, this distance is not homogenous with respect to dilation on $T^N(\mathbb{R}^d)$.

Definition 2.32 (Inhomogenous variation and Hölder distance, Definition 8.6 of [15]). *Given $\mathbf{x},\mathbf{y} \in T^N(\mathbb{R}^d)$ we define*

(i) for $k = 1,\ldots,N$,

$$\rho^{(k)}_{p-var;[0,T]}(\mathbf{x},\mathbf{y}) = \sup_{(t_i)\subseteq[0,T]} \left(\sum_i |\pi_k(x_{t_i,t_{i+1}} - \mathbf{y}_{t_i,t_{i+1}})|^{p/k}\right)^{k/p}$$

and

$$\rho_{p-var;[0,T]}(\mathbf{x},\mathbf{y}) := \sum_{k=1}^N \rho^{(k)}_{p-var;[0,T]}(\mathbf{x},\mathbf{y});$$

(ii) for any control function ω on $[0,T]$, for $k=1,\dots,N$,

$$\rho^{(k)}_{p-\omega;[0,T]}(\mathbf{x},\mathbf{y}) = \sup_{0\le s<t\le T} \frac{|\pi_k(\mathbf{x}_{s,t}-\mathbf{y}_{s,t})|}{\omega(s,t)^{k/p}}$$

and

$$\rho_{p-\omega;[0,T]}(\mathbf{x},\mathbf{y}) = \max_{k=1}^{N} \rho^{(k)}_{p-\omega;[0,T]}(\mathbf{x},\mathbf{y});$$

(iii) As a particular case (with $\omega(s,t)\equiv t-s$) for $k=1,\dots,N$,

$$\rho^{(k)}_{1/p-Hol;[0,T]}(\mathbf{x},\mathbf{y}) = \sup_{0\le s<t\le T} \frac{|\pi_k(\mathbf{x}_{s,t}-\mathbf{y}_{s,t})|}{(t-s)^{k/p}}$$

and

$$\rho_{1/p-Hol;[0,T]}(\mathbf{x},\mathbf{y}) = \max_{k=1}^{N} \rho^{(k)}_{1/p-Hol;[0,T]}(\mathbf{x},\mathbf{y}).$$

2.4. **Geometric rough paths.** Let us denote $[p]$ the integer part of $p\in[0,\infty)$.

When dealing with rough paths, we will generally assume α in $(1/3,1/2]$ unless otherwise stated.

Let us make an essential point about notation of rough path spaces. The two most common notations are the one of [15] and of [12]. We present both of them in this section and give the correspondence between the coexisting notations for geometric rough path spaces. For our purposes, we will have to use both notations depending on the context.

Definition 2.33 (Section 2 of [7]). *A (step-2) α-Hölder rough path on $\mathbb{R}^d$ is a triple $\boldsymbol{X}=(X_0,X,\mathbb{X})\in T^2(\mathbb{R}^d)$, with X_0 point in $\mathbb{R}^d$, $X=(X_{s,t}), s<t$ a two-index $\mathbb{R}^d$-valued map and $\mathbb{X}=(\mathbb{X}_{s,t}), s<t$ two-index $\mathbb{R}^{d\times d}$-valued map (we always suppose $0\le s,t\le T$ when not specified), satisfying the following conditions:*

1. algebraic conditions (Chen's relation): for any $s<u<t$,

$$X_{s,t}=X_{s,u}+X_{u,t} \quad \text{and} \quad \mathbb{X}_{s,t}=\mathbb{X}_{s,u}+\mathbb{X}_{u,t}+X_{s,u}\otimes X_{u,t}; \tag{15}$$

2. analytic conditions:

$$\sup_{0\le s<t\le T}\frac{|X_{s,t}|}{|t-s|^{\alpha}}<\infty \quad \text{and} \quad \sup_{0\le s<t\le T}\frac{|\mathbb{X}_{s,t}|}{|t-s|^{2\alpha}}<\infty. \tag{16}$$

Here X_0 represents the initial condition; it is not included in the standard definition (Definition 2.1 in [12]*), but we need to keep track of it because we will work with paths starting from a generic probability measure (and not just from a single point). However, with some abuse of notation, we will usually write $\boldsymbol{X}=(X,\mathbb{X})$, without X_0, when this is not relevant for our purposes, as for example when the initial point is fixed.*

The space of α-Hölder rough paths on $\mathbb{R}^d$ is denoted by $\mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$. Let us endow $\mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$ with a variant of the homogenous α-Hölder distance defined in Remark 2.31 for a priori non-geometric rough paths, that is instead of looking at the Hölder regularity of the antisymmetric part of the second level, look at the whole second level $\mathbb{X}$. This yields

$$|||\boldsymbol{X}|||_{\alpha} := |X|_{\alpha-Hol}+\sqrt{|\mathbb{X}|_{2\alpha-Hol}}.$$

$\mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$ is not a vector space since the sum of two paths does not respect Chen's relation.

Definition 2.34 (Definition 2.4 of [12]). *Given rough paths $\boldsymbol{X},\boldsymbol{Y}\in\mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$, we define the (inhomogeneous) α-Hölder rough path metric*

$$\rho_{\alpha-Hol}(\boldsymbol{X},\boldsymbol{Y}) := |X-Y|_{\alpha-Hol}+|\mathbb{X}-\mathbb{Y}|_{2\alpha-Hol}.$$

This is similar to the definition of the inhomogenous norm in 2.32 with a sum instead of a maximum.

Remark 2.35. *While* (15) *does capture the most basic (additivity) property that one expects any decent theory of integration to respect, it does not imply any form of integration by parts/chain rule. Then for any pair e_i^*, e_j^* of elements in $(\mathbb{R}^d)^*$, writing $X_t^i=e_i^*(X_t)$ and $\mathbb{X}^{ij}_{s,t}=(e_i^*\otimes e_j^*)(\mathbb{X}_{s,t})$, by Stratonovitch Itô formula or integration by parts for Riemann-Stieltjes integrals 2.9 one has*

$$\begin{aligned}\int_s^t X^i_{s,r}dX^j_r+\int_s^t X^j_{s,r}dX^i_r &= \int_s^t d(X^iX^j)_r - X^i_sX^j_{s,t}-X^j_sX^i_{s,t}\\ &= (X^iX^j)_{s,t}X^i_sX^j_{s,t}-X^j_sX^i_{s,t} = X^i_{s,t}\otimes X^j_{s,t},\end{aligned}$$

Now, if one looks for a first order calculus setting, such as is valid in the context of smooth paths or the Stratonovitch stochastic calculus, one would expect to have $\mathbb{X}^{ij}_{s,t}+\mathbb{X}^{ij}_{s,t}=X^i_{s,t}\otimes X^j_{s,t}$ such that the symmetric part of $\mathbb{X}$ is determined by X. In other words, for all times s,t we would like to have the "first order calculus" condition

$$Sym(\mathbb{X}_{s,t})=\frac{1}{2}X_{s,t}\otimes X_{s,t}. \tag{17}$$

However, if we take X to be a multidimensional Brownian path and define $\mathbb{X}$ by Itô integration, then (16) *still holds, but* (17) *certainly does not. The Itô enhanced Brownian motion is in $\mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$ but it is not a geometric rough path.*

There are two natural ways to define a set of "geometric" rough paths for which (17) *holds. On the one hand, we can define the space of weakly geometric (α-Hölder) rough paths*

$$\mathcal{C}^{\alpha}_g([0,T],\mathbb{R}^d)\subseteq\mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$$

by stipulating that $(X,\mathbb{X}) \in \mathcal{C}_g^\alpha([0,T],\mathbb{R}^d)$ *if and only if* $(X,\mathbb{X}) \in \mathcal{C}^\alpha([0,T],\mathbb{R}^d)$ *and* (17) *holds as equality in* $\mathbb{R}^d \otimes \mathbb{R}^d$*, for every* $s,t \in [0,T]$*. Note that* $\mathcal{C}_g^\alpha([0,T],\mathbb{R}^d)$ *is a closed subset of* $\mathcal{C}^\alpha([0,T],\mathbb{R}^d)$*.*

On the other hand, we have already seen in 2.29 that every smooth (bounded variation continuous) path can be lifted canonically to an element in $C_o^{1-var}([0,T],G^N(\mathbb{R}^d))$*. This choice of* X *then obviously satisfies* (17) *and we can define the space of geometric (α-Hölder) rough paths ,*

$$\mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^d) \subseteq \mathcal{C}^\alpha([0,T],\mathbb{R}^d),$$

as the closure of $C_o^{1-var}([0,T],G^N(\mathbb{R}^d))$ *in* $\mathcal{C}^\alpha([0,T],\mathbb{R}^d)$ *with respect to the homogenous distance.*

Definition 2.36 (Definition 9.15 of [15]). *(i) A weak geometric p-rough path is a continuous path of finite p-variation with values in the free nilpotent group of step $[p]$ over $\mathbb{R}^d$, i.e. an element of* $C^{p-var}([0,T],G^{[p]}(\mathbb{R}^d))$*.*

(ii) A geometric p-rough path is a continuous path with values in the free nilpotent group of step $[p]$ over $\mathbb{R}^d$ which is in the p-variation closure of the set of bounded variation paths, i.e. an element of $C^{0,p-var}([0,T],G^{[p]}(\mathbb{R}^d))$*.*

(iii) A weak geometric $1/p$-Hölder rough path is a $1/p$-Hölder path with values in the free nilpotent group of step $[p]$ over $\mathbb{R}^d$, i.e. an element of $C^{1/p-Hol}([0,T],G^{[p]}(\mathbb{R}^d))$*.*

(iv) A geometric $1/p$-Hölder rough path is a continuous path with values in the free nilpotent group of step $[p]$ over $\mathbb{R}^d$ which is in the $1/p$-Hölder closure of the set of 1-Hölder paths, i.e. an element of $C^{0,1/p-Hol}([0,T],G^{[p]}(\mathbb{R}^d))$*.*

Recall from the interpolation results of the previous chapter that

$$C^{(p+\epsilon)-var}([0,T],G^{[p]}(\mathbb{R}^d)) \subseteq C^{0,p-var}([0,T],G^{[p]}(\mathbb{R}^d)) \subseteq C^{p-var}([0,T],G^{[p]}(\mathbb{R}^d))$$

and for this reason the difference between weak and genuine geometric p-rough paths is important only when we care about very precise results.

2.5. **Rough differential equations.** This subsection is mostly inspired from Section 10 of [15].

Definition 2.37 (Definition 3.3 of [15]). *Given a collection of continuous vector fields* $V=(V_1,...,V_d)$ *on* $\mathbb{R}^e$ *, a driving signal* $x \in C^{1-var}([0,T],\mathbb{R}^d)$*, and an initial condition* $y_0 \in \mathbb{R}^e$ *, we write* $\pi_{(V)}(0,y_0;x)$ *for the set of all solutions to the ODE*

$$dy_t = \sum_{i=1}^d V_i(y_t)dx_t^i \equiv V(y_t)dx_t \tag{18}$$

for $t \in [0,T]$ *started at* y_0*. The above ODE is understood as a Riemann- Stieltjes integral equation, i.e.* $y_{0,t} := y_t - y_0 = \int_0^t V(y_s)dx_s$ *. In case of uniqueness* $y = \pi_{(V)}(0,y_0;x)$ *denotes <u>the</u> solution. If necessary,* $\pi(0,y_0;x)$ *is only considered up to some explosion time. Similarly,* $\pi_{(V)}(s,y_s;x)$ *stands for solutions of* (18) *started at time s from a point* $y_s \in \mathbb{R}^e$ *.*

As explained in the beginning of page 213 of [15], To prepare for the following definition, given a real $\gamma \in (0,\infty)$, we agree that $\lfloor\gamma\rfloor$ is the largest integer strictly smaller than γ so that $\gamma = \lfloor\gamma\rfloor + \{\gamma\}$ with $\lfloor\gamma\rfloor \in \mathbb{N}$ and $\{\gamma\} \in (0,1]$.

Definition 2.38 (Lipschitz map, Definition 10.2 of [15]). *A map* $V : E \to F$ *between two normed spaces* E,F *is called γ-Lipschitz (in the sense of E. Stein), in symbols* $V \in Lip^\gamma(E,F)$ *or simply* $V \in Lip^\gamma(E)$ *if* $E=F$*, if V is* $\lfloor\gamma\rfloor$ *times continuously differentiable and such that there exists a constant* $0 \le M < \infty$ *such that the supremum norm of its kth derivatives,* $k=0,\dots,\lfloor\gamma\rfloor$*, and the* $\{\gamma\}$*-Hölder norm of its* $\lfloor\gamma\rfloor$*th derivative are bounded by M. The smallest M satisfying the above conditions is the γ-Lipschitz norm of V, denoted* $|V|_{Lip^\gamma}$*.*

Remark 2.39. *It should be noted that* Lip^N *maps have* $(N-1)$ *bounded derivatives, with the* $(N-1)$*th derivative being Lipschitz, but need not be N times continuously differentiable.*

Davie's lemma (Lemma 10.7 of [15]) is the main result of allowing to deal with RDEs. It is a p-variation estimate of an ODE solution in terms of the p-variation of the driving signal. Its proof is based on two types of lemmas, say type A and B. Here is the idea behind each of these two types of lemmas. They are quantitative estimates of:

(A) the difference of ODE solutions started at the same point, with different driving signals (but with common iterated integrals up to a given order)

(B) the difference of ODE solutions started at different points but with identical driving signals.

Davie's lemma gives uniform estimates for ODE solutions which depend only on the rough path regularity (p-variation or $1/p$-Hölder) of the canonical lift of a "nice" driving signal $x \in C^{1-var}([0,T],\mathbb{R}^d)$. Unsurprisingly, a careful passage to the limit will yield a sensible notion of differential equations driven by a "generalized" driving signal, given as a limit of nice driving signals (in p-variation or $1/p$-Hölder rough path sense). This class of generalized driving signals is precisely the class of weak geometric p-rough paths introduced previously. Indeed, we saw we can deduce from Section 8.2 of [15] that for any $x \in C^{p-var}([0,T],G^{[p]}(\mathbb{R}^d)$, there exist $(x_n) \subseteq C^{1-var}([0,T],\mathbb{R}^d)$ which approximate $\mathbf{x}$ uniformly with uniform p-variation bounds,

$$\lim_{n\to\infty} d_{0;[0,T]}(S_{[p]}(x_n),\mathbf{x}) = 0 \text{ and } \sup_n \|S_{[p]}(x_n)\|_{p-var;[0,T]} < \infty. \tag{19}$$

The proof of the statement right above is given in the Appendix 15.

Definition 2.40 (RDE solution, Definition 10.17 of [15]). *Let $\mathbf{x} \in C^{p-var}([0,T], G^{[p]}(\mathbb{R}^d))$ be a weak geometric p-rough path. We say that $y \in C([0,T], \mathbb{R}^e)$ is a solution to the rough differential equation (RDE solution) driven by $\mathbf{x}$ along the collection of $\mathbb{R}^e$-vector fields $(V_i)_{i=1,\dots,d}$ and started at $y_0 \in \mathbb{R}^e$ if there exists a sequence $(x^n)_n$ in $C^{1-var}([0,T], \mathbb{R}^d)$ such that (19) holds, and ODE solutions $y_n \in \pi_{(V)}(0, \pi_1(\mathbf{y}_0); x^n)$ such that*

$$y_n \text{ converges uniformly to } y \text{ when } n \to \infty.$$

The formal equation

$$dy = V(y)d\mathbf{x} \tag{20}$$

is referred to as an RDE.

Definition 2.41 (RDE, Lyons definition, Remark 10.19 of [15]). *As one expects, RDE solutions can also be defined as the solution to a "rough" integral equation and this is Lyons' original approach* [34]. *To this end, one first needs a notion of rough integration (cf. Definition 2.43) which allows, for sufficiently smooth $\varphi = (\varphi_1, \dots, \varphi_d)$, defined on $\mathbb{R}^d$, the definition of an (indefinite) rough integral*

$$\int_0^{\cdot} \varphi(z)d\mathbf{z} \quad \text{with } z = \pi_1(\mathbf{z})$$

such that, in the case when $\mathbf{z} = S_{[p]}(z)$ for some $z \in C^{1-var}([0,T], \mathbb{R}^d)$, it coincides with $S_{[p]}(\xi)$ where ξ is the classical Riemann–Stieltjes integral $\int_0^{\cdot} \varphi(z)dz$. Note that (20) *cannot be rewritten as an integral equation of the above form (for y is not part of the integrating signal x). Nonetheless, the "enhanced" differential equation (in which the input signal is carried along to the output)*

$$dx = dx$$
$$dy = V(y)dx$$

can be written in the desired form $z_{0,\cdot} = \int_0^{\cdot} \varphi(z)dz$ provided we set $z = (x, y)$ and

$$\varphi(z) = \begin{pmatrix} 1 & 0 \\ V(y) & 0 \end{pmatrix}.$$

The above integral equation indeed makes sense as a rough integral equation (with implicit regularity assumption on V so the rough integral is well-defined) replacing z by a genuine geometric p-rough path $\mathbf{z} \in C^{p-var}([0,T], G^{[p]}(\mathbb{R}^d \oplus \mathbb{R}^e))$ and solutions can be constructed, for instance, by a Picard- iteration [34]. *An $\mathbb{R}^e$-valued solution is then recovered by projection $\mathbf{z} \mapsto \pi_1(\mathbf{z}) = z = (x,y) \mapsto y$ and again one can see that an RDE solution in the sense of Definition 2.40 is also a solution in this sense.*

The RDEs we considered in 2.40 map weak geometric p-rough paths to $\mathbb{R}^e$-valued paths of bounded variation. We shall now see that one can construct a "full" solution as a weak geometric p-rough path in its own right. This allows to use a solution to a first RDE to be a driving signal for a second RDE. Relatedly, RDE solutions can then be used (as integrators) in rough integrals, cf. 2.43 below. Let us remark that in Lyons' original work, existence and uniqueness were established by Picard iteration and so it was a necessity to work with full RDE solutions.

Definition 2.42 (Full RDE solution, Definition 10.34 of [15]). *Let $\mathbf{x} \in C^{p-var}([0,T], G^{[p]}(\mathbb{R}^d))$ be a weak geometric p-rough path. We say that $\mathbf{y} \in C([0,T], G^{[p]}(\mathbb{R}^e))$ is a solution to the full rough differential equation (full RDE solution) driven by $\mathbf{x}$ along the vector fields $(V_i)_i$ and started at $\mathbf{y}_0 \in G^{[p]}(\mathbb{R}^e)$ if there exists a sequence $(x^n)_n$ in $C^{1-var}([0,T], \mathbb{R}^d)$ such that (19) holds, and ODE solutions $y_n \in \pi_{(V)}(0, \pi_1(\mathbf{y}_0); x^n)$ such that*

$$\mathbf{y}_0 \otimes (S_{[p]}(y_n)) \text{ converges uniformly (in } d_\infty) \text{ to } \mathbf{y} \text{ when } n \to \infty.$$

The formal equation $d\mathbf{y} = V(y)d\mathbf{x}$ is referred to as a full RDE.

With our main interest in rough differential equation we constructed RDEs directly as limits of ODEs. In the same spirit, we now define "rough integrals" as limits of Riemann-Stieltjes integration. Given the work already done, we can take a short-cut and derive existence, uniqueness and continuity properties quickly from the previous RDE results.

Definition 2.43 (Rough integrals, Definition 10.44 of [15]). *Let $\mathbf{x} \in C^{p-var}([0,T], G^{[p]}(\mathbb{R}^d))$ be a geometric p-rough path, and $\varphi = (\varphi)_{i=1,\dots,d}$ a collection of maps from $\mathbb{R}^d$ to $\mathbb{R}^e$ (thus φ is a map from $\mathbb{R}^d$ to $L(\mathbb{R}^d, \mathbb{R}^e)$). We say that $\mathbf{y}$ is a rough integral of φ along $\mathbf{x}$, if there exists a sequence (x_n) in $C^{1-var}([0,T], G^{[p]}(\mathbb{R}^d))$ such that*

$$\forall n, x_0^n = \pi_1(\mathbf{x}_0) \text{ (initial condition)}$$
$$\lim_{n\to\infty} d_{0;[0,T]}(S_{[p]}(x^n), \mathbf{x}) = 0 \text{ (convergence of } S_{[p]}(x^n) \text{ to } \mathbf{x} \text{ in } d_{0;[0,T]} \text{ distance)}$$
$$\sup_n \|S_{[p]}(x^n)\|_{p-var;[0,T]} < \infty \text{ (uniform boundedness of } S_{[p]}(x^n) \text{ in p-var)}$$

and

$$\lim_{n\to\infty} d_\infty\Big(S_{[p]}\Big(\int_0^{\cdot} \varphi(x_u^n)dx_u^n\Big), \mathbf{y}\Big) = 0 \text{ (convergence of } \int_0^{\cdot} \varphi(x_u^n)dx_u^n \text{ to } \mathbf{y} \text{ in } d_\infty \text{ distance)}$$

We will write $\int \varphi(x)dx$ for the set of rough path integrals of φ along $\mathbf{x}$. If this set is a singleton it will denote the rough path integral of φ along $\mathbf{x}$.

The theorem giving existence and uniqueness/continuity of the rough path integral, and therefore of the Itô-Lyons map in Lyons' sense is the following. Remember $\rho_{p-\omega}$ is the inhomogenous $p-\omega$ distance between rough paths defined in 2.32.

Theorem 2.44 (Theorem 10.47 of [15]). *Existence: Assume that*

(i) $\varphi=(\varphi_j)_{j=1,\dots,d}$ *is a collection of* $Lip^{\gamma-1}(\mathbb{R}^d,\mathbb{R}^e)$*-maps where* $\gamma>p\geq 1$;

(ii) $\mathbf{x}$ *is a geometric p-rough path in* $C^{p-var}([0,T],G^{[p]}(\mathbb{R}^d))$.

Then, for all $s<t\in[0,T]$, *there exists a unique rough-path integral of* φ *along* $\mathbf{x}$. *The indefinite integral* $\int\varphi(x)d\mathbf{x}$ *is a geometric rough path: there exists a constant* C_1 *depending only on p and* γ *such that, for all* $s<t\in[0,T]$

$$\|\int\varphi(x)d\mathbf{x}\|_{p-var;[s,t]}\leq C_1\|\varphi\|_{Lip^{\gamma-1}}(\|\mathbf{x}\|_{p-var;[s,t]}\vee\|\mathbf{x}\|^p_{p-var;[s,t]}).$$

Also, if $x^{s,t}:[s,t]\to\mathbb{R}^d$ *is any continuous bounded variation path such that*

$$x^{s,t}_s=\pi_1(\mathbf{x}_s),\ S_{\lfloor\gamma\rfloor}(\mathbf{x})_{s,t}\text{ and }\int_s^t|dx^{s,t}_u|\leq K\|\mathbf{x}\|_{p-var;[s,t]}$$

for some constant K, we have for all $s<t\in[0,T]$ *with* $K\|\mathbf{x}\|_{p-var;[s,t]}<1$, *and all* $k\in\{1,\dots,\lfloor\gamma\rfloor\}$,

$$\Big|\pi_k\Big\{\Big(\int\varphi(x)d\mathbf{x}\Big)_{s,t}-S_{\lfloor\gamma\rfloor}\Big(\int\varphi(x^{s,t}_u)dx^{s,t}_u\Big)_{s,t}\Big\}\Big|\leq C_2\|\varphi\|^k_{Lip^{\gamma-1}}(K\|\mathbf{x}\|_{p-var;[s,t]})^{\gamma+k-1}$$

where C_2 *depends on p and* γ.

Uniqueness, continuity: There exists a unique element in $\int\varphi(x)d\mathbf{x}$. *More precisely, if* ω *is a fixed control,*

$$\max_{i=1,2}\{|\varphi^i|_{Lip^{\gamma-1}},\|\mathbf{x}^i\|_{p-\omega;[0,T]}\}<R$$

and

$$\epsilon=|x^1_0-x^2_0|+\rho_{p-\omega;[0,T]}(\mathbf{x}^1,\mathbf{x}^2)+|\varphi^2-\varphi^1|_{Lip^{\gamma-1}},$$

then for some constant $\beta=\beta(\gamma,p)>0$ *and* $C=C(R,\gamma,p)\geq 0$,

$$\rho_{p-\omega;[0,T]}\Big(\int\varphi^1(x^1)d\mathbf{x}^1,\int\varphi^2(x^2)d\mathbf{x}^2\Big)\leq C\epsilon^\beta$$

Thanks to Theorem 8.10 of [15] we can "locally uniformly" switch from the inhomogenous path-space metrics $(\rho_{p-\omega},\rho_{p-var})$ to the homogenous ones $(d_{p-\omega},d_{p-var})$. In fact, we can state the following.

Corollary 2.44.1 (Corollary 10.48 of [15]). *Let* $\varphi_i:\mathbb{R}^d\to\mathbb{R}^e, i=1,\dots,d$ *be some* $(\gamma-1)$*-Lipschitz maps where* $\gamma>p$. *For any fixed control* $\omega, R>0$ *and* $p\leq p'$ *, the maps*

$$Lip^{\gamma-1}\times(\{\|\mathbf{x}\|_{p-\omega}\leq R\},d_{p'-\omega})\to\Big(C^{p-\omega}([0,T],G^{[p]}(\mathbb{R}^e)),d_{p'-\omega}\Big)$$
$$(\varphi,\mathbf{x})\mapsto\int\varphi(\mathbf{x})d\mathbf{x}$$

and

$$Lip^{\gamma-1}\times(\{\|\mathbf{x}\|_{p-var}\leq R\},d_{p'-var})\to\Big(C^{p-var}([0,T],G^{[p]}(\mathbb{R}^e)),d_{p'-var}\Big)$$
$$(\varphi,\mathbf{x})\mapsto\int\varphi(\mathbf{x})d\mathbf{x}$$

and

$$Lip^{\gamma-1}\times(\{\|\mathbf{x}\|_{p-var}\leq R\},d_\infty)\to\Big(C^{p-var}([0,T],G^{[p]}(\mathbb{R}^e)),d_\infty\Big)$$
$$(\varphi,\mathbf{x})\mapsto\int\varphi(\mathbf{x})d\mathbf{x}$$

are uniformly continuous.

Remark 2.45. *In particular, if* $\varphi^1=\varphi^2$ *and* $x^1_0=x^2_0$, *we end up with the same notion of continuity of the rough path integral as in (ii) of Theorem 2.5 of* [17], *that is local Lipschitzness (in the sense of being Lipschitz on bounded sets).*

Like in [17], *continuity gives uniqueness here: if* $\varphi^1=\varphi^2$ *and* $\mathbf{x}^1=\mathbf{x}^2$, *then* $\epsilon=0$ *and*

$$\rho_{p-\omega;[0,T]}\Big(\int\varphi^1(x^1)d\mathbf{x}^1,\int\varphi^2(x^2)d\mathbf{x}^2\Big)=0,$$

which means, according to Definition 8.6 of [15]; *that, for all* $k=1,\dots,N$,

$$\sup_{0\leq s<t\leq T}|\pi_k(\mathbf{x}^1_{s,t}-\mathbf{x}^2_{s,t})|=0.$$

2.6. **Gubinelli's controlled rough path theory.**

Definition 2.46 (Gubinelli derivative, Definition 4.6 of [12]). *Given a path* $X\in C^{\alpha-Hol}([0,T],\mathbb{R}^d)$, *we say that* $Y\in C^{\alpha-Hol}([0,T],L(\mathbb{R}^d,\mathbb{R}^e))$ *is controlled by X if there exists* $Y'\in C^{\alpha-Hol}\Big([0,T],L(\mathbb{R}^d,L(\mathbb{R}^d,\mathbb{R}^e))\Big)$ *so that the remainder term* R^Y *given implicitly through the relation*

$$Y_{s,t}=Y'_sX_{s,t}+R^Y_{s,t}\tag{21}$$

satisfies $|R^Y|_{2\alpha-Hol}<\infty$. *This defines the space of controlled rough paths,*

$$(Y,Y')\in\mathcal{D}^{2\alpha}_X([0,T],L(\mathbb{R}^d,\mathbb{R}^e)).$$

Although Y' *is not, in general, uniquely determined from Y (cf. Remark 4.7 and Section 6 of* [12]*) we call any such* Y' *the Gubinelli derivative of Y (with respect to X).*

We endow the space $\mathcal{D}_X^{2\alpha}([0,T],L(\mathbb{R}^d,\mathbb{R}^e))$ *with the seminorm*

$$\|Y,Y'\|_{X,2\alpha-Hol} := |Y'|_{\alpha-Hol} + |R^Y|_{2\alpha-Hol}.$$

As in the case of classical Hölder spaces, $\mathcal{D}_X^{2\alpha}([0,T],L(\mathbb{R}^d,\mathbb{R}^e))$ *is a Banach space under the norm* $(Y,Y') \mapsto |Y_0| + |Y'_0| + \|Y,Y'\|_{X,2\alpha-Hol}$. *This quantity also controls the* α*-Hölder regularity of* Y.

Let us define the rough integral of Y against X in terms of compensated Riemann sums as

$$\int_0^1 Y d\mathbf{X} := \lim_{|D|\to 0} \sum_{[t_i,t_{i+1}]\in D} Y_{t_i} X_{t_i,t_{i+1}} + Y'_{t_i}\mathbb{X}_{t_i,t_{i+1}} \tag{22}$$

Theorem 2.47 (Gubinelli, Theorem 4.10 of [12])**.** *Let* $T>0$*, let* $\boldsymbol{X} = (X,\mathbb{X}) \in \mathcal{C}^\alpha([0,T],\mathbb{R}^d)$ *(defined in 2.33) for some* $\alpha \in \left(\frac{1}{3},\frac{1}{2}\right]$*, and let* $(Y,Y') \in \mathcal{D}_X^{2\alpha}([0,T],L(\mathbb{R}^d,\mathbb{R}^e))$*. Then, there exists a constant* C *depending only on* α *such that*

(i) The integral defined in (22) *exists and, for every pair* s,t*, one has the bound*

$$\left|\int_s^t Y_r d\boldsymbol{X}_r - Y_s X_{s,t} - Y'_s\mathbb{X}_{s,t}\right| \le C\left(|X|_{\alpha-Hol}|R^Y|_{2\alpha-Hol} + |\mathbb{X}|_{2\alpha-Hol}|Y'|_{\alpha-Hol}\right)|t-s|^{3\alpha}.$$

(ii) The map from $\mathcal{D}_X^{2\alpha}([0,T],L(\mathbb{R}^d,\mathbb{R}^e))$ *to* $\mathcal{D}_X^{2\alpha}([0,T],\mathbb{R}^e))$ *given by*

$$(Y,Y') \mapsto (Z,Z') := \left(\int_0^\cdot Y_t d\boldsymbol{X}_t, Y\right),$$

is a continuous linear map and

$$\|Z,Z'\|_{X,2\alpha-Hol} \le |Y|_{\alpha-Hol} + |Y'|_\infty|\mathbb{X}|_{2\alpha-Hol} + CT^\alpha(|X|_{\alpha-Hol}|R^Y|_{2\alpha-Hol} + |\mathbb{X}|_{2\alpha-Hol}|Y'|_{\alpha-Hol})$$

It seems natural in the light of the previous subsections that if $\alpha \le 1/3$, a controlled rough path should have a kind of "Taylor expansion" up to order $N\alpha$.

Definition 2.48 (Definition 4.18 of [12])**.** *Let* $\alpha \in (0,1)$*, let* $N = \lfloor 1/\alpha \rfloor$*, and let* $\boldsymbol{X}$ *be a geometric* α*-Hölder rough path as defined in 2.36. A controlled rough path is a* $T^{N-1}(\mathbb{R}^d)^*$*-valued function* $\boldsymbol{Y}$ *such that, for every word* w *with* $|w| \le N-1$*, one has the bound*

$$|\langle e_w, \mathbf{y}_t\rangle - \langle \mathbb{X}_{s,t}\otimes e_w, \mathbf{y}_s\rangle| \le C|t-s|^{(N-|w|)\alpha},$$

where, given a word $w = w_1\dots w_k$ *with letters in* $1,\dots,d$*, we write* $e_w = e_1\otimes\cdots\otimes e_k$ *for the corresponding basis vector of* $T^N(\mathbb{R}^d)$*. We then identify the words themselves as the dual basis of* $T^N(\mathbb{R}^d)^*$*. Note that* $e_\emptyset = 1 \in \mathbb{R} \simeq (\mathbb{R}^d)^{\otimes 0} \subseteq T^N(\mathbb{R}^d)$.

The main theorem about RDEs in [12] is formulated using controlled rough paths.

Theorem 2.49 (Theorem 8.3 of [12])**.** *Given* $\xi \in \mathbb{R}^e, f \in C^3(W,L(\mathbb{R}^d,\mathbb{R}^e))$ *and a rough path* $\boldsymbol{X} = (X,\mathbb{X}) \in C^\beta([0,T],\mathbb{R}^d)$ *with* $\beta \in \left(\frac{1}{3},\frac{1}{2}\right)$*, there exists* $0 < T_0 \le T$ *and a unique element* $(Y,Y') \in D_X^{2\beta}([0,T_0],\mathbb{R}^e)$*, with* $Y' = f(Y)$*, such that, for all* $0 \le t \le T_0$*,*

$$Y_t = \xi + \int_0^t f(Y_s)d\boldsymbol{X}_s.$$

Here, the integral is interpreted in the sense of Theorem 2.47 and $f(Y) \in D_X^{2\beta}$ *is built from* Y *by Lemma 7.3 of* [12]*. Moreover, if* f *is linear or* $f \in C_b^3$*, we may take* $T_0 = T$*, and thus global existence holds on* $[0,T]$.

There should be a generalisation of Theorem 2.49 to controlled rough paths of lower regularity.

Even the continuity of the Itô-Lyons map (Theorem 8.5 of [12]) is given in terms of Theorem 2.49, which is a sign of the importance and generality of the notion of controlled rough paths.

Remark 2.50. *In* [12]*, the large deviations are only stated for Itô and Stratonovitch Brownian rough paths in Theorem 9.5, they are not stated for Gaussian rough paths in general.*

Their proof is based on Theorem 9.1 of [12]*, which proves uniqueness of a controlled RDE solution* $(Y,f(Y)) \in \mathcal{D}_{B(\omega)}^{2\alpha}$.

This shows that, even the essential corollaries that are large deviations statements are based on controlled rough paths theorems in [12].

3. Large Deviations on SDEs using rough paths

3.1. **Introducing Large Deviations Principles.** Most of this section is taken from [6].

The large deviation principle (LDP) characterizes the limiting behavior, as $\epsilon \to 0$, of a family of probability measures $\{\mu_\epsilon\}$ on $(\mathcal{X},\mathcal{B})$ in terms of a rate function. This characterization is via asymptotic upper and lower exponential bounds on the values that μ_ϵ assigns to measurable subsets of $\mathcal{X}$. Throughout, $\mathcal{X}$ is a topological space so that open and closed subsets of $\mathcal{X}$ are well-defined, and the simplest situation is when elements of $\mathcal{B}_\mathcal{X}$, the Borel σ-field on $\mathcal{X}$, are of interest.

Definition 3.1 (Rate function, Section 1.2 of [6])**.** *A rate function* I *is a lower semicontinuous mapping* $I : \mathcal{X} \to [0,\infty]$ *(such that for all* $\alpha \in [0,\infty)$*, the level set* $\psi_I(\alpha) := \{x : I(x) \le \alpha\}$ *is a closed subset of* $\mathcal{X}$*). A good rate function is a rate function for which all the level sets* $\psi_I(\alpha)$ *are compact subsets of* $\mathcal{X}$*. The effective domain of* I*, denoted* $\mathcal{D}_I$*, is the set of points in* $\mathcal{X}$ *of finite rate, namely,* $\mathcal{D}_I := \{x : I(x) < \infty\}$*. When no confusion occurs, we refer to* $\mathcal{D}_I$ *as the domain of* I*.*

The following standard notation is used throughout this report. For any set Γ, $\bar{\Gamma}$ denotes the closure of Γ, Γ^o the interior of Γ, and Γ^c the complement of Γ. The infimum of a function over an empty set is interpreted as $+\infty$.

Definition 3.2 (LDP, Section 1.2 of [6]). *μ_ϵ satisfies the large deviation principle (or full large deviation principle) with a rate function I if, for all $\Gamma \in \mathcal{B}$,*

$$-\inf_{x\in\Gamma^o} I(x) \leq \liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq \limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(\Gamma) \leq -\inf_{x\in\bar{\Gamma}} I(x). \tag{23}$$

The right- and left-hand sides of (23) are referred to as the large deviation upper and lower bounds, respectively.

Remark 3.3. *Sometimes it is practical to parametrize the family of probability measures so as to consider $\epsilon^2 \log \mu_\epsilon(\Gamma)$.*

Definition 3.4 (Exponential tightness, Section 1.2 of [6]). *Suppose that all the compact subsets of $\mathcal{X}$ belong to $\mathcal{B}$. A family of probability measures $\{\mu_\epsilon\}$ on $\mathcal{X}$ is exponentially tight if for every $\alpha < \infty$, there exists a compact set $K_\alpha \subseteq \mathcal{X}$ such that*

$$\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(K_\alpha^c) < -\alpha$$

Let us now depict, following Section 4.2 of [6] transformations that preserve the LDP, although, possibly, changing the rate function. Once the LDP with a good rate function is established for μ_ϵ, the basic contraction principle yields the LDP for $\mu_\epsilon \circ f^{-1}$, where f is any continuous map. The inverse contraction principle deals with f which is the inverse of a continuous bijection, and this is a useful tool for strengthening the topology under which the LDP holds. Eventually, exponentially good approximations have essential implications; for example, it is shown that when two families of measures defined on the same probability space are exponentially equivalent, then one can infer the LDP for one family from the other. A direct consequence is Theorem 4.2.23, which extends the contraction principle to "approximately continuous" maps.

Theorem 3.5 (Contraction principle, Theorem 4.2.1 of [6]). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Consider a good rate function $I : \mathcal{X} \to [0,\infty]$.*

(a) For each $y \in \mathcal{Y}$, define

$$I'(y) := \inf\{I(x) : x \in \mathcal{X}, y = f(x)\}.$$

Then I' is a good rate function on $\mathcal{Y}$, where as usual the infimum over the empty set is taken as ∞.

(b) If I controls the LDP associated with a family of probability measures $\{\mu_\epsilon\}$ on $\mathcal{X}$, then I' controls the LDP associated with the family of probability measures $\{\mu_\epsilon \circ f^{-1}\}$ on $\mathcal{Y}$.

Theorem 3.6 (Inverse contraction principle, Theorem 4.2.4 of [6]). *Let $\mathcal{X}$ and $\mathcal{Y}$ be Hausdorff topological spaces. Suppose that $g : \mathcal{Y} \to \mathcal{X}$ is a continuous bijection, and that $\{\nu_\epsilon\}$ is an exponentially tight family of probability measures on $\mathcal{Y}$. If $\{\nu_\epsilon^{-1}\}$ satisfies the LDP with the rate function$I : \mathcal{X} \to [0,\infty]$, then $\{\nu_\epsilon\}$ satisfies the LDP with the good rate function $I'(\cdot) := I(g(\cdot))$.*

In order to extend the contraction principle beyond the continuous case, it is obvious that one should consider approximations by continuous functions. Let us define the notion of exponential approximations.

Theorem 3.7 (Definition 4.2.14 of [6]). *Let $(\mathcal{Y}, d)$ be a metric space and, for each $\delta > 0$, define*

$$\Gamma_\delta := \{(\tilde{y}, y) : d(\tilde{y}, y) > \delta\} \subseteq \mathcal{Y} \times \mathcal{Y}.$$

For each $\epsilon > 0$ and all $m \in \mathbb{N}$, let $(\Omega, \mathcal{B}_\epsilon, P_{\epsilon,m})$ be a probability space, and let the $\mathcal{Y}$-valued random variables $\tilde{Z}_\epsilon$ and $Z_{\epsilon,m}$ be distributed according to the joint law $P_{\epsilon,m}$, with marginals $\tilde{\mu}_\epsilon$ and $\mu_{\epsilon,m}$, respectively. $\{Z_{\epsilon,m}\}$ are called exponentially good approximations of $\tilde{Z}_\epsilon$ if, for every $\delta > 0$, the set $\{\omega : (\tilde{Z}_\epsilon, Z_{\epsilon,m}) \in \Gamma_\delta\}$ is $\mathcal{B}_\epsilon$-measurable and

$$\lim_{m\to\infty} \limsup_{\epsilon\to 0} \epsilon \log P_{\epsilon,m}(\Gamma_\delta) = -\infty.$$

Similarly, the measures $\{\mu_{\epsilon,m}\}$ are exponentially good approximations of $\tilde{\mu}_\epsilon$ if one can construct probability spaces $(\Omega, \mathcal{B}_\epsilon, P_{\epsilon,m})$ as above.

The so-called "extended contraction principle" is the following extension of the contraction principle to maps that are not continuous, but that can be approximated well by continuous maps.

Theorem 3.8 (Extended contraction principle, Theorem 4.2.23 of [6]). *Let $\{\mu_\epsilon\}$ be a family of probability measures that satisfies the LDP with a good rate function I on a Hausdorff topological space $\mathcal{X}$, and for $m \in \mathbb{N}$ let $f_m : \mathcal{X} \to \mathcal{Y}$ be continuous functions, with $(\mathcal{Y}, d)$ a metric space. Assume there exists a measurable map $f : \mathcal{X} \to \mathcal{Y}$ such that for every $\alpha < \infty$,*

$$\limsup_{m\to\infty} \sup_{\{x : I(x)\leq\alpha\}} d(f_m(x), f(x)) = 0.$$

Then any family of probability measures $\{\tilde{\mu}_\epsilon\}$ for which $\{\mu_\epsilon \circ f_m^{-1}\}$ are exponentially good approximations satisfies the LDP in $\mathcal{Y}$ with the good rate function $I'(y) = \inf\{I(x) : y = f(x)\}$.

3.2. **Some famous types of LDPs.** Let Σ be a Polish space and $M(\Sigma)$ denote the space of signed measures on Σ. Let $M_1(\Sigma)$ denote the space of probability measures on Σ. Endow $M(\Sigma)$ with the topology generated by the sets

$$\left\{\beta \in M(\Sigma) : \left|\int \phi d(\beta - \alpha)\right| < r\right\},$$

where $\alpha \in M(\Sigma)$, $\phi \in C_b(\Sigma, \mathbb{R})$ (the set of continuous bounded functions from Σ to $\mathbb{R}$) and $r > 0$. As is shown in 3.9, the Lévy metric on $M_1(\Sigma)$ is a complete separable metric, which is consistent with the restriction of this topology to $M_1(\Sigma)$.

Clearly, the topology inherited by $M_1(\Sigma)$ as a subset of $M(\Sigma)$ is the weak topology (i.e., the topology corresponding to convergence against bounded continuous test functions).

Lemma 3.9. *The Lévy metric ρ (defined in (3.2.1) of* [8]*) is compatible with the weak topology on $M_1(\Sigma)$, and $(M_1(\Sigma), \rho)$ is Polish.*

Let $\tilde{\mu}$ be the distribution of $\sigma \in \Sigma \mapsto \delta_\sigma \in M_1(\Sigma)$ under some $\mu \in M_1(\Sigma)$. Define the distribution $\tilde{\mu}_n$ of the empirical distribution functional

$$\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_n) \in \Sigma^n \mapsto \mathrm{L}_n(\boldsymbol{\sigma}) := \frac{1}{n} \sum_{m=1}^{n} \delta_{\sigma_m} \tag{24}$$

under $\mu^n \in M_1(M_1(\Sigma)^n)$.

Let $\nu, \mu \in M_1(\Sigma)$. The relative entropy of ν with respect to μ is defined as:

$$H(\nu|\mu) = \begin{cases} \int_\Sigma f \log f d\mu & \text{if } \nu << \mu \text{ and } f = \frac{d\nu}{d\mu} \\ \infty & \text{otherwise} \end{cases} \tag{25}$$

Let us state the classical Sanov theorem in the weak convergence topology.

Theorem 3.10 (Sanov, Theorem 3.2.17 of [8])**.** *Let μ be a probability measure on the Polish space Σ and let $\tilde{\mu}_n \in M_1(M_1(\Sigma))$ be the distribution under μ_n of the function L_n, in* (24)*. Also, define $H(\cdot|\mu)$ as in* (25)*. Then $H(\cdot|\mu)$ is a good, convex rate function on $M_1(\Sigma)$ and $\{\tilde{\mu}_n : n \geq 1\}$ satisfies the full large deviation principle with rate function $H(\cdot|\mu)$.*

Corollary 3.10.1. *Consider the i.i.d sequence $\{(X^n, \hat{X}^n)\}_{n\in\mathbb{N}}$ where each X^n is a standard Brownian motion and $\hat{X}^n$ is a fractional Brownian motion with Hurst parameter H.*

Let $\Sigma = C^{\alpha-Hol}([0,T], \mathbb{R}) \times C^{\beta-Hol}([0,T], \mathbb{R})$ for any $\alpha < 1/2$ and any $\beta < H$. It is a Polish space in which each random variable $(X^n, \hat{X}^n)$ takes their values. Denote μ the law on Σ.

Let $\tilde{\mu}_n \in M_1(M_1(\Sigma))$ be the distribution under μ_n of the function L_n, in (24)*. Also, define $H(\cdot|\mu)$ as in* (25)*. Then $H(\cdot|\mu)$ is a good, convex rate function on $M_1(\Sigma)$ and $\{\tilde{\mu}_n : n \geq 1\}$ satisfies the full large deviation principle with rate function $H(\cdot|\mu)$.*

We state a general fact of Gaussian analysis, Theorem D.2 of [15], in a Brownian context. The Cameron-Martin space for d-dimensional Brownian motion is given by (cf. Section 1.4.1 of [15])

$$\mathcal{H} = W_0^{1,2}([0,T], \mathbb{R}^d) = \left\{ \int_0^\cdot \dot{h}_t dt : \dot{h} \in L^2([0,1], \mathbb{R}^d) \right\}, \tag{26}$$

and has Hilbert structure given by $\langle h, g\rangle_{\mathcal{H}} = \langle \dot{h}, \dot{g}\rangle_{L^2}$.

Let B denote a d-dimensional standard Brownian motion on $[0,T]$. If $P_\epsilon := (\epsilon B)_* \mathbb{P}$ denotes the law of ϵB, viewed as a Borel measure on $C_0([0,T], \mathbb{R}^d)$, the next theorem can be summarized in saying that $(P_\epsilon)_{\epsilon>0}$ satisfies a large deviation principle on the space $C_0([0,T], \mathbb{R}^d)$ with rate function I. (When no confusion arises, we shall simply say that $(\epsilon B)_{\epsilon>0}$ satisfies a large deviation principle.) Let us define the good rate function

$$I(h) = \begin{cases} \frac{1}{2}\langle h, h\rangle_{\mathcal{H}} & \text{if } h \in \mathcal{H} \\ +\infty & \text{otherwise} \end{cases} \tag{27}$$

where $\mathcal{H}$ denotes the Cameron–Martin space for B as defined in (26). We now state that $(\epsilon B)_{\epsilon>0}$ satisfies a large deviation principle in uniform topology with good rate function I. This is nothing other than a special case of the general large deviation result for Gaussian measures on Banach spaces, see Section D.2 of [15].

Theorem 3.11 (Schilder, Theorem 13.38 of [15])**.** *Let B be a d-dimensional Brownian motion on $[0,T]$. For any measurable $A \subseteq C_0([0,T], \mathbb{R}^d)$ we have*

$$-I(A^o) \leq \liminf_{\epsilon\to 0} \epsilon^2 \log \mathbb{P}[\epsilon B \in A] \leq \limsup_{\epsilon\to 0} \epsilon^2 \log \mathbb{P}[\epsilon B \in A] \leq -I(\bar{A}).$$

3.3. LDPs on Brownian rough paths.

The following is inspired from Section 13 of [15] and Section 3 of [12].

Definition 3.12 (Lévy's area, Definition 13.2 of [15])**.** *Given a d-dimensional Brownian motion $B = (B^1, \dots, B^d)$, we define the d-dimensional Lévy area $A = (A^{i,j} : i,j \in \{1, \dots, d\})$ as the continuous process*

$$t \mapsto A_t^{i,j} = \frac{1}{2}\left(\int_0^t B_s^i dB_s^j - B_s^j dB_s^i \right).$$

We also define the Lévy area increments as, for any $s < t$ in $[0,T]$,

$$\begin{aligned} A_{s,t}^{i,j} &= A_t^{i,j} - A_s^{i,j} - \frac{1}{2}\left(B_s^i B_{s,t}^j - B_s^j B_{s,t}^i \right) \\ &= \frac{1}{2}\left(\int_s^t B_{s,r}^i dB_r^j - B_{s,r}^j dB_r^i \right). \end{aligned}$$

Remark 3.13. *Let $\beta, \tilde{\beta}$ be two independent 1-dimensional Brownian motions. Consider a "building block" of Lévy's area of form $\int_0^1 \beta d\tilde{\beta}$. We observe that, conditional on $\beta(\cdot)$, we can view $\int_0^1 \beta d\tilde{\beta}$ as if the integrand β were deterministic and from a very basic form of Itô's isometry,*

$$\int_0^1 \beta_s d\tilde{\beta}_s \sim \mathcal{N}\left(0, \int_0^1 \beta_s^2 ds\right).$$

Definition 3.14 (Enhanced Brownian motion, Definition 13.9 of [15]). *Let B and A denote a d-dimensional Brownian motion and its Lévy area process. The continuous $G^2(\mathbb{R}^d)$-valued process $\boldsymbol{B}$, defined by*

$$\boldsymbol{B}_t := \exp[B_t + A_t], \qquad t \geq 0,$$

is called enhanced Brownian motion; if we want to stress the underlying process we call $\boldsymbol{B}$ the natural lift of B. Sample path realizations of $\boldsymbol{B}$ are called Brownian rough paths. This definition is motivated by 3.18.

We also write $\mathbf{B}_{s,t} = \mathbf{B}_s^{-1} \otimes \mathbf{B}_t \in G^2(\mathbb{R}^d)$ and observe that this is consistent with

$$\mathbf{B}_{s,t} = \exp[B_{s,t} + A_{s,t}].$$

Remark 3.15 (Exercise 13.10 of [15]). *Almost surely,*

$$\boldsymbol{B}_t = \left(1, B_t, \int_0^t B \otimes \circ dB\right) \in G^2(\mathbb{R}^d),$$

where $\circ dB$ denotes Stratonovich integration.

Proposition 3.16 (Proposition 3.4 of [12]). *For any $\alpha \in (1/2, 1/2]$, with probability one*

$$\boldsymbol{B}^{It\hat{o}} = (B, \mathbb{B}^{It\hat{o}}) \in \mathcal{C}^\alpha([0,T], \mathbb{R}^d).$$

In fact, the homogenous rough path norm $|\boldsymbol{B}|_{\alpha-Hol}$ has Gaussian tails.

Remark 3.17. *Observe that Brownian motion enhanced with its iterated Ito integrals yields a (random) rough path but not a geometric rough path which is, by definition, an object with hardwired first order behaviour. Indeed, Itô formula yields the identity*

$$d(B^i B^j) = B^i dB^j + B^j dB^i + \langle B^i, B^j \rangle dt, \qquad i, j = 1, \dots, d,$$

so that, writing Id for the identity matrix in d dimensions, we have for $s < t$,

$$Sym(\mathbb{B}^{It\hat{o}}_{s,t}) = \frac{1}{2}\mathbb{B}_{s,t} \otimes \mathbb{B}_{s,t} - \frac{1}{2} Id(t-s) \neq \frac{1}{2}\mathbb{B}_{s,t} \otimes \mathbb{B}_{s,t}$$

Proposition 3.18 (Corollary 13.14 of [15]). *Define $\boldsymbol{B}$ as in 3.14.*
Let $\alpha \in [0, 1/2)$. Then there exists $\eta > 0$, not dependent on T, such that

$$\mathbb{E}\left[\exp\left(\frac{\eta}{T^{1-2\alpha}} \|\boldsymbol{B}\|^2_{\alpha-Hol;[0,T]}\right)\right] < \infty.$$

Thus, $\boldsymbol{B}$ is a weak geometric α-Hölder rough path as defined in 2.36, i.e $\boldsymbol{B} \in C^{\alpha-Hol}([0,T], G^2(\mathbb{R}^d))$.

Proposition 3.19 (Proposition 13.20 of [15]). *Let D be a dissection of $[0,T]$ and $1/r \in [0, 1/2]$. Then, there exists $C = C(T)$ such that, for $k = 1, 2$ and all $0 \leq s < t \leq T$, $q \geq 1$,*

$$\left|\pi_k(\boldsymbol{B}_{st} - S_2(B^D)_{st})\right|_{L^q(\mathbb{P})} \leq C|D|^{1/2-1/r}(\sqrt{q}|t-s|^{1/r})^k$$

Corollary 3.19.1 (Corollary 13.21 of [15]). *Let $\alpha \in [0, 1/2)$. Then, for every $\eta \in (0, 1/2 - \alpha)$, there exists a constant $C = C(\alpha, \eta, T)$ such that, for all $q \in [1, \infty)$,*

$$\left|d_{\alpha-Hol,[0,T]}(\boldsymbol{B}, S_2(B^D))\right|_{L^q(\mathbb{P})} \leq C|D|^{\eta/2}\sqrt{q}$$

and, also for $k \in \{1, 2\}$,

$$\left|\rho^{(k)}_{\alpha-Hol,[0,T]}(\boldsymbol{B}, S_2(B^D))\right|_{L^q(\mathbb{P})} \leq C|D|^\eta \sqrt{q}^k$$

Remark 3.20 (Exercise 13.22 of [15]). *Let $(D_n) \subseteq \mathcal{D}([0,T])$ be a sequence of dissections of $[0,T]$. We have that, $S_2(B^{D_n}) \to \boldsymbol{B}$ almost surely with respect to α-Hölder rough path topology, $\alpha \in [0, 1/2)$, provided mesh $|D_n| \to 0$ fast enough, for example with dyadic dissections, as shown in Proposition 3.6 of* [12].

Let B denote d-dimensional standard Brownian motion. It is rather obvious that $\epsilon B \to 0$ in distribution as $\epsilon \to 0$. The same can be said for enhanced Brownian motion $\mathbf{B}$ provided scalar multiplication by ϵ on $\mathbb{R}^d$ is replaced by dilation δ_ϵ on $G^2(\mathbb{R}^d)$, i.e. $\delta_\epsilon \mathbf{B} \to o$ in distribution as $\epsilon \to 0$. It turns out that, to leading order, the speed of this convergence can be computed very precisely. This is a typical example of a large deviations statement for sample paths. Adopting standard terminology, the goal of this section is prove a large deviation principle for enhanced Brownian motion $\mathbf{B}$ in suitable rough path metrics. There is an obvious motivation for all this. The contraction principle will imply - by continuity of the Itô-Lyons maps and without any further work - a large deviation principle for rough differential equations driven by enhanced Brownian motion. Combined with the fact that RDEs driven by enhanced Brownian motion are exactly Stratonovich stochastic differential equations, this leads directly to large deviations for SDEs, better known as Freidlin-Wentzell estimates.

Theorem 3.21 (Theorem 13.42 of [15]). *For any $\alpha \in [0, 1/2)$, the family $(\delta_\epsilon \boldsymbol{B} : \epsilon > 0)$ satisfies a large deviation in homogenous α-Hölder topology. More precisely, viewing $\boldsymbol{P}_\epsilon := (\delta_\epsilon \boldsymbol{B})_* \mathbb{P}$ as a Borel measure on the Polish space $(C_0^{0,\alpha-Hol}([0,T], G^2(\mathbb{R}^d)), d_{\alpha-Hol})$, the family $(\boldsymbol{P}_\epsilon : \epsilon > 0)$ satisfies a large deviation principle on this space with good rate function, defined for $x \in C_0^{0,\alpha-Hol}([0,T], G^2(\mathbb{R}^d))$, given by*

$$J(x) = \frac{1}{2}\langle \pi_1(x), \pi_1(x) \rangle_{\mathcal{H}} \quad \text{if } \pi_1(y) \in \mathcal{H},$$

where $\mathcal{H}$ is defined in (26).

3.4. LDPs on Gaussian rough paths.

Theorem 3.22 (Enhanced Gaussian process, Theorem 15.33 of [15]). *Assume $X = (X^1, \dots, X^d)$ is a centred continuous Gaussian process with independent components. Let $\rho \in [1,2)$ and assume the covariance of X is of finite ρ-variation dominated by a 2D control ω with $\omega([0,1]^2) \le K$. Then, there exists a unique continuous $G^3(\mathbb{R}^d)$ -valued process $\boldsymbol{X}$, such that:*

(i) $\boldsymbol{X}$ "lifts" the Gaussian process X in the sense $\pi_1(\boldsymbol{X}_t) = X_t - X_0$;

(ii) there exists $C = C(\rho)$ such that for all $s < t$ in $[0,1]$ and $q \in [1, \infty)$,

$$|d(\boldsymbol{X}_s, \boldsymbol{X}_t)|_{L^q} \le C\sqrt{q}\omega([s,t]^2)^{1/2\rho};$$

(iii) (Fernique-estimates) for all $p > 2\rho$ and $\omega([0,1]^2) \le K$, there exists $\eta = \eta(p, \rho, K) > 0$, such that

$$\mathbb{E}[\exp(\eta \|\boldsymbol{X}\|^2_{p-var;[0,1]})] < \infty$$

and if $\omega([s,t]^2) \le K|t-s|$ for all $s < t$ in $[0,1]$, then we may replace $\|\boldsymbol{X}\|_{p-var;[0,1]}$ by $\|\boldsymbol{X}\|_{1:p-Hol;[0,1]}$;

(iv) the lift $\boldsymbol{X}$ is natural in the sense that it is the limit of $S_3(X^n)$ where X^n is any sequence of piecewise linear or mollifier approximations to X such that $d_\infty(X^n, X)$ converges to 0 almost surely.

Definition 3.23 (Definition 15.34 of [15]). *A $G^3(\mathbb{R}^d)$-valued process $\boldsymbol{X}$ as constructed above is called an enhanced Gaussian process; if we want to stress the underlying Gaussian process we call $\boldsymbol{X}$ the natural lift of X. Sample path realizations of $\boldsymbol{X}$ are called Gaussian rough paths, as is motivated by*

(i) for $\rho \in [1, 3/2)$, we see that $\boldsymbol{X}$ has almost surely finite p-variation, for any $p \in (2\rho, 3)$, and hence so does its projection to $G^2(\mathbb{R}^d)$, which is therefore almost surely a geometric p-rough path;

(ii) for $\rho \in [3/2, 2)$, we see that $\boldsymbol{X}$ has almost surely finite p-variation, for any $p \in (2\rho, 4)$, and is therefore almost surely a geometric p-rough path.

Remark 3.24. *Theorem 3.22 asserts in particular that d-dimensional Brownian motion can be naturally lifted to an enhanced Gaussian process, easily identified as enhanced Brownian motion (see in Appendix 16). Other examples are obtained by considering d independent (continuous, centred) Gaussian processes, each of which satisfies the condition that its covariance is of finite ρ-variation, for some $\rho < 2$. For example (see in Appendix 17.1) one may take d independent copies of fractional Brownian motion: the resulting $\mathbb{R}^d$-valued fractional Brownian motion B^H can be lifted to an enhanced Gaussian process ("enhanced fractional Brownian motion", $\boldsymbol{B}^H$) provided $H > 1/4$. Further examples are constructed by consulting the list of Gaussian processes in Section 15.2 of* [15].

Remark 3.25. *By Proposition 17.2, a fractional BM with Hurst parameter $H \le 1/2$ has its covariance function of finite $1/(2H)$-variation. If $H \le 1/4$, $2H \le 1/2$ we have $1/(2H) \ge 2$. Thus, the covariance function is only of finite $1/(2H) \ge 2$-variation. That is why we cannot enhance an fBM with $H \le 1/4$.*

Theorem 3.26 (Theorem 15.42 of [15]). *Assume that $X = (X^1, \dots, X^d)$ is a centred continuous Gaussian process with independent components and covariance R of finite ρ-variation, $\rho \in [1,2)$, controlled by some 2D control ω. Fix an arbitrary $p \in (2\rho, 4)$, $\eta \in \left(0, \frac{1}{2\rho} - \frac{1}{p}\right)$ and write $\boldsymbol{X}$ for the natural lift of X. Then*

(i) *if $\omega\Big([0,1]^2\Big) \le K$, there exists some constant $C_1 = C(\rho, p, K, \theta)$, such that for all $D \in \mathcal{D}([0,1])$ and $q \in [1, \infty)$,*

$$|d_{p-var;[0,1]}(\boldsymbol{X}, S_3(X^D))|_{L^q(\mathbb{P})} \le C_1\sqrt{q} \max_{t_i \in D} \omega([t_i, t_{i+1}]^2)^{\eta/3}$$

and also

$$\forall n \in \{1,2,3\}, \left|\rho^{(n)}_{p-var;[0,1]}(\boldsymbol{X}, S_3(X^D))\right|_{L^q(\mathbb{P})} \le C_1 q^{n/2} \max_{t_i \in D} \omega([t_i, t_{i+1}]^2)^{\eta}$$

(ii) *if $\omega(s,t) \le K|t-s|$ for all $s < t$ in $[0,1]$ then $d_{p-var;[0,1]}, \rho^{(n)}_{p-var;[0,1]}$ in the above estimates may be replaced by $d_{1/p-Hol;[0,1]}, \rho^{(n)}_{1/p-Hol;[0,1]}$ respectively.*

Let $X = (X^1, \dots, X^d)$ denote a centred continuous Gaussian process on $[0,1]$, with independent components, each with covariance of finite ρ-variation for some $\rho \in [1,2)$ and dominated by some 2D control ω. We write $\mathcal{H}$ for its associated Cameron–Martin space. Since the law of X induces a Gaussian measure on $C([0,1], \mathbb{R}^d)$, it follows from general principles (see Section D.2 of [15]) that $(\epsilon X : \epsilon > 0)$ satisfies a large deviation principle with a ϵ^2 instead of ϵ as a factor of the log (like in 3.11) with good rate function I in uniform topology, where I is given by

$$I(x) = \begin{cases} \frac{1}{2}\langle x, x\rangle_{\mathcal{H}} & \text{if } x \in \mathcal{H} \subseteq C([0,1], \mathbb{R}^d) \\ +\infty & \text{otherwise .} \end{cases} \tag{28}$$

We write ϕ_m for the piecewise linear approximations along the dissection $D_m = \{i/m : i = 0, \dots, m\}$. It is clear that

$$S_3 \circ \phi_m : (C([0,1], \mathbb{R}^d), |\cdot|_\infty) \to (C([0,1], G^3(\mathbb{R}^d)), d_\infty)$$

is continuous. By the contraction principle, $S_3(\epsilon\phi_m(X))$ satisfies a large deviation principle with good rate function

$$J_m(y) = \inf\{I(x), x \text{ such that } S_3(\phi_m(x)) = y\},$$

the infimum of the empty set being $+\infty$. Essentially, a large deviation principle for $\delta_\epsilon \mathbf{X}$ is obtained by sending m to infinity. To this end one must prove that $S_3(\phi_m(X))$ is an exponentially good approximation to $\mathbf{X}$. This being proven, one deduces the following theorem.

Theorem 3.27 (Theorem 15.55 of [15]). *Assume that*
(i) $X=(X^1,\dots,X^d)$ *is a centred continuous Gaussian process on* $[0,1]$ *with independent components;*
(ii) $\mathcal{H}$ *denotes the Cameron-Martin space associated with* X*;*
(iii) the covariance of X *is of finite* ρ*-variation dominated by some 2D control* ω*, for some* $\rho\in[1,2)$*;*
(iv) $\boldsymbol{X}$ *denotes the natural lift of* X *to a* $G^3(\mathbb{R}^d)$*-valued process.*
Then, for any $p\in(2\rho,4)$*, the family* $(\delta_\epsilon \mathbf{x})_{\epsilon>0}$ *satisfies a large deviation principle in* p*-variation topology with good rate function, defined for* $x\in C_o^{0,p-var}([0,1],G^3(\mathbb{R}^d))$ *, given by*

$$J(x)=\frac{1}{2}\langle \pi_1(\mathbf{x}),\pi_1(\mathbf{x})\rangle_{\mathcal{H}} \quad \text{if } \pi_1(\mathbf{x})\in\mathcal{H}\}.$$

If ω *is Hölder-dominated then the large deviation principle holds in* $1/p$*-Hölder topology.*

3.5. **LDPs on SDEs.** In Theorem 3.21, we saw that Schilder's theorem holds for enhanced Brownian motion. Since $\pi_{(V)}(0,y_0;\mathbf{B})$, a Stratonovich solution to $dY=V(Y)d\mathbf{B}$, depends continuously on $\mathbf{B}$ in this α-Hölder rough path topology, we can apply the contraction principle to deduce (without any further work) a large deviation principle for solution of stochastic differential equations; better known as Freidlin-Wentzell estimates. More precisely, also including a drift term, we have

Theorem 3.28 (Freidlin–Wentzell large deviations, Theorem 19.9 of [15]). *Assume that* $V=(V_1,\dots,V_d)$ *is a collection of* Lip^2*-vector fields on* $\mathbb{R}^e$*, and* V_0 *is a* Lip^1*-vector field on* $\mathbb{R}^e$*. Let* B *be a* d*-dimensional Brownian motion and consider the unique (up to indistinguishability) Stratonovich SDE solution on* $[0,T]$ *to*

$$dY^\epsilon=\sum_{i=1}^d V_i(Y)\circ\epsilon dB^i+V_0(Y)dt$$

started at y_0*. Let* $\alpha\in[0,1/2)$*. Then* Y^ϵ *satisfies a large deviation principle (in* α*-Hölder topology) with good rate function given by*

$$J(y)=\inf\{I(h):\pi_{(V,V_0)}(0,y_0;(h,t))=y\}$$

where I *is given in* (27).

Eventually, we can state a large deviations result for SDEs driven by more general stochastic processes than the Brownian motion. Using the large deviation results for enhanced Gaussian and Markov processes established in Section 15.7 resp. Section 16.7 of [15], we can generalize the previous section to RDEs driven by Gaussian and Markovian signals. Let us state the result for Gaussian processes.

Proposition 3.29 (Proposition 19.14 of [15]). *Assume that*
(i) $X=(X^1,\dots,X^d)$ *is a centred continuous Gaussian process on* $[0,1]$ *with independent components;*
(ii) $\mathcal{H}$ *denotes the Cameron-Martin space associated with* X*;*
(iii) the covariance of X *is of finite* ρ*-variation dominated by some 2D control* ω*, for some* $\rho\in[1,2)$*;*
(iv) $\boldsymbol{X}$ *denotes the natural lift of* X *to a* $G^{[2\rho]}(\mathbb{R}^d)$*-valued process (with geometric rough sample paths);*
(v) $V=(V_1,\dots,V_d)$ *is a collection of* Lip^γ*-vector fields on* $\mathbb{R}^e$*, with* $\gamma>2\rho$*;*
(vi) $Y^\epsilon=\pi_{(V)}(0,y_0;\delta_\epsilon\boldsymbol{X})$ *is the RDE solution to*

$$dY^\epsilon=\epsilon V(Y^\epsilon)d\boldsymbol{X},\quad Y(0)=y_0\in\mathbb{R}^e.$$

Then, for any $p>2\rho$*,* $(Y^\epsilon:\epsilon>0)$ *satisfies a large deviation principle in* p*-variation topology, with good rate given by*

$$J(y)=\inf\left\{\frac{1}{2}|h|^2_{\mathcal{H}}:\pi_{(V)}(0,y_0;h)=y\right\}$$

where we agree that $|h|^2_{\mathcal{H}}=+\infty$ *when* $h\notin\mathcal{H}$*.*
If ω *is Hölder-dominated, then the above large deviation principle also holds in* $1/p$*-Hölder topology.*

4. Large deviations on rough volatility models

Recall the simple rough volatility model (2) and Remark 1.1. This allows to write the model as

$$dS_t=S_tf(\hat{X}_t)dX_t=S_tf(\hat{X}_t)(\rho dW_t+\bar{\rho}d\bar{W}_t),$$

with $\hat{X}_t:=\int_0^t k_H(t-s)dW_s$ and $X:=\rho W+\bar{\rho}\bar{W}$ with $\rho\in(-1,1)$ and $\bar{\rho}:=\sqrt{1-\rho^2}$ and $(W,\bar{W})$ a 2-dimensional Brownian motion.

Why do we need LDPs on rough volatility ? The literature ([4],[11],[3],[13],[14],[22],[23],[24],[29],[19],[30]) has extensively investigated LDPs under rough volatility in order to derive a power law and seek a precise approximation formula.

Here, we give an overview of why it is complex to show large deviations for the price (or log-price) of an asset following the rough volatility model. We depict a few approaches that have worked and why a classical rough path approach cannot work (off-the-shelf).

4.1. **Complications with rough volatility.** This subsection is inspired from Sections 1.1 and 1.2 of [3].

For comparison with rough volatility below, we first mention a selection of tools and methods well-known for Markovian Stochastic volatility models.

- PDE methods are ubiquitous in (low-dimensional) pricing problems, as are
- Monte Carlo methods, noting that knowledge of strong (resp. weak) rate 1/2 (resp. 1) is the grist in the mills of modern multilevel methods (MLMC);

- Quasi Monte Carlo (QMC) methods are widely used; related in spirit we have the Kusuoka–Lyons–Victoir cubature approach, popularized in the form of Ninomiya–Victoir (NV) splitting scheme, nowadays available in standard software packages;
- Freidlin-Wentzell theory of small noise large deviations is essentially immediately applicable, as are various "strong" large deviations (a.k.a. exact asymptotics) results, used e.g. the derive the famous SABR formula.

For several reasons it can be useful to write model dynamics in Stratonovich form: From a PDE perspective, the operators then take sum-square form which can be exploited in many ways (Hörmander theory, naturally linked to Malliavin calculus ...). From a numerical perspective, we note that the cubature / NV scheme [36] also requires the full dynamics to be rewritten in Stratonovich form. Another financial example that requires a Stratonovich formulation comes from interest rate model validation [5], based on the Stroock-Varadhan support theorem. We further note, that QMC (e.g. Sobol') works particularly well if the noise has a multiscale decomposition, as obtained by interpreting a (piece-wise) linear Wong-Zakai approximation, as Haar wavelet expansion of the driving white noise.

Due to loss of Markovianity, PDE methods are not applicable to rough volatility, and neither are (off-the-shelf) Freidlin-Wentzell large deviation estimates but Forde and Zhang [11] managed to prove some large deviations under the strong assumption of global Hölder-regularity of f in (2). Moreover, rough volatility is not a semi-martingale, which complicates, to say the least, the use of several established stochastic analysis tools. In particular, rough volatility admits no Stratonovich form. Closely related, one lacks a (Wong-Zakai type) approximation theory for rough volatility. To see this, focus on the "simple" rough volatility model, that is (2) so that

$$S_t/S_0 = \mathcal{E}\left(\int_0^{\cdot} f(\hat{X}_s)dX_s\right)(t).$$

Inside the (classical) stochastic exponential of a continuous semimartingale M, $\mathcal{E}(M)(t) = \exp(M_t - \frac{1}{2}[M]_t)$ we have the logarithm of the martingale term

$$\int_0^t f(\hat{X}_s)dX_s = \rho\int_0^t f(\hat{X}_s)dW_s + \bar{\rho}\int_0^t f(\hat{X}_s)d\bar{W}_s$$

and, in essence, the trouble is due to the Itô integral $\int_0^t f(\hat{X}_s)dW_s$. Indeed, any naive attempt to put it in Stratonovich form,

$$\text{“}\int_0^t f(\hat{X}_s)\circ dW_s := \int_0^t f(\hat{X}_s)dW_s + \text{ (Itô-Stratonovich correction) ”}$$

or, in the spirit of Wong-Zakai approximations,

$$\text{“}\int_0^t f(\hat{X}_s)\circ dW_s := \lim_{\epsilon\to 0}\int_0^t f(\hat{X}^\epsilon)dW^\epsilon\text{ ”}$$

must fail whenever $H < 1/2$. The Itô-Stratonovich correction is given by the quadratic covariation, defined (whenever possible) as the limit, in probability, of

$$\sum_{[u,v]\in D}(f(\hat{X}_v) - f(\hat{X}_u))(W_v - W_u),$$

along any sequence (D^n) of partitions with mesh-size $|D^n|$ tending to zero. But, disregarding trivial situations, this limit does not exist. For instance, when $f(x) = x$ fractional scaling immediately gives divergence (at rate $H - 1/2$) of the above bracket approximation.

4.2. **An approach to LDPs for rough volatility not using rough paths.** In [11]:"Using the large deviation principle (LDP) for a rescaled fractional Brownian motion B_t^H, where the rate function is defined via the reproducing kernel Hilbert space, they compute small-time asymptotics for a correlated fractional stochastic volatility model of the form $dS_t = S_t\sigma(Y_t)(\bar{\rho}dW_t + \rho dB_t), dY_t = dB_t^H$, where σ is α-Hölder continuous for some $\alpha \in (0,1]$; in particular, we show that $t^{H-1/2}\log S_t$ satisfies the LDP as $t \to 0$."

In [11], they denote $S_t = e^{X_t}$ and set

$$\begin{cases} dS_t = S_t\sigma(Y_t)(\bar{\rho}dW_t + \rho dB_t), \\ dY_t = dB_t^H. \end{cases}$$

with $B_t^H := \int_0^t k_h(t-s)dW_s$, and σ is α-Hölder continuous on $\mathbb{R}^+$ and (W, B) a 2-dimensional Brownian motion. They introduce the small-noise process

$$\begin{cases} dX_t^\epsilon = -\frac{1}{2}\epsilon\sigma(Y_t^\epsilon)^2dt + \sqrt{\epsilon}\sigma(Y_t^\epsilon)[\bar{\rho}dW_t + \rho dB_t], \\ dY_t^\epsilon = \epsilon^H dB_t^H. \end{cases} \tag{29}$$

with $X_0^\epsilon = 0, Y_0^\epsilon = 0$. Therefore, they indeed look at the same rough volatility we depicted in 2, that is with correlation between the standard Brownian integrator and the fractional Brownian motion driving underlying the variance process. Let us the notation of [11] in this subsection for clarity.

Theorem 4.1 (Theorem 4.8 of [11]). *$t^{H-1/2}X_t$ satisfies the LDP as $t \to 0$ with speed $\frac{1}{t^{2H}}$ and good rate function given by*

$$I(x) = \inf_{f\in H^1}\left[\frac{(x-\rho G(f))^2}{2\bar{\rho}^2F(\boldsymbol{K}_H f')} + \frac{1}{2}\|f\|_{\mathcal{H}}^2\right] \le \frac{x^2}{2\bar{\rho}^2\sigma(0)^2},$$

where $F(f) = \int_0^1 \sigma((\boldsymbol{K}_H f')(s))^2 ds, G(f) = \int_0^1 \sigma((\boldsymbol{K}_H f')(s)) f'(s) ds$, $\mathcal{H}$ *defined in* (26) *is the usual Cameron-Martin space for Brownian motion and the operator* $\boldsymbol{K}_H$ *is defined by* $(\boldsymbol{K}_H f)(t) = \int_0^t K_H(s,t) f(s) ds$ *for* $t \in [0,1]$ *and*

$$K_H(s,t) = \begin{cases} (c_+ s^{1/2-H} \int_s^t (u-s)^{H-3/2} u^{H-1/2} du & \text{if } H \in (1/2,1), \\ c_- \left[\left(\frac{t}{s}\right)^{H-1/2} (t-s)^{H-1/2} - (H-\frac{1}{2}) s^{1/2-H} \int_s^t u^{H-3/2} 2(u-s)^{H-1/2} du \right] & \text{if } H \in (0,1/2), \end{cases}$$

$c_+ = \left[\frac{H(2H-1)}{\beta(2-2H,H-\frac{1}{2})}\right]^{1/2}$ *and* $c_- = \left[\frac{2H}{(1-2H)\beta(1-2H,H+\frac{1}{2})}\right]^{1/2}$ *, and* $\beta(\cdot,\cdot)$ *denotes the beta function.* $I(x)$ *attains its minimum value of zero at* $x = \rho G(0) = 0$.

Proof. Given that (B, B^H) is a Gaussian process, they prove that the Reproducing Kernel Hilbert space (RKHS) for (B, B^H) is $\mathcal{H}_H^2 := \{(f,g) \in C_0([0,1],\mathbb{R}^2) : f(t) = \int_0^t \dot{h}(s) ds, g(t) = \int_0^t K_H(s,t)\dot{h}(s) ds, \dot{h} \in L^2[0,1]\}$. Using the general LDP for Gaussian processes in Subsection 3.3 of [11] or Section D.2 of [15], we know that $\epsilon^H(B^H, B)$ satisfies a joint LDP on $C_0([0,1],\mathbb{R}^2)$ as $\epsilon \to 0$ with speed $1/\epsilon^{2H}$ and rate function defined as in (28)

$$I_H(f,g) = \begin{cases} \frac{1}{2}\int_0^1 \dot{h}(s)^2 ds \text{ if } (f,g) \in \mathcal{H}_H^2, \\ +\infty \text{ otherwise.} \end{cases}$$

Then, Forde and Zhang only make use of arguments of Large Deviations theory, Gaussian analysis and the definition of the fractional Brownian motion as an integral. They build exponentially good approximations to X^ϵ defined in (29) and the extended contraction principle.

□

As we can see, [11] do not need to use any rough path theory to prove their main theorem. Rough path theory allows to prove Freidlin-Wentzell LDPs as pure corollaries of the continuity of the Itô-Lyons map but other ways to show such LDPs exist.

Remark 4.2. *Yet, this model makes a strong assumption on the regularity and growth of* σ*. Indeed, as one sees in Lemma 4.3 of* [11]*, they need* σ *to be globally Hölder, so that they get a condition of at most linear growth of* σ*.*

4.3. **The problem with classical rough path theory for rough volatility when $H \leq 1/4$.** We can only apply Theorem 3.29 with the 2-dimensional Gaussian process $(W, \hat{B})$, with W being a Brownian motion independent of the fractional Brownian motion $\hat{B}$.

A rough path lift (as defined in Theorem 15.33 of [15]) of the Gaussian process $(X, \hat{X})$ would require to be able to give a meaning to iterated integrals of $\hat{X}$. Yet,

The usual way to lift a multidimensional continuous semi-martingale or Gaussian process (3.23) is based on independence of the components. In our case, let $X = \rho W + \bar{\rho}\bar{W}$ with $W, \bar{W}$ independent standard Brownian motions, $\hat{X}_t = \int_0^t k_H(t-s) dW_s$.

If we look at iterated integrals of $(X, \hat{X})$, we will look at

$$\int \hat{X} d\hat{X}$$

which we can only manage if $H > 1/4$ according to remark 3.24. This does not work in the case of $H \leq 1/4$ for reasons depicted in 3.25.

The second type of iterated integral is

$$\int \hat{X} dX = \rho \int \hat{X} dW + \bar{\rho} \int \hat{X} d\bar{W}.$$

This can be managed because we can define $\int \hat{X} dW$ as an Itô integral with deterministic integrand as in 3.13 and $\int \hat{X} d\bar{W}$ as an Itô integral on

$$\mathbb{H}^2_{loc}[0,T] := \{\phi \text{ measurable}, \mathcal{F}-\text{adapted such that } \int_0^T |\phi_s|^2 ds < \infty \text{ a.s. }\}$$

A third type of iterated integral is

$$\int X dX = \rho \int W dX + \bar{\rho} \int \bar{W} dX = \rho^2 \int W dW + \rho\bar{\rho} \int W d\bar{W} + \rho\bar{\rho} \int \bar{W} dW + \bar{\rho}^2 \int \bar{W} d\bar{W}$$

which is a sum of Itô integrals of standard BMs.

The fourth and final type of iterated integral is

$$\int X d\hat{X} = \rho \int W d\hat{X} + \bar{\rho} \int \bar{W} d\hat{X}$$

$\int \bar{W} d\hat{X}$ can be defined as a conditional Wiener integral with respect to an fBM, since it is easier than defining $\int \hat{X} d\hat{X}$ (which we can do with 3.24).

$\int W d\hat{X}$ is the only problematic iterated integral, that is we do not know how to define it with: if these were independent it would be well-defined by 3.22 but they are not. One other possibility is to use Gubinelli's controlled rough path theory, this is treated in the next subsection.

4.4. **The controlled rough paths approach fails.** As mentioned by Fukasawa (Introduction of [17]), one cannot control the integrand of the rough volatility model $f(\hat{X})$ by the integrator X in 2 in the sense of Definition 2.46.

Indeed, as seen in the previous subsection, we can treat $\int f(\hat{X})dX = \rho \int f(\hat{X})dW + \bar{\rho} \int f(\hat{X})d\bar{W}$ as a sum of Itô integrals, which (for f smooth enough) have the Hölder regularity (Proposition 3.1 of [17] or using another inequality like Burkholder-Davis-Gundy combined with the Kolmogorov-Chentsov theorem), $(H+1/2)^-$ with H being the Hurst parameter of the fractional Brownian motion $\hat{X}$. If f is less smooth, we can at least say that $\int f(\hat{X})dX$ is $(1/2)^-$-Hölder regular.

If we want to have consistence between a stochastic integral formulation and a controlled rough path formulation of $\int f(\hat{X})dX$ we need the Hölder regularity of the rough integral $\int f(\hat{X})d\mathbf{X}$ defined via Theorem 2.47 to be $(H+1/2)^-$ (or at least $(1/2)^-$). Thus, we need $(\hat{X}, \hat{X}') \in \mathcal{D}_X^{\beta+1/2}([0,T], L(\mathbb{R}^d, \mathbb{R}^d))$ for all $\beta \in (0, H)$ or at least in $\mathcal{D}_X^{\alpha}([0,T], L(\mathbb{R}^d, \mathbb{R}^d))$ for all $\alpha \in (0, 1/2)$.

Thus, we want to have (21) for all $\alpha \in (0, 1/2)$ and for $Y = \hat{X}$, that is

$$\hat{X}_{s,t} = (\hat{X})'_s X_{s,t} + R^{\hat{X}}_{s,t},$$

with $(\hat{X})'$ being α-Hölder and $R^{\hat{X}}_{s,t}$ being 2α-Hölder. We know that the Hölder regularity of a product of two functions is at least the minimum of the Hölder regularities of the functions. Thus, in the equation above, $(\hat{X})'_s X_{s,\cdot}$ must be α-Hölder. Thus, the right-hand-side $(\hat{X})'_s X_{s,\cdot} + R^{\hat{X}}_{s,\cdot}$ must be α-Hölder for all $\alpha \in (0, 1/2)$, in particular for $\alpha \in [H, 1/2)$.

Yet, the left-hand-side $\hat{s, \cdot}$ is only H^--Hölder. Hence a contradiction.

Therefore, $\hat{X}$ is not controlled by X and we cannot apply Theorem 2.47 and 2.49.

4.5. **Regularity structures approach.** The regularity structures approach [3] allows to improve the Forde-Zhang (simple rough volatility) short-time large deviations [11] such as to include f of exponential type, a defining feature in the works of Gatheral and coauthors [18],[2]. Indeed, in Remark 4.2, we can see that Forde-Zhang make a very strong assumption on f of (2) (or σ in their notation). The regularity structures approach allows to get rid of this assumption.

As outlined in Section 1.3 of [12], recently, a new theory of "regularity structures" was introduced [26], unifying various flavours of the theory of rough paths (including Gubinelli's controlled rough paths [20], as well as his branched rough paths [21]), as well as the usual Taylor expansions. While it has its conceptual roots in the theory of rough paths, the main advantage of this new theory is that it is no longer tied to the one-dimensionality of the time parameter, which makes it also suitable for the description of solutions to stochastic partial differential equations, rather than just stochastic ordinary differential equations. The main achievement of the theory of regularity structures is that it allows to give a (pathwise!) meaning to ill-posed stochastic PDEs that arise naturally when trying to describe the macroscopic behaviour of models from statistical mechanics near criticality.

5. Fukasawa and Takano's contribution to treat rough volatility

Let us first give an overview as outlined by Fukasawa and Takano in the introduction of [17]. In [3], the LDP for rough volatility models is obtained using the continuity of Hairer's reconstruction map. Fukasawa and Takano take a similar to [3] in spirit but different approach. In stead of embedding a rough volatility model into the abstract framework of regularity structure, they develop a minimal extension of the rough path theory to incorporate rough volatility models. The advantage of our approach is, besides its relatively elementary construction, that we can prove the continuity of the integration map between rough path spaces. This enables to treat a more general case than [3] using merely that the composition of continuous maps is continuous.

The rough volatility model Fukasawa focuses on (equation 1.1 of [17]) is a bit more general than the "simple" rough volatility model defined above in (2) and has the following Itô differential form:

$$dS_t = \theta(S_t, t) f(\hat{X}_t) dX_t, \qquad S_0 \in \mathbb{R}. \tag{30}$$

Indeed, the "simple" rough volatility model corresponds to $\theta(s,t) = s$.

5.1. **The (α, β) rough path space and rough integral.** Throughout the article, they fix $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right]$ and denote

$$\Delta_T := \{(s,t) : 0 \le s \le t \le T\}, \qquad I := \{i \in \mathbb{N} : i\beta + \alpha \le 1\},$$

and

$$J := \{(j,k) \in \mathbb{N} \times \mathbb{N} : (j+k)\beta + 2\alpha \le 1\},$$

Extending the classical notion of α-Hölder rough path of 2.33, they define an (α, β) rough path.

Definition 5.1 (Definition 2.1 of [17]). *An (α, β) rough path $\mathbb{X} = (\hat{X}, X^{(i)}, X^{(jk)})_{i \in I, (j,k) \in J}$ is a triplet of functions on Δ_T satisfying the following conditions; for any $i \in I$, $(j,k) \in J$ and $s \le u \le t$,*

(i) $\hat{X}$ is $\mathbb{R}$-valued, $X^{(i)}$ is $\mathbb{R}^d$-valued, and $X^{(jk)}$ is $\mathbb{R}^d \otimes \mathbb{R}^d$-valued.

(ii) Modified Chen's relation: $\hat{X}_{st} = \hat{X}_{su} + \hat{X}_{ut}$, and

$$X^{(i)}_{st} = X^{(i)}_{su} + \sum_{p=0}^{i} \frac{1}{(i-p)!} (\hat{X}_{su})^{i-p} X^{(p)}_{ut} \tag{31}$$

and

$$X_{st}^{(jk)} = X_{su}^{(jk)} + \sum_{q=0}^{k} \frac{1}{(k-q)!}(\hat{X}_{su})^{k-q} X_{su}^{(j)} \otimes X_{ut}^{(q)} + \sum_{p=0}^{j}\sum_{q=0}^{k} \frac{1}{(j-p)!(k-q)!}(\hat{X}_{su})^{j+k-p-q} X_{ut}^{(pq)}. \tag{32}$$

(iii) Hölder regularity: there exists a constant C such that, for all $(s,t) \in \Delta_T$,

$$|\hat{X}_{s,t}| \leq C(t-s)^{\beta}, \quad |X_{s,t}^{(i)}| \leq C(t-s)^{i\beta+\alpha}, \quad |X_{s,t}^{(jk)}| \leq C(t-s)^{(j+k)\beta+2\alpha}.$$

Let $\Omega_{(\alpha,\beta)-Hol}$ denote the set of the (α,β) rough paths.

We define an inhomogenous metric $d_{(\alpha,\beta)-Hol}$ on $\Omega_{(\alpha,\beta)-Hol}$ and a homogenous norm $\|\cdot\|_{(\alpha,\beta)}$ by

$$\rho_{(\alpha,\beta)-Hol}(\mathbb{X},\mathbb{Y}) := |\hat{X}-\hat{Y}|_{\beta-Hol} + \sum_{i\in I} |X^{(i)} - Y^{(i)}|_{i\beta+\alpha-Hol} + \sum_{(j,k)\in J} |X^{(jk)} - Y^{(jk)}|_{(j+k)\beta+2\alpha-Hol},$$

and

$$\|\!|\mathbb{X}|\!\|_{(\alpha,\beta)-Hol} := |\hat{X}|_{\beta-Hol} + \sum_{i\in I} |X^{(i)}|^{1/(i+1)}_{i\beta+\alpha-Hol} + \sum_{(j,k)\in J} |X^{(jk)}|^{1/(j+k+2)}_{(j+k)\beta+2\alpha-Hol}.$$

Extending the rough path integration, here we introduce an integration with respect to an (α,β) rough path.

Definition 5.2 (Definition 2.4 of [17])**.** *Fix $\mathbb{X} \in \Omega_{(\alpha,\beta)-Hol}$. We define $Y^{(1)}$ and $Y^{(2)}$ as follows if they exist:*

$$Y_{st}^{(1)} = \lim_{|\mathcal{P}|\searrow 0} \sum_{p=1}^{N}\sum_{i\in I} \nabla^i f(\hat{x}_{t_{p-1}}) X^{(i)}_{t_{p-1}t_p}$$

$$Y_{st}^{(2)} = \lim_{|\mathcal{P}|\searrow 0} \sum_{p=1}^{N} \left(Y^{(1)}_{t_0 t_{p-1}} \otimes Y^{(1)}_{t_{p-1}t_p} + \sum_{(j,k)\in J} \nabla^j f(\hat{x}_{t_{p-1}}) \nabla^k f(\hat{x}_{t_{p-1}}) \boldsymbol{X}^{(jk)}_{t_{p-1}t_p} \right)$$

where $\hat{x}_s := \hat{X}_{0s}$ and $\mathcal{P} = \{s = t_0 < t_1 < \cdots < t_N = t\}$ is a partition of $[s,t]$. The mesh size $|\mathcal{P}|$ is defined by $|\mathcal{P}| = \max_j |t_j - t_{j-1}|$. If they exist on Δ_T, we denote $(Y^{(1)}, Y^{(2)})$ by $\int f(\hat{\mathbb{X}})d\mathbb{X}$ and call it the (α,β) rough path integral of f.

Consider $\mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$ as defined in 2.33 the space of α-Hölder rough paths and the inhomogenous distance on this space $\rho_{\alpha-Hol}$ defined in 2.34.

Theorem 5.3 (Theorem 2.5 of [17])**.** *Let $n := \max I$ and assume that $f \in C^{n+1}(\mathbb{R},\mathbb{R})$ ($n+1$ times continuously differentiable).*

(i) For any $\mathbb{X} \in \Omega_{(\alpha,\beta)-Hol}$, the (α,β) rough path integral $\int f(\hat{\mathbb{X}})d\mathbb{X}$ is well-defined, and $\int f(\hat{\mathbb{X}})d\mathbb{X} \in \mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$.

(ii) The integration map $\int : \Omega_{(\alpha,\beta)-Hol} \to \mathcal{C}^{\alpha}([0,T],\mathbb{R}^d)$ is Lipschitz on bounded sets (like in 2.44.1). More precisely, for any $M > 0$, the map $\int_{B_M}$, restricted on the set $B_M := \{\mathbb{X} \in \Omega_{(\alpha,\beta)-Hol} : \|\!|\mathbb{X}|\!\|_{(\alpha,\beta)-Hol} \leq M\}$ is Lipschitz continuous, that is for all $\mathbb{V},\mathbb{W} \in B_M$,

$$\rho_{\alpha-Hol}\left(\int f(\hat{\mathbb{V}})d\mathbb{V}, \int f(\hat{\mathbb{W}})d\mathbb{W} \right) \leq C\rho_{(\alpha,\beta)-Hol}(\mathbb{V},\mathbb{W}).$$

For the purposes of our contribution, which was to generalise (α,β) rough paths and try to prove an enhanced Sanov theorem on them, we had to introduce the following distance, which is not in [17].

Definition 5.4. *We can also define a homogenous distance between (α,β) rough paths as*

$$d_{(\alpha,\beta)-Hol}(\mathbb{X},\mathbb{Y}) := |\hat{X}-\hat{Y}|_{\beta-Hol} + \sum_{i\in I} |X^{(i)} - Y^{(i)}|^{1/(i+1)}_{i\beta+\alpha-Hol} + \sum_{(j,k)\in J} |X^{(jk)} - Y^{(jk)}|^{1/(j+k+2)}_{(j+k)\beta+2\alpha-Hol}. \tag{33}$$

We now construct an (α,β) rough path lift, which is the motivation of the definition of the modified Chen's relations (31) and (32) and of the exponents on each term of the sum in the definition of the homogenous norm $\|\!|\cdot|\!\|_{(\alpha,\beta)-Hol}$ and distance $d_{(\alpha,\beta)-Hol}$.

Proposition 5.5 (Proposition 3.1 of [17])**.** *As set in the 1.1, let $(\Omega,\mathcal{F},\mathbb{P},\{\mathcal{F}_t\}_{t\geq 0})$ be a filtered probability space, and let $H \in (0,1/2)$. Suppose that $X = (X^1,\ldots,X^d)$ is a d-dimensional (possibly with correlation between its components) Brownian motion, and W is a one-dimensional Brownian motion possibly correlated to X (like in 2 but in the current definition we allow for X to be d-dimensional instead of 1-dimensional). Using Itô integration, define $\hat{X}, X^{(i)}, X^{(jk)}$ as*

$$\hat{X}_{st} := \int_0^t k_H(t-r)dW_r - \int_0^s k_H(s-r)dW_r, \quad X_{st}^{(i)} := \frac{1}{i!}\int_s^t (\hat{X}_{sr})^i dX_r,$$

$$X_{st}^{(jk)} := \frac{1}{k!}\int_s^t (\hat{X}_{sr})^k X_{sr}^{(j)} \otimes dX_r$$

for all $(s,t) \in \Delta_T$ with k_H defined in 2. Then, we have the following.

(i) For almost every $\omega \in \Omega$, $\mathbb{X}(\omega) := (\hat{X}(\omega), X^{(i)}(\omega), X^{(jk)}(\omega))_{i\in I,(j,k)\in J}$ is an (α,β) rough path for any $\beta < H$ and $\alpha \in \left(\frac{1}{3},\frac{1}{2}\right)$.

(ii) It holds

$$\left(\int f(\hat{\mathbb{X}})d\mathbb{X}\right)^{(1)}_{0t} = \int_0^t f(\hat{X}_{0r})dX_r, \ \ a.s.$$

where the left-hand-side is the first level of the (α,β) rough path integral defined in 5.2 and the right-hand-side is defined as an Itô integral, as we have seen in 4.3. Thus, the definitions of the (α,β) rough path integral and the (α,β) lifting of [17] *allow for consistence between the stochastic (Itô) integral and the (α,β) rough path (first level) definition of $\int_0^t f(\hat{X}_{0r})dX_r$.*

Having these definitions in mind, we can now discuss the LDP on $\Omega_{(\alpha,\beta)-\mathrm{Hol}}$. As always, we consider the simple rough volatility model 2, so we have $X = \rho W + \bar{\rho}\bar{W}, \rho \in [-1,1]$ and $\bar{\rho} := \sqrt{1-\rho^2}$ and $(W,\bar{W})$ a 2-dimensional Brownian motion.

Theorem 5.6 (Theorem 3.2 of [17])**.** *Define $\hat{X}, X^{(i)}, X^{(jk)}$ as in 5.5 but with $d=1$, so that we really are in the framework of simple rough volatility 2. Let $\mathbb{X} := (\hat{X}, X^{(i)}, X^{(jk)})_{i\in I,(j,k)\in J}$ be the random variable taking values in $(\Omega_{(\alpha,\beta)-Hol}, \rho_{(\alpha,\beta)})$ defined as above. Then, the sequence of triplets*

$$\mathbb{X}^\epsilon := \left(\epsilon^H \hat{X}, \epsilon^{(i+1)H} X^{(i)}, \epsilon^{(j+k+2)H} X^{(jk)}\right)$$

satisfies the LDP on $(\Omega_{(\alpha,\beta)-Hol}, \rho_{(\alpha,\beta)})$ with speed ϵ^{-2H} with good rate function

$$I^{\#\#}(\hat{x}, x^{(i)}, x^{(jk)}) := \inf\{I^{\#}(u,v) : u,v \in C([0,T],\mathbb{R}), v \in C^{1-var}([0,T],\mathbb{R}), (\hat{x}, x^{(i)}, x^{(jk)}) = \mathbb{L}(u,v)\},$$

where $\mathbb{L}(u,v) := (\delta u, u\cdot v, u * v)$ for all $u,v \in C([0,T],\mathbb{R}), v \in C^{1-var}([0,T],\mathbb{R})$ where we have

$$u\cdot v = (u \cdot_i v), \ \ u * v = (u *_{jk} v), \ \ (\delta u)_{st} := u_t - u_s,$$

and

$$(u\cdot_i v)_{st} := \int_s^t (u_r - u_s)^i dv_r, \quad (u\cdot_{jk} v)_{st} := \int_s^t (u\cdot_j v)_{sr}(u_r-u_s)^k dv_r.$$

Here, $I^{\#}$ is the same as in [19]*:*

$$I^{\#}(u,v) := \begin{cases} \frac{1}{2}\|(u,v)\|^2_{\mathcal{H}^\psi}, & (u,v) \in \mathcal{H}^\psi, \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{H}^\psi := \{\mathcal{I}^\psi g : g \in L^2([0,T],\mathbb{R}^2)\}$ with inner product

$$\langle \mathcal{I}^\psi g, \mathcal{I}^\psi h\rangle := \langle g,h\rangle_{L^2},$$

and $\mathcal{I}^\psi : L^2([0,T],\mathbb{R}^2) \to L^2([0,T],\mathbb{R}^2)$ is defined, for all $g \in L^2([0,T],\mathbb{R}^2)$ by

$$\mathcal{I}^\psi g := \int_0^{\cdot} \psi(\cdot - u)g(u)du$$

with $\psi : \mathbb{R}^+ \to \mathbb{R}^{2\times 2}$ defined by

$$\psi := \begin{pmatrix} k_H & 0 \\ \rho & \bar{\rho}\end{pmatrix}.$$

Theorem 5.7 (Theorem 3.3 of [17])**.** *The sequence of processes $\{Y^\epsilon := \int f(\hat{\mathbb{X}}^\epsilon)d\mathbb{X}^\epsilon\}_{\epsilon\geq 0}$ satisfies the LDP on $(\mathcal{C}^\alpha([0,T],\mathbb{R}), \rho_{\alpha-Hol})$ with speed ϵ^{-2H} with good rate function*

$$I^{\#\#\#}(y) := \inf\left\{I^{\#\#}(\mathbb{X}) : \mathbb{X} \in \Omega_{(\alpha,\beta)-Hol}, y = \int f(\hat{\mathbb{X}})d\mathbb{X}\right\}$$
$$= \inf\left\{I^{\#}(u,v) : u,v \in C([0,T],\mathbb{R}), v \in C^{1-var}([0,T],\mathbb{R}), y = \int f(\hat{\mathbb{L}}(u,v))d\mathbb{L}(u,v)\right\}.$$

where $I^{\#\#}$ is defined in 5.6.

Proof. By Theorems 5.3 and 5.6 together with the contraction principle 3.5, we have the claim. □

We now discuss the following type of RDE in Lyons' sense (2.41 or Section 8.8 of [12]):

$$\bar{S}_t = \int_0^t \bar{\sigma}(\bar{S}_u, u)dY_u, \qquad Y := \int f(\hat{\mathbb{X}})d\mathbb{X} \in \mathcal{C}^\alpha([0,T],\mathbb{R}^d), \tag{34}$$

where $\bar{S}_t = S_t - S_0, \bar{\sigma}(s,t) = \sigma(S_0+s,t)$.

In the following theorems of [17], we are back to the more general case of d not necessarily equal to 1. Since, as we said in the beginning of this section, this is a more general type of SDE than the simple rough volatility model, we will not give more details about these results. But we state them to give the idea of the approach, that is to proceed as in the classical rough path theory, starting from an LDP on the driving noise and using the continuity of the solution map of RDEs to derive LDPs on the solution of the SDE.

Theorem 5.8 (Theorem 3.4 of [17])**.** *Let $\sigma \in C_b^3(\mathbb{R}^{d+1},\mathbb{R})$ (the set of 3 times continuously differentiable bounded functions from $\mathbb{R}^{d+1}$ to $\mathbb{R}$).*

(i) The RDE (34) *driven (in Lyons' sense) by $Y = \int f(\hat{\mathbb{X}})d\mathbb{X}$ has a unique solution. Moreover, the solution map ϕ,*

$$\phi : \mathcal{C}^\alpha([0,T],\mathbb{R}^d)\times\mathbb{R} \to \mathcal{C}^\alpha([0,T],\mathbb{R}^{d+1})$$

is Lipschitz on bounded sets with respect to $\rho_{\alpha-Hol}$.

(ii) The first level (in the sense of Lyons) of the solution to RDE (34) *is the solution to the Itô SDE* (30)*.*

Theorem 5.9 (Theorem 3.5 of [17])**.** *Let $\sigma \in C_b^3(\mathbb{R}^{d+1},\mathbb{R})$ and $\bar{S}^\epsilon := \phi(Y^\epsilon)$, where ϕ is the solution map defined in Theorem 5.8. Then the sequence of processes $\{\bar{S}^\epsilon\}_{\epsilon\geq 0}$ satisfies the LDP on $\mathcal{C}^\alpha([0,T],\mathbb{R}^d)$ with speed ϵ^{-2H} with the good rate function I, defined as*

$$I(\bar{s}) := \inf\{I^{\#\#\#}(Y) : Y \in \mathcal{C}^\alpha([0,T],\mathbb{R}^d), \bar{s} = \phi(Y)\}$$
$$= \inf\left\{I^{\#}(u,v) : u,v \in C([0,T],\mathbb{R}), v \in C^{1-var}([0,T],\mathbb{R}), \bar{s} = \int \bar{\sigma}(\bar{s},\cdot) f(\hat{\mathbb{L}}(u,v)) d\mathbb{L}(u,v)\right\}.$$

Proof. Since, by Theorem 5.8, the solution map ϕ is continuous, Theorem 5.7 and the contraction principle imply the claim. □

The rest of the results of [17] are short asymptotics statements deduced from the above LDP on the sequence $\{Y^\epsilon := \int f(\hat{\mathbb{X}}^\epsilon) d\mathbb{X}^\epsilon\}_{\epsilon\geq 0}$ of Theorem 5.7, and the extended contraction principle together with exponentially good approximations of the sequence $\{Y^\epsilon\}_{\epsilon\geq 0}$.

6. An enhanced Sanov Large Deviation Principle

The framework of Deuschel et al's paper [7] is the same as 1.1. Fix a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$.

6.1. **Summary of** [7]**.** Given a Polish space (Σ, d_Σ), we denote by $P_1(\Sigma)$ the space of probability measures on Σ with finite first moment, i.e. the probability measures μ that satisfy $\int_\Sigma d_\Sigma(x,x_0)\mu(dx) < +\infty$ for some (equivalently for all) $x_0 \in \Sigma$. It is a Polish space endowed with the 1-Wasserstein distance d_W, namely

$$d_W(\mu,\nu) = \inf_{\pi\in\Gamma(\mu,\nu)}\left\{\int_{\Sigma\times\Sigma} d_\Sigma(x^1,x^2)\pi(d(x^1,x^2))\right\},$$

where $\Gamma(\mu,\nu)$ is the set of all probability measures on $\Sigma\times\Sigma$ with the first marginal and the second marginal equal respectively to μ and ν. Whenever Σ is some (Polish) space of α-Hölder continuous (rough) paths, cf. 2.33, we shall write $d_{W,\alpha}$ for the corresponding 1-Wasserstein distance.

Here is a new version of Sanov's theorem 3.10 with a topology different from the original weak convergence topology. The interest for this new topology is that it allows to derive LDPs on the empirical measures of interacting particle systems, which is done in [7]. Any LDP can be stated equivalently talking about measures or talking about probability measures and the following Sanov theorem is stated with random variables.

Theorem 6.1 (Sanov Theorem in Wasserstein metric, Theorem 3.2 of [7])**.** *Let (Σ,d) be a Polish space and let $(X^i)_i$ be a sequence of Σ-valued i.i.d. random variables with same law μ. Assume that μ satisfies the following condition: there exists a function $G: E\to[0,+\infty]$, with compact sublevel sets (in particular lower semi-continuous), with more than linear growth (i.e., for some x_0, $\frac{G(x)}{d(x,x_0)} \to +\infty$ as $d(x,x_0)\to\infty$) such that*

$$\int_\Sigma e^G d\mu < \infty. \tag{35}$$

Then the sequence of laws of the empirical measures (same definition as in (24)

$$L_n^X := \frac{1}{n}\sum_{i=1}^n \delta_{X^i}$$

satisfies a large deviation principle on $P_1(\Sigma)$, endowed with the 1-Wasserstein metric, with rate (speed) n and good rate function $H(\cdot|\mu)$.

This result differs from the classical Sanov theorem by the fact that it involves the 1-Wasserstein metric, while classical Sanov theorem involves C_b-weak topology. In this, the statement above is stronger, but does need the additional condition on the measure μ.

Remark 6.2 (Remark 3.3 of [7])**.** *In the case $\Sigma = \mathcal{C}^\alpha([0,T],\mathbb{R}^d), \alpha < 1/2$, the assumption* (35) *above is satisfied by $\{B^i : i\in\mathbb{N}\}$ (independent Brownian motions starting from the distribution λ), if λ verifies Condition* (36) *of* [7]*. Indeed one can take*

$$G(\gamma) := c\left(\sup_{0\leq s<t\leq T}\frac{d(\gamma(t),\gamma(s))}{|t-s|^\beta}\right)^{1+\epsilon} + c|\gamma(0)|^{1+\epsilon},$$

where β is in $(\alpha,1/2)$ and c,ϵ are the same of Condition (36) *of* [7]*. This G has compact sublevel sets and more than linear growth; Condition* (35) *is verified since ($B^1(x=0)$ is the Brownian motion starting at 0)*

$$\mathbb{E}[e^{G(B^1)}] = \mathbb{E}[\exp(c\|B^1(x=0)\|_{\beta-Hol}^{1+\epsilon})]\int_{\mathbb{R}^d} e^{c|x|^{1+\epsilon}}\lambda(dx) < \infty$$

by Condition (36) *of* [7] *and exponential integrability of $c\|B^1(x=0)\|_{\beta-Hol}^{1+\epsilon}$ (a consequence for example of 3.18.*

Proof. The assertion of 6.1 is a consequence of classical Sanov theorem in the weak convergence topology 3.10 and the inverse contraction principle, provided we prove exponential tightness, in 1-Wasserstein metric, of the laws of the empirical measures L_n^X, which is done in [7] using Markov inequality. □

Define $C^{0,\alpha-Hol}([0,T],\mathbb{R}^d)$ as the closure of smooth α-Hölder paths with respect to the homogenous α-Hölder distance and recall we defined the space of geometric α-Hölder rough paths in 2.35 as $\mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^d)$.

Definition 6.3. *Define $S^{\{d\}} : C^{0,\alpha-Hol}([0,T],\mathbb{R}^d) \to \mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^d)$ as follows. Consider the sequence of piecewise linear approximations X^k along the dyadic partitions $D_k := \{2^{-k}i, 2^{-k}(i+1), i \in \mathbb{N}\}$ for $k \in \mathbb{N}$ and set*

$$A^k_{s,t} = \int_s^t X^k_{s,r} \otimes dX^k_r.$$

Whenever $S^k := (X^k, A^k)$ is Cauchy in ρ_α metric, set

$$S(X) := (X, A(X)) := \lim_{k\to\infty} (X^k, A^k)$$

and zero elsewhere. By construction, $S(X)$ is in $\mathcal{C}^\alpha([0,T],\mathbb{R}^d)$ and actually in $\mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^d)$ (because $X^k \in C^{1-var}([0,T],\mathbb{R}^d)$) and $X \mapsto S(X)$ is a well-defined measurable (but in general discontinuous) map on the path space.

Remark 6.4. *For $X = B$ being a Brownian motion is equivalent (see 3.15) to define $S(B)$ as the Stratonovitch enhancement of the Brownian motion. Indeed, we know, from 3.20, that, if $\{B^k\}$ is a sequence of piecewise linear approximations of Brownian motions along dyadic partitions then we have $\lim_{k\to\infty} S(B^k) = (B, \mathbb{B}^{Strat})$ with $\mathbb{B}^{Strat}$ being the Stratonovitch enhancement of B, that is $\mathbb{B}^{Strat}_t = \int 0^t B \otimes \circ dB$.*

Remark 6.5. *It is also equivalent to define S as the lifting of a d-dimensional Brownian motion to a geometric rough path in 3.14, that is*

$$S(B) = \exp(B + A),$$

where A denotes the (antisymmetric, careful to not make the confusion with the A^k in 6.3) Lévy area (defined as an Itô integral) associated with B.

Remark 6.6. *An essential idea in* [7] *is that this enhancement of the Brownian motion, in the Stratonovitch way or as a limit of enhancement smooth approximations, can be generalised to a concatenation of any number of Brownian motions. We will focus on the case of the concatenation of two Brownian motions first.*

6.2. **The enhanced Sanov LDP for the 2-enhanced empirical measure.** We can now state and depict the steps of the proof of the large deviation result for the sequence of the enhanced empirical measure $\{\mathbf{L}^{\mathbf{B}}_n : n \in \mathbb{N}\}$.

Let B^i, B^j be two independent d-dimensional Brownian motions. As announced in 6.6, we call B^{ij} the path (B^i, B^j) (which is the concatenation of B^i, B^j) and $\mathbf{B}^{ij} = (B^{ij}, \mathbb{B}^{ij}) := S(B^{ij})$ the corresponding rough path lift of B^{ij} with S defined in 6.3.

The enhanced empirical measure $\mathrm{L}^{\mathrm{B}}_n$ is defined as

$$\mathrm{L}^{\mathrm{B}}_n := \frac{1}{n^2} \sum_{i,j=1}^n \delta_{(B^{ij}, \mathbb{B}^{ij})}.$$

Let us define, based on the definition of S in 6.3, the map F as

$$\begin{aligned} F : P_1(C^{0,\alpha}([0,T],\mathbb{R}^d)) &\to P_1(\mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^{2d})) \\ Q &\mapsto (Q \otimes Q) \circ S^{-1} \end{aligned}$$

where $(Q \otimes Q) \circ S^{-1}$ is the pushforward measure $S_*(Q \otimes Q)$.

Let us also define the projection $\pi_1 : \mathbf{X} \mapsto X^1$ for any element $\mathbf{X} = ((X^1, X^2), \mathbb{X}) \in \mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^{2d})$.

Let us denote the d-dimensional Wiener measure $P^{\{d\}}$.

We can now state the main result of [7] in the case of the concatenation of 2 (instead of $k \in \mathbb{N}$) paths.

Theorem 6.7 (2-enhanced Sanov theorem, Theorem 3.6 of [7])**.** *Let $\{B^i : i \in \mathbb{N}\}$ be a family of independent d-dimensional Brownian motions, with initial measure λ and assume that there exists $c, \epsilon > 0$ such that*

$$\int_{\mathbb{R}^d} e^{c|x|^{1+\epsilon}} \lambda(dx) < \infty. \tag{36}$$

The family $\{\boldsymbol{L}^{\boldsymbol{B}}_n{}_*\mathbb{P} : n \in \mathbb{N}\}$ satisfies a LDP on $P_1(\mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^{2d}))$ endowed with the 1-Wasserstein metric, with scale (or rate or speed) n and good rate function $\boldsymbol{I}$ given by

$$\boldsymbol{I}(\mu) = \begin{cases} H(\mu \circ \pi_1^{-1} | P^{\{d\}}), & \text{if } \mu = F(\mu \circ \pi_1^{-1}), \\ \infty, & \text{otherwise.} \end{cases} \tag{37}$$

The basic fact that invites the authors to use the extended contraction principle to prove 6.7 is the following lemma

Lemma 6.8 (Lemma 3.7 of [7])**.** *The enhanced empirical measure $\boldsymbol{L}^{\boldsymbol{B}}_n$ is almost surely the image of the (true) empirical measure L^B_n under the map $F : Q \mapsto (Q \otimes Q) \circ S^{-1} = S_*(Q \otimes Q)$.*

Proof. We have that $\mathbf{L}_n^{\mathbf{B}}$ is the image of $\mathrm{L}_n^{\mathrm{B}}$ under F. In the proof of 6.8, when they say "the image measure of $\mathrm{L}_n^{\mathrm{B}}$ under F is given by", one should understand "$F(\mathrm{L}_n^{\mathrm{B}})$". Then, there is a subtlety in understanding the formula giving F. $F(Q) = (Q \otimes Q) \circ S^{-1}$ for $Q \in \mathcal{P}_1(\mathcal{C}^{0,\alpha}([0,T],\mathbb{R}^d))$ means for $\mathrm{L}_n^{\mathrm{B}}$, that $\mathrm{L}_n^{\mathrm{B}} \otimes \mathrm{L}_n^{\mathrm{B}} = \frac{1}{n^2}\sum_{i,j=1}^n \delta_{\mathrm{B}^{ij}}$ and then, we consider $\mathrm{L}_n^{\mathrm{B}} \otimes \mathrm{L}_n^{\mathrm{B}}$ like a function on $\mathcal{C}^{0,\alpha}([0,T],\mathbb{R}^{2d})$. So

$$\mathrm{L}_n^{\mathrm{B}} \otimes \mathrm{L}_n^{\mathrm{B}} \circ S^{-1} = \frac{1}{n^2}\sum_{i,j=1}^n \delta_{\mathrm{B}^{ij}} \circ S^{-1} = \frac{1}{n^2}\sum_{i,j=1}^n \delta_{S(\mathrm{B}^{ij})} = \frac{1}{n^2}\sum_{i,j=1}^n \delta_{\mathbf{B}^{ij}},$$

that is to say that the formula true for almost all ω: $\mathbf{L}_n^{\mathbf{B}(\omega)} = \frac{1}{n^2}\sum_{i,j=1}^n \delta_{S(\mathrm{B}^{ij}(\omega))}$. □

In order to apply the extended contraction principle, one must introduce a continuous approximation F_m to the map F. Given a trajectory Y, we define its piecewise linear approximation along the uniform dissection of $[0,T]$ in $\lfloor Tm \rfloor$ intervals as

$$Y^{(m)}(t) = Y\left(\frac{\lfloor mt \rfloor}{m}\right) + m\left(Y\left(\frac{\lfloor mt \rfloor + 1}{m}\right) - Y\left(\frac{\lfloor mt \rfloor}{m}\right)\right)\left(t - \frac{\lfloor mt \rfloor}{m}\right).$$

The iterated integral of $Y^{(m)}$ is defined as a Riemann-Stieltjes integral, more precisely

$$(\mathbb{Y}^{(m)}(t))^{ij} = \int_0^t Y_s^{(m),i} dY_s^{(m),j}$$

Now we set F_m as

$$\begin{aligned} F_m : P_1(C^{0,\alpha}([0,T],\mathbb{R}^d)) &\to P_1(\mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^{2d})) \\ Q &\mapsto (Q \otimes Q) \circ (S^{(m)})^{-1} \end{aligned}$$

where

$$\begin{aligned} S^{(m)} : C^{0,\alpha}([0,T],\mathbb{R}^{2d}) &\to \mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^{2d}) \\ Y &\mapsto (Y^{(m)},\mathbb{Y}^{(m)}) := S(Y^{(m)}) \end{aligned} \tag{38}$$

Note that this $S^{(m)}$ is defined as S^k, but replacing the dyadic approximation with the approximation at step $1/m$. We denote $\mathbf{L}_n^{\mathbf{B}(m)}$ the enhanced empirical measure associated with $B^{(m)}$, namely $\mathbf{L}_n^{\mathbf{B}(m)} = F_m(\mathrm{L}_n^{\mathrm{B}})$. Notice that, for each m, $S^{(m)}$ is continuous with at most linear growth (this is due to the use of the homogenous distance $d_{\alpha-Hol}$) and the map $Q \mapsto Q \otimes Q$ is continuous with respect to the 1-Wasserstein metric on $P_1(C^{0,\alpha}([0,T],\mathbb{R}^d))$ and $P_1(C^{0,\alpha}([0,T],\mathbb{R}^{2d}))$ (Lemma A.4 of [7]). So F_m is continuous in the 1-Wasserstein metric (by Corollary A.2 of [7]).

In the proceeding lemmata, Deuschel et al. show that the approximation given by F_m is indeed exponentially good, in the sense 3.7 of to be able to apply the extended contraction principle 3.8.

The main tool is the following lemma, which follows from 3.19.1, which gives an exponential bound for the approximation.

Lemma 6.9 (Lemma 3.8 of [7]). *Let B be a Brownian motion on on $\mathbb{R}^e$ and $\boldsymbol{B}$ be its Stratonovitch enhanced Brownian motion, let $\boldsymbol{B}^{(m)}$ be the lifting of the piecewise linear approximation $B^{(m)}$ to a geometric rough path, defined in* (38). *Fix $\alpha \in (0,1/2)$. Then, for every $\eta \in (0,1/2-\alpha)$, there exists $c > 0$ such that*

$$\sup_{m\geq 1} \mathbb{E}[\exp(cm^{\eta/2} d_{\alpha-Hol}(\boldsymbol{B},\boldsymbol{B}^{(m)})] < \infty.$$

As a first step, we establish the exponential tightness of the approximation $\mathbf{L}_n^{\mathbf{B}(m)}$ of $\mathbf{L}_n^{\mathbf{B}}$.

Lemma 6.10 (Lemma 3.9 of [7]). *For any $\delta > 0$, it holds*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}[d_{W,\alpha}(\boldsymbol{L}_n^{\boldsymbol{B}},\boldsymbol{L}_n^{\boldsymbol{B}(m)}) > \delta] = -\infty.$$

Proof. The proof is based on a Hoeffding's decomposition

We expect all the argument to still hold for our multidimensional enhancement of (α,β) rough paths. □

Lemma 6.11 (Lemma 3.10 of [7]). *For every $a < \infty$, it holds*

$$\lim_{m\to\infty} \sup_{Q:H(Q|P^{\{d\}})\leq a} d_{W,\alpha}(F_m(Q),F(Q)) = 0.$$

Proof. The proof is done using Orlicz spaces and the classical Orlicz-Birnbaum estimate. We expect the arguments to still hold in our (α,β) rough paths framework. □

We can now prove Theorem 6.7. Sanov theorem in Wasserstein distance 6.1 is true for an i.i.d sequence of d-dimensional Brownian motions (with common law the Wiener measure $P^{\{d\}}$) by 6.2. Thus, we have that $\mathrm{L}_n^{\mathrm{B}}$ satisfies an LDP in Wasserstein distance with good rate function $H(\cdot|P^{\{d\}})$. Condition 3.8 of the extended contraction principle 3.8 is satisfied using 6.11. The exponential tightness of the approximations assumption of the extended contraction principle is proven in 6.10. Thus, by extended contraction principle, $\{\mathbf{L}_n^{\mathbf{B}}{}_*\mathbb{P} : n \in \mathbb{N}\}$ satisfies a LDP on $P_1(\mathcal{C}_g^{0,\alpha}([0,T],\mathbb{R}^{2d}))$ endowed with the 1-Wasserstein metric, with scale (or rate or speed) n and good rate function $\boldsymbol{I}$ given by

$$\mu \mapsto \inf\left\{H(P|Q)\middle| Q \in P_1(C^{0,\alpha}([0,T],\mathbb{R}^d)) \text{ and } F(Q) = \mu\right\}$$

We check easily that this function coincides with the one given in (37).

Remark 6.12. *All of the above generalises to a k (instead of 2)-layer enhanced Brownian motion, where the k-layer process is $B^{\{k\};i_1,\dots,i_k} := (B^{i_1},\dots,B^{i_k})$, which has the rough path lift $\boldsymbol{B}^{\{k\};i_1,\dots,i_k}$ and the enhanced k-layer empirical measure*

$$\boldsymbol{L}_n^{\boldsymbol{B};\{k\}}(\omega) := \frac{1}{n^k}\sum_{1\le i_1,\dots,i_k}\delta_{\boldsymbol{B}^{\{k\};i_1,\dots,i_k}(\omega)}$$

The proof of the k-enhanced Sanov theorem for is done in Section 5 of [7].

Part 2. Proving an enhanced Sanov LDP for (α,β) rough paths

The main objective of this research project was to prove an enhanced Sanov theorem analogous to 6.7 with (α,β) rough path enhancement instead of the usual Brownian rough path enhancement.

The first step was to define a N-layer driving noise generalising $(\hat{X},X)$ of 2 and the lifting to an (α,β) rough path of such a N-layer noise. In the case of Brownian rough paths, it was straightforward from the definition of the lifting of a d-dimensional Brownian motion to define the lifting of the concatenation of i.i.d. Brownian paths. This imposed to define a N-dimensional Chen's relation and the Hölder regularity of a N-dimensional (α,β) rough path, which is the same as in dimension 1. Our generalisation is conceptually non-trivial because it allows to generalise the SDEs considered in rough volatility to i.i.d. sequences of driving noises.

That being done, our objective has been to prove a strong approximation theorem similar to 3.19 (and 3.19.1) or 3.26 in the case of N-dimensional (α,β) rough paths.

This would allow to show a strong approximation lemma, analogous to 6.9 but for (α,β) rough paths. Then, as we have mentioned in the summary of the proofs of 6.10 and 6.11, we expect the arguments allowing to prove these lemmas to also hold, without too much work, provided we have an analog of 6.9 for (α,β) rough paths.

The most promising path to prove such a strong approximation lemma seemed to be to follow the steps of the proof of 3.19 (Proposition 13.20 in [15]). We were able to prove the analog to Step 2 of Proposition 13.20 in [15], that is the small interval case.

Trying to prove the analog to Step 1, that is the case where $s,t\in D$ has led me down a blind alley, to such an extent that I am not anymore sure the property I am trying to prove is true.

The other promising way to prove our strong approximation lemma was to inspire ourselves from the stochastic integration technique developed in Chapter 15 of [15], which allows to lift Gaussian processes with independent components whose covariance's ρ-variation is controlled by a 2D-control ω as in 3.22 for $\rho\in[1,2)$. Yet, it seems that controlling the covariance of the powers $(\hat{X}_{sr})^i$ for $i\in I$ defined in [17] is not natural and also leads down a blind alley.

After that, I tried other paths, like proving a weak approximation estimate instead of a strong one.

7. Definition of N-dimensional (α,β) lifting

In what follows, fix $\alpha\in(0,1/2),\beta\in(0,1/2)$ and define $I:=I(\alpha,\beta)$ and $J:=J(\alpha,\beta)$ as in [17].

Definition 7.1. *An N-dimensional (α,β) rough path $\tilde{\mathbb{X}}=(\hat{X},\tilde{X}^{(i)},\tilde{X}^{(jk)})_{i\in I,(j,k)\in J}$ is a triplet of functions on Δ_T satisfying the following conditions; for any $i\in I,(j,k)\in J$ and $s\le u\le t$,*

(i) X is $\mathbb{R}^N$-valued, $\hat{X}$ is $\mathbb{R}^N$-valued, $X^{(i)}$ is $(\mathbb{R}^N)^{\otimes 2}$-valued, and $X^{(jk)}$ is $(\mathbb{R}^N)^{\otimes 4}$-valued.

(ii) Modified Chen's relation: $X_{st}=X_{su}+X_{ut},\hat{X}_{st}=\hat{X}_{su}+\hat{X}_{ut}$, and for all $a,b\in[\![1,N]\!]$

$$\tilde{X}^{(i),ab}_{st}=\tilde{X}^{(i),ab}_{su}+\sum_{p=0}^{i}\frac{1}{(i-p)!}(\hat{X}^a_{su})^{i-p}\tilde{X}^{(p),ab}_{ut}.$$

and for all $a,b,c,d\in[\![1,N]\!]$

$$\tilde{X}^{(jk),ab,cd}_{st}-\tilde{X}^{(jk),ab,cd}_{su}=\tilde{X}^{(j),cd}_{su}\sum_{q=0}^{k}\frac{(\hat{X}^a_{su})^{k-q}}{(k-q)!}\tilde{X}^{(q),ab}_{ut}+\sum_{p=0}^{j}\frac{(\hat{X}^c_{su})^{j-p}}{(j-p)!}\sum_{q=0}^{k}\frac{(\hat{X}^a_{su})^{k-q}}{(k-q)!}\tilde{X}^{(pq),ab,cd}_{ut}.$$

(iii) Hölder regularity: there exists a constant C such that, for all $(s,t)\in\Delta_T$,

$$|\hat{X}_{s,t}|\le C(t-s)^{\beta},\quad |X^{(i)}_{s,t}|\le C(t-s)^{i\beta+\alpha},\quad |X^{(jk)}_{s,t}|\le C(t-s)^{(j+k)\beta+2\alpha}.$$

Let $\Omega^N_{(\alpha,\beta)-Hol}$ denote the set of the N-dimensional (α,β) rough paths.

$\tilde{X}^{(0)}$ is a $N\times N$ matrix where each line is the same and will correspond to $(X^1,\dots,X^N)$ in 7.3. On the other hand, $\tilde{X}^{(00)}$ is made of N^2 identical matrices of size $N\times N$, each of them will be equal to the matrix of the iterated integrals of X against itself in 7.3.

We want to stress the fact that $(\hat{X},\tilde{X}^{(0)})$ is the process we lift (but with an N times repetition of half of the original process), the first level of the (α,β) rough path. One might improve the construction in 7.1 to avoid such a repetition or give meaning to it. Yet, it does not bar us to propose a meaningful generalisation of (α,β) rough paths.

Definition 7.2. *Let us denote by $\pi_i:\mathbb{R}^{2N}\bigoplus_{i\in I\backslash\{0\}}(\mathbb{R}^N)^{\otimes 2}\bigoplus_{(jk)\in J}(\mathbb{R}^N)^{\otimes 4}\to(\mathbb{R}^N)^{\otimes 2}$ the projection of an element of $\Omega^N_{(\alpha,\beta)-Hol}$ onto its ith level for $i\in I\backslash\{0\}$ level and $\pi_{jk}:\mathbb{R}^{2N}\bigoplus_{i\in I\backslash\{0\}}(\mathbb{R}^N)^{\otimes 2}\bigoplus_{(jk)\in J}(\mathbb{R}^N)^{\otimes 3}\to(\mathbb{R}^N)^{\otimes 3}$ the projection onto its (jk)th level and π_0 the projection onto the first level driving noise $((\hat{X}^1,X^1),\dots,(\hat{X}^N,X^N))$. That is a new meaning for the notation π_i instead of the usual projection on the ith level of the truncated tensor algebra that it always means in* [15]. *Recall that it meant π_k was the projection from $T^N(\mathbb{R}^d)\to(\mathbb{R}^d)^{\otimes k}$.*

Let $g \in \Omega^N_{(\alpha,\beta)-Hol}$. We denote $\pi_{00}(g)$ the N components of $\pi_0(g)$ that must satisfy the β-Hölder regularity condition, and $\pi_{01}(g)$ the N components of $\pi_0(g)$ that must satisfy the $(0\beta+\alpha=)\alpha$-Hölder regularity condition.

Proposition 7.3. *Denote:*

$$X = (X^1, \dots, X^N), \qquad \hat{X} = (\hat{X}^1, \dots, \hat{X}^N)$$

with $(X^i, \hat{X}^i)_{i\le N}$ i.i.d. and, for each $i \in [\![1,N]\!]$, $\hat{X}^i$ a 1-dimensional fractional Brownian motion possibly correlated to X^i, a 1-dimensional standard Brownian motion. We denote $X^i = \rho W^i + \bar{\rho}\bar{W}^i$ with $(W^i, \bar{W}^i)$ being a 2-dimensional Brownian motion, $\bar{\rho} = \sqrt{1-\rho^2}$ and $\hat{X}^i_t := \int_0^t k_H(t-s)dW^i_s$, $H \in (0,1)$ similarly to 2. Define, for all $(s,t) \in \Delta_T$, $i \in I, (j,k) \in J$

$$\tilde{X}^{(i)}_{st} = \begin{pmatrix} \tilde{X}^{(i),11}_{st} & \dots & \tilde{X}^{(i),1N}_{st} \\ \vdots & & \vdots \\ \tilde{X}^{(i),N1}_{st} & \dots & \tilde{X}^{(i),NN}_{st} \end{pmatrix} = \begin{pmatrix} \frac{1}{i!}\int_s^t (\hat{X}^1_{sr})^i dX^1_r & \dots & \frac{1}{i!}\int_s^t (\hat{X}^1_{sr})^i dX^N_r \\ \vdots & & \vdots \\ \frac{1}{i!}\int_s^t (\hat{X}^N_{sr})^i dX^1_r & \dots & \frac{1}{i!}\int_s^t (\hat{X}^N_{sr})^i dX^N_r \end{pmatrix} \in (\mathbb{R}^N)^{\otimes 2},$$

and, similarly, define

$$\tilde{X}^{(jk)}_{st} := \frac{1}{k!}\int_s^t (\hat{X}_{sr})^k \otimes \tilde{X}^{(j)}_{sr} \otimes dX_r \in (\mathbb{R}^N)^{\otimes 4},$$

that is, for all $a,b,c,d \in [\![1,N]\!]$

$$\tilde{X}^{(jk),ab,cd}_{st} = \frac{1}{k!}\int_s^t (\hat{X}^a_{sr})^k \tilde{X}^{(j),cd}_{sr} dX^b_r.$$

Then we have that, for almost every $\omega \in \Omega$, $\tilde{\mathbb{X}}(\omega) := (\hat{X}(\omega), \tilde{X}^{(i)}(\omega), \tilde{X}^{(jk)}(\omega))_{i\in I(\alpha,\beta),(j,k)\in J(\alpha,\beta)}$ is an N-dimensional (α,β) rough path for any $\alpha \in (0,1/2), \beta \in (0,H)$.

This allows us, for all $\alpha \in (0,1/2), \beta \in (0,H)$ to define the lift of $(X,\hat{X})$ to a N-layer (α,β) rough path as:

$$\tilde{S}^N : C^{\alpha-Hol}([0,T],\mathbb{R}^N) \times C^{\beta-Hol}([0,T],\mathbb{R}^N) \to \Omega^N_{(\alpha,\beta)-Hol}$$
$$(X,\hat{X}) = (X^1,\dots,X^N,\hat{X}^1,\dots,\hat{X}^N) \mapsto (\hat{X}, \tilde{X}^{(i)}_{i\in I}, \tilde{X}^{(jk)}_{(j,k)\in J})$$

where $\Omega^N_{(\alpha,\beta)-Hol}$ is the set of the N-layer (α,β)-rough paths defined above.

Lemma 7.4. *Consider the $\mathbb{R}^{2N}$-valued continuous centred Gaussian process $(X,\hat{X}) := ((\hat{X}^1,X^1),\dots,(\hat{X}^N,X^N))$ and its lifting to an N-dimensional (α,β) rough path $\tilde{S}^N(X,\hat{X}) \in \Omega^N_{(\alpha,\beta)-Hol}$. Then*

(i) for $i \in I\setminus\{0\}$, $\pi_i(\tilde{S}^N(X,\hat{X}))$ belongs to the $i+1$-Wiener chaos
(ii) for $(jk) \in J$, $\pi_{jk}(\tilde{S}^N(X,\hat{X}))$ belongs to the $k+j+2$-Wiener chaos
(iii) $\pi_0(\tilde{S}^N(X,\hat{X}))$ belongs to the 1-Wiener chaos

Proof. (i) In the proof of Proposition 3.1 of [17], we have:

$$(\hat{X}_{sr})^i = (Z^r_{sr})^i = \sum_{q=0}^{\lfloor \frac{i}{2}\rfloor} \frac{i!}{2^q} F^{(i-2q)}_r G^{(q)}_r$$

Thus, with $X = (X^1,\dots,X^N)$ and $\hat{X} = (\hat{X}^1,\dots,\hat{X}^N)$ we have that F and G are now N-dimensional, each component corresponding to each $\hat{X}^a, a \in [\![1,N]\!]$ and:

$$\pi_i(\tilde{S}^N(X,\hat{X})) = \tilde{X}^{(i)}_{st} = \sum_{q=0}^{\lfloor \frac{i}{2}\rfloor} \frac{1}{2^q}\int_s^t F^{(i-2q)}_r \odot G^{(q)}_r \otimes dX_r.$$

In our case, $\hat{X}$ and X are respectively N-dimensional, hence the tensor product and the $\odot$ in the definition.

For each term, it suffices to count the number of "noises" that appear in the expression. Here $F^{(i-2q)}$ has $i-2q$ multiples of Z, G is just a multiple Riemann integral so 0 multiples and then the whole thing is integrated again wrt X so 1 more noise. In total, this results to Wiener chaos of order $i-2q+1$. Therefore, $\pi_i(\tilde{S}^N(X,\hat{X}))$ belongs to the Wiener chaos of order $i-2\times 0+1 = i+1$.

(ii) Recall that, with $X = (X^1,\dots,X^N)$ and $\hat{X} = (\hat{X}^1,\dots,\hat{X}^N)$ we have defined $\pi_{jk}(\tilde{S}^N(X,\hat{X})) := \tilde{X}^{(jk)}$ as:

$$\tilde{X}^{(jk)}_{st} := \frac{1}{k!}\int_s^t (\hat{X}_{sr})^k \otimes \tilde{X}^{(j)}_{sr} \otimes dX_r$$

and

$$\begin{aligned}
\tilde{X}^{(jk)}_{st} := \frac{1}{k!}\int_s^t (\hat{X}_{sr})^k \otimes \tilde{X}^{(j)}_{sr} \otimes dX_r &= \frac{1}{k!}\int_s^t \sum_{p=0}^{\lfloor \frac{k}{2}\rfloor} \frac{k!}{2^p} F^{(k-2p)}_r G^{(p)}_r \otimes \sum_{q=0}^{\lfloor \frac{j}{2}\rfloor} \frac{1}{2^q}\int_s^r F^{(j-2q)}_v G^{(q)}_v \otimes dX_v \otimes dX_r \\
&= \int_s^t \sum_{p=0}^{\lfloor \frac{k}{2}\rfloor} \frac{1}{2^p} F^{(k-2p)}_r G^{(p)}_r \otimes \sum_{q=0}^{\lfloor \frac{j}{2}\rfloor} \frac{1}{2^q}\int_s^r F^{(j-2q)}_v G^{(q)}_v \otimes dX_v \otimes dX_r \\
&= \sum_{p=0}^{\lfloor \frac{k}{2}\rfloor}\sum_{q=0}^{\lfloor \frac{j}{2}\rfloor} \frac{1}{2^{p+q}} \int_s^t F^{(k-2p)}_r G^{(p)}_r \otimes \Big(\int_s^r F^{(j-2q)}_v G^{(q)}_v \otimes dX_v\Big) \otimes dX_r
\end{aligned}$$

The superscript on each $X, \hat{X}$ is in $\{1, \ldots, N\}$ and we do not write them for readability.

In a (p,q) term of the sum, $F^{(i-2q)}$ has $i-2q$ multiples of Z, G is just a multiple Riemann integral so 0 multiples. So we integrate $k-2p+j-2q = k+j-2(p+q)$ noise with respect to X twice so $k+j-2(p+q)+2$. So each (p,q) term of the sum is a $k+j-2(p+q)+2$-Wiener chaos.

Then, the one of highest order is for $p=q=0$ and is of order $k+j+2$.

(iii) We can extend what has been done in (i) to $i=0$, we get that $(\hat{X}_{sr})^1$ is a Wiener chaos of order $1-2\times 0+1$ because the sum $\sum_{q=0}^{\lfloor \frac{i}{2} \rfloor}$ only has one term.

Also, if we take $i=0$, we have that $\tilde{X}_{st}^{(0)} = X_{st}$ is a Wiener chaos of order $0+1$. □

Proof. The proof of the modified Chen's relations is done using binomial theorem and linearity of the integration just like in the introduction of [17].

For the regularity conditions, we proceed just like in the proof of (i) of Proposition 3.1 of [17] but endowing the space $(\mathbb{R}^N)^{\otimes 2}$ with the euclidean norm (2-norm) for the i levels and endowing the space $(\mathbb{R}^N)^{\otimes 4}$ with the euclidean norm (2-norm).

With $X = (X^1, \ldots, X^N)$ and $\hat{X} = (\hat{X}^1, \ldots, \hat{X}^N)$ we have that F and G of the proof of (i) of Proposition 3.1 of [17] are now N-dimensional, each component corresponding to each $\hat{X}^a, a \in [\![1, N]\!]$. Those N components of F and of G are i.i.d because the $\hat{X}^a, a \in [\![1, N]\!]$ are i.i.d and:

$$\pi_i(\tilde{S}^N(X, \hat{X})) = \tilde{X}_{st}^{(i)} = \sum_{q=0}^{\lfloor \frac{i}{2} \rfloor} \frac{1}{2^q} \int_s^t F_r^{(i-2q)} \odot G_r^{(q)} \otimes dX_r.$$

In our case, $\hat{X}$ and X are respectively N-dimensional, hence the tensor product and the $\odot$ in the definition.

Thus, using the Wiener chaos decomposition of the (α, β) rough path levels 7.4 we have that $\int_s^t F_r^{(i-2q),a} G_r^{(q),a} dX_r^b$ is a $i-2q+1$-Wiener chaos and, by hypercontractivity (Lemma 15.20 of [15]):

$$\left\| \int_s^t F_r^{(i-2q),a} G_r^{(q),a} dX_r^b \right\|_p \le \sqrt{i-2q+1}(p-1)^{(i-2q+2)/2} \left\| \int_s^t F_r^{(i-2q),a} G_r^{(q),a} dX_r^b \right\|_2$$

Then, we proceed just like in [17], that is using successive Itô isometries or, by using once Itô isometry and then the Gaussian integrability equivalence of Condition A.18 of [15], that is restated in 8.1.

We have the exact same approach for the (j,k) levels. □

8. Proving a strong approximation lemma of N-dimensional (α, β) rough paths

In what follows, we will sometimes, for notational simplicity, write $\|\cdot\|_p$ or $\|\cdot\|_{L^p}$ for the norm $\|\cdot\|_{L^p(\mathbb{P})}$.

8.1. **Notations.** To simplify, we try to prove an analog of Proposition 13.20 of [15] with the uniform dissection $D^m = \{\frac{kT}{m} : k = 0, \ldots, m\}$ with $m \in \mathbb{N}^*$ of $[0,T]$. The meshsize of such a dissection is $\frac{1}{m} = |D^m|$. Since we fix m, we will only note D instead of D^m in the proofs. Without loss of generality, let us consider $T=1$. For any point $t \in D^m$ of the dissection, the notation of [15] is that t_D and t^D are the closest points belonging to D such that $D \ni t_D \le t \le t^D \in D$.

Let $D = D^m$, we have, define the piecewise linear approximation $(X^D, \hat{X}^D)$ to $(X, \hat{X})$ as follows. For all $D \ni t_D \le t \le t^D \in D$:

$$X_t^D = X_{t_D} + \frac{t - t_D}{t^D - t_D} X_{t_D, t^D} = X_{t_D} + m(t - t_D) X_{t_D, t^D}$$

and likewise for the fractional BM part:

$$\hat{X}_t^D = \hat{X}_{t_D} + \frac{t - t_D}{t^D - t_D} \hat{X}_{t_D, t^D} = \hat{X}_{t_D} + m(t - t_D) \hat{X}_{t_D, t^D}$$

Thus, for all $D \ni r_D \le r \le r^D \in D$:

$$dX_r^D = \frac{dr}{r^D - r_D} X_{r_D, r^D} = m X_{r_D, r^D} dr \ , \ d\hat{X}_r^D = \frac{dr}{r^D - r_D} \hat{X}_{r_D, r^D} = m X_{r_D, r^D} dr$$

Denote $\tilde{X}_{s,t}^{D,(i),ab} := \pi_i(\tilde{S}^N(X^D, \hat{X}^D)_{s,t})^{a,b}$ for all $s < t$.

8.2. **Statement of the desired lemma.** If we fix $H \in (0, 1/2)$, $\alpha \in \left(\frac{1}{3}, \frac{1}{2}\right), \beta \in (0, H)$ ($H < \frac{1}{4}$ is the case we are interested in). We want to show the following property.

Property. *Fix $\alpha \in (0, \frac{1}{2})$ and $\beta \in (0, H)$. Then fix $\alpha' \in (\alpha, \frac{1}{2})$. For all $i \in I, (j,k) \in J, s < t \in [0,1]$ and $p \in [1, \infty)$,*

$$\|X_{st}^{(i)} - X_{st}^{D,(i)}\|_{L^p} \le p^{\frac{i+1}{2}} C(\alpha, \beta, H) |D|^{\frac{1}{2} - \alpha'} (t-s)^{\frac{2i\beta + 2\alpha'}{2}}$$
$$\|X_{st}^{(jk)} - X_{st}^{D,(jk)}\|_{L^p} \le p^{\frac{j+k+2}{2}} C(\alpha, \beta, H) |D|^{\frac{1}{2} - \alpha'} (t-s)^{(j+k)\beta + 2\alpha'}.$$

A similar property with any strictly positive power on $|D|$ would also be sufficient for our final objective. In this property, we have omitted the ˜ over the levels of the (α, β) rough path for notational simplicity as the proof for a N-dimensional (α, β) rough path should follow trivially from the proof for a "usual" 1-dimensional (α, β) rough path as defined in [17].

Using 18.6.1, this would allow to show

$$\|d_{(\alpha,\beta)-Hol}(\mathbf{X}, \tilde{S}^N(X^D, \hat{X}^D))\|_{L^p} \le C'(\alpha, \beta, H) |D|^{\frac{1}{2} - \alpha'} p^{\frac{1}{2}}$$

In what follows, we explain section our attempts to prove the desired property 8.2. We focus on the i levels, that is showing that, for all $i \in I, (j,k) \in J, s<t \in [0,1]$ and $p \in [1,\infty)$,

$$\|X^{(i)}_{st} - X^{D,(i)}_{st}\|_{L^p} \leq p^{\frac{i+1}{2}} C(\alpha,\beta,H)|D|^{\frac{1}{2}-\alpha'}(t-s)^{\frac{2i\beta+2\alpha'}{2}}.$$

Indeed, we were already stuck on this case and expect the proof for the (j,k) levels to be straightforward if we are able to show the 8.2 for the i levels.

8.3. **Gaussian integrability and trick for the definition of the sets I and J.** We will make extensive use of the following lemma in what follows without mentioning it.

Lemma 8.1. *The fBM β^H with Hurst parameter H satisfies Condition A.18 of* [15] *with $1/p = H$, that is, there exists $\eta > 0$ such that*

$$\sup_{0\leq s<t\leq T} \mathbb{E}\exp\left(\eta\left[\frac{d(X_s,X_t)}{|t-s|^{1/p}}\right]^2\right) < \infty,$$

or, equivalently (by Lemma A.17 of [15]*)*

$$\sup_{0\leq s<t\leq T}\left|\frac{d(X_s,X_t)}{|t-s|^{1/p}}\right|_{L^q} = O(\sqrt{q}) \quad \text{as } q\to\infty.$$

Proof. Let $0 \leq s < t$. By stationarity of the increments and self-similarity $\beta^H_{st} \sim \beta^H_{t-s} \sim (t-s)^H\beta^H_1$. By definition of a Gaussian process, β^H_1 is a Gaussian random variable. By Fernique's theorem, it has Gaussian tail. By Lemma A.17 of [15], it has square root growth of moments: $\|\beta^H_1\|_{L^q} \leq c_H\sqrt{q}$. Thus $\frac{\|\beta^H_{st}\|_{L^q}}{(t-s)^H} = \|\beta^H_1\|_{L^q} \leq c_H\sqrt{q}$. □

Remark 8.2. *Define $n(\alpha,\beta) = \max I(\alpha,\beta)$ with $I(\alpha,\beta)$ defined in Section 2.1 of* [17].

We have $n\beta+\alpha = 1$ if and only if $\frac{1-\alpha}{\beta}\in\mathbb{N}$. If we want to look at a Hölder regularity with a pair (α,β), in (iii) of Definition 2.1 of [17]*, and have $\frac{1-\alpha}{\beta}\in\mathbb{N}$, we can take $\frac{1}{2} > H > \beta' > \beta$ such that $\frac{1-\alpha}{\beta'}\notin\mathbb{N}$. This way, the Hölder regularity with the pair(α,β') will cover the Hölder regularities of the levels with the pair (α,β).*

Similarly, remember $J := \{(j,k)\in\mathbb{N}^2 : (j+k)\beta+2\alpha\leq 1\}$. Define $m(\alpha,\beta) = \max_{(j,k)\in J}(j+k)$. We have $m(\alpha,\beta)\in\mathbb{N}$ if and only if $\frac{1-2\alpha}{\beta}\in\mathbb{N}$.

If $\frac{1-\alpha}{\beta}\in\mathbb{N}$ or $\frac{1-2\alpha}{\beta}\in\mathbb{N}$, we take $\frac{1}{2} > H > \beta' > \beta$ such that $\frac{1-\alpha}{\beta'}\notin\mathbb{N}$ and $\frac{1-2\alpha}{\beta'}\notin\mathbb{N}$.

This way, without loss of generality, we can assume that we have $I(\alpha,\beta)$ such that $n(\alpha,\beta), m(\alpha,\beta) < 1$.

8.4. **Proof of small interval case.** Consider the case $s_D \leq s < t \leq s^D$. In the following we consider we are in the case of Remark 8.2.

We have:

$$\pi_i(\tilde{S}^N(X^D,\hat{X}^D)_{st})^{a,b} = \frac{1}{i!}\int_s^t \left(\hat{X}^{D,a}_{sr}\right)^i dX^{D,b}_r = \frac{1}{i!}mX^b_{s_D,s^D}\int_s^t\left(m(r-s)\hat{X}^a_{s_D,s^D}\right)^i dr$$
$$= \frac{1}{i!}m^{i+1}X^b_{s_D,s^D}(\hat{X}^a_{s_D,s^D})^i\int_s^t (r-s)^i dr = \frac{1}{i!}m^{i+1}X^b_{s_D,s^D}(\hat{X}^a_{s_D,s^D})^i\frac{(t-s)^{i+1}}{i+1}$$

By Cauchy-Schwarz, we have

$$\|X^b_{s_D,s^D}(\hat{X}^a_{s_D,s^D})^i\|_p \leq \|X^b_{s_D,s^D}\|_{2p}\|(\hat{X}^a_{s_D,s^D})^i\|_{2p}$$
$$\leq C(H,\alpha,\beta)p^{i/2}m^{-iH}p^{1/2}m^{-1/2} = C(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{iH+1/2}$$

with $C(H,\alpha,\beta)$ a constant depending only on H,α,β (and implicitly on T). Therefore:

$$\|\pi_i(\tilde{S}^N(X^D,\hat{X}^D)_{st})^{a,b}\|_p \leq C_2(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{iH+1/2}m^{i+1}(t-s)^{i+1}$$
$$= C_2(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{i(H-1)-1/2}(t-s)^{i+1} \leq C_2(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{iH-1/2}(t-s)$$

Remark 8.3. *We can interpolate: $t-s = (t-s)^{i\beta+\alpha'}(t-s)^{1-(i\beta+\alpha')} \leq (t-s)^{i\beta+\alpha'}(m^{-1})^{1-(i\beta+\alpha')}$ with $\alpha<\alpha'<\frac{1}{2}$ such that $n(\alpha,\beta)\beta+\alpha'<1$, remembering that we defined $n(\alpha,\beta) := \max I(\alpha,\beta)$.*

This gives

$$\|\pi_i(\tilde{S}^N(X^D,\hat{X}^D)_{st})^{a,b}\|_p \leq C_2(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{i(H-\beta)+\frac{1}{2}-\alpha'}(t-s)^{i\beta'+\alpha'}$$
$$C_2(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{\frac{1}{2}-\alpha'}(t-s)^{i\beta'+\alpha'},$$

with $\eta = \frac{1}{2}-\alpha'$ in $\big(0\vee(n(\alpha,\beta)\beta-\frac{1}{2});\frac{1}{2}-\alpha\big)$.

By triangle inequality and the previous estimates, we have

$$\|\tilde{X}^{(i),ab}_{st} - \tilde{X}^{D,(i),ab}_{st}\|_p \leq \|\tilde{X}^{(i),ab}_{st}\|_p + \|\tilde{X}^{D,(i),ab}_{st}\|_p$$
$$\leq C(H,\alpha,\beta)p^{\frac{i+1}{2}}(t-s)^{iH+1/2} + C(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{iH-1/2}(t-s)$$
$$\leq C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[(t-s)^{iH+1/2} + \Big(\frac{1}{m}\Big)^{iH-1/2}(t-s)\right].$$

We can now proceed to a second interpolation: $(t-s)^{iH+1/2} = (t-s)^{i\beta+\alpha'}(t-s)^{i(H-\beta)+\frac{1}{2}-\alpha'} \leq (t-s)^{i\beta+\alpha'}\left(\frac{1}{m}\right)^{i(H-\beta)+\frac{1}{2}-\alpha'}$. Therefore, we have

$$\begin{aligned} \|\tilde{X}^{(i),ab}_{st} - \tilde{X}^{D,(i),ab}_{st}\|_p &\leq C_2(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{i(H-\beta)+\frac{1}{2}-\alpha'}(t-s)^{i\beta+\alpha'} \\ &\leq C_2(H,\alpha,\beta)p^{(i+1)/2}\Big(\frac{1}{m}\Big)^{\frac{1}{2}-\alpha'}(t-s)^{i\beta+\alpha'}. \end{aligned} \tag{39}$$

We end up with a satisfying upper bound. Indeed, we have a positive power of the mesh $\left(\frac{1}{m}\right)^{\frac{1}{2}-\alpha'}$ because $\alpha' \in \left(0, \frac{1}{2}\right)$ and a Hölder exponent $i\beta + \alpha' > i\beta + \alpha$.

Remark 8.4. *It is essential to understand that the interpolation step allows us to "trade" some positive power of $(t-s)$ for some positive power of $|D| = \frac{1}{m}$. Thus, for the other cases, we just want to upper bound $\|\tilde{X}^{(i),ab}_{st} - \tilde{X}^{D,(i),ab}_{st}\|_2$ by $C(H,\alpha,\beta)\left(\frac{1}{m}\right)^{\theta}(t-s)^{\epsilon}$ with ϵ, θ big enough to allow to trade some positive power of $\min(|D|, t-s)$ for some positive power of $\max(|D|, t-s)$ to end up with the exponents on $|D|, t-s$ needed in 8.2.*

9. Problem with the case of endpoints belonging to dissection

Remark 9.1. *By the modified Chen's relation 7.1 applied to each term of the telescopic sum, we can prove the following relation for any $s, t \in [0,T]$:*

$$\tilde{X}^{(i),ab}_{st} = \sum_{l=c}^{d-1}\left(\tilde{X}^{(i),ab}_{s,t_{l+1}} - \tilde{X}^{(i),ab}_{s,t_l}\right) = \sum_{l=c}^{d-1}\sum_{j=0}^{i}\frac{1}{(i-j)!}\tilde{X}^{(j),ab}_{t_l,t_{l+1}}\left(\hat{X}^a_{t_c,t_l}\right)^{i-j}$$

The same reasoning gives, if $s = t_c, t = t_d$ belonging to the uniform dissection of $[0,1]$ in m intervals with $c < d \in \mathbb{N}^*$, that:

$$\tilde{X}^{D,(i),ab}_{st} = \sum_{l=c}^{d-1}\sum_{j=0}^{i}\frac{1}{(i-j)!}\tilde{X}^{D,(j),ab}_{t_l,t_{l+1}}\left(\hat{X}^{D,a}_{t_c,t_l}\right)^{i-j} = \sum_{l=c}^{d-1}\sum_{j=0}^{i}\frac{1}{(i-j)!}\tilde{X}^{D,(j),ab}_{t_l,t_{l+1}}\left(\hat{X}^{a}_{t_c,t_l}\right)^{i-j}$$

9.1. Naive heuristic attempt.

In this subsection, we are in the case $s, t \in D$. Let $s = t_c$ and $t = t_d$ for some $c < d$. We have $(d-c) = m(t-s)$.

From Remark 9.1 we deduce, for $s = t_c, t = t_d$ belonging to the uniform dissection of $[0,1]$ in m intervals with $c < d \in \mathbb{N}^*$,

$$\tilde{X}^{(i),ab}_{st} - \tilde{X}^{D,(i),ab}_{st} = \sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\left(\hat{X}^a_{t_c,t_l}\right)^{i-k}\left(\tilde{X}^{(k),ab}_{t_l,t_{l+1}} - \tilde{X}^{D,(k),ab}_{t_l,t_{l+1}}\right).$$

Remark 9.2. *Remember that $\tilde{X}^{D,(k),ab}_{t_l,t_{l+1}} = \frac{1}{(k+1)!}\left(\hat{X}^a_{t_l,t_{l+1}}\right)^k X^b_{t_l,t_{l+1}}$, therefore we have*

$$\begin{aligned} \tilde{X}^{(i),ab}_{st} - \tilde{X}^{D,(i),ab}_{st} &= \sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\left(\hat{X}^a_{t_c,t_l}\right)^{i-k}\left(\tilde{X}^{(k),ab}_{t_l,t_{l+1}} - \tilde{X}^{D,(k),ab}_{t_l,t_{l+1}}\right) \\ &= \sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\left(\hat{X}^a_{t_c,t_l}\right)^{i-k}\left(\tilde{X}^{(k),ab}_{t_l,t_{l+1}} - \frac{1}{(k+1)!}\left(\hat{X}^a_{t_l,t_{l+1}}\right)^k X^b_{t_l,t_{l+1}}\right). \end{aligned}$$

Thus, using Cauchy-Schwarz, the small interval case (39) and 8.1, this yields

$$\begin{aligned} \|\tilde{X}^{(i),ab}_{st} - \tilde{X}^{D,(i),ab}_{st}\|_p &\leq \sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\|\left(\hat{X}^a_{t_c,t_l}\right)^{i-k}\left(\tilde{X}^{(k),ab}_{t_l,t_{l+1}} - \tilde{X}^{D,(k),ab}_{t_l,t_{l+1}}\right)\|_p \\ &\leq \sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\|\left(\hat{X}^a_{t_c,t_l}\right)^{i-k}\|_{2p}\|\tilde{X}^{(k),ab}_{t_l,t_{l+1}} - \tilde{X}^{D,(k),ab}_{t_l,t_{l+1}}\|_{2p} \\ &= \sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\|\left(\hat{X}^a_{t_c,t_l}\right)\|^{i-k}_{2p(i-k)}\|\tilde{X}^{(k),ab}_{t_l,t_{l+1}} - \tilde{X}^{D,(k),ab}_{t_l,t_{l+1}}\|_{2p} \\ &\leq \sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\left(c_H(t_l-t_c)^H\sqrt{2p(i-k)}\right)^{i-k}C(H,\alpha,\beta)\frac{1}{m^{kH+\frac{1}{2}}}p^{\frac{k+1}{2}} \\ &\leq C(H,\alpha,\beta)p^{\frac{i+1}{2}}\sum_{l=c}^{d-1}\sum_{k=0}^{i}(t_l-t_c)^{H(i-k)}\frac{1}{m^{kH+\frac{1}{2}}}. \end{aligned}$$

Now let us compute without upper bounds. We have

$$
\begin{aligned}
&C(H,\alpha,\beta)p^{\frac{i+1}{2}}\sum_{l=c}^{d-1}\sum_{k=0}^{i}(t_l-t_c)^{H(i-k)}\frac{1}{m^{kH+\frac{1}{2}}}\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\sum_{k=0}^{i}(0)^{H(i-k)}\frac{1}{m^{kH+\frac{1}{2}}}+\sum_{l=c+1}^{d-1}\sum_{k=0}^{i}(t_l-t_c)^{H(i-k)}\frac{1}{m^{kH+\frac{1}{2}}}\right]\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\frac{1}{m^{iH+\frac{1}{2}}}+\sum_{l=c+1}^{d-1}\sum_{k=0}^{i}(t_l-t_c)^{H(i-k)}\frac{1}{m^{kH+\frac{1}{2}}}\right]\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\frac{1}{m^{iH+\frac{1}{2}}}+\sum_{l=c+1}^{d-1}\frac{(t_l-t_c)^{iH}}{m^{1/2}}\sum_{k=0}^{i}(t_l-t_c)^{-kH}\frac{1}{m^{kH}}\right]\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\frac{1}{m^{iH+\frac{1}{2}}}+\sum_{l=c+1}^{d-1}\frac{(t_l-t_c)^{iH}}{m^{1/2}}\sum_{k=0}^{i}\frac{1}{(l-c)^{kH}}\right]\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\frac{1}{m^{iH+\frac{1}{2}}}+\frac{i+1}{m^{iH+\frac{1}{2}}}+\sum_{l=c+2}^{d-1}\frac{(t_l-t_c)^{iH}}{m^{1/2}}\sum_{k=0}^{i}\frac{1}{(l-c)^{kH}}\right]\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\frac{1}{m^{iH+\frac{1}{2}}}+\frac{i+1}{m^{iH+\frac{1}{2}}}+\frac{1}{m^{1/2}}\sum_{l=c+2}^{d-1}(t_l-t_c)^{iH}\frac{(l-c)^{(i+1)H}-1}{(l-c)^{(i+1)H}-(l-c)^{iH}}\right]\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\frac{1}{m^{iH+\frac{1}{2}}}+\frac{i+1}{m^{iH+\frac{1}{2}}}+\frac{1}{m^{iH+1/2}}\sum_{l=c+2}^{d-1}(l-c)^{iH}\frac{(l-c)^{(i+1)H}-1}{(l-c)^{(i+1)H}-(l-c)^{iH}}\right]\\
&=C(H,\alpha,\beta)p^{\frac{i+1}{2}}\left[\frac{1}{m^{iH+\frac{1}{2}}}+\frac{i+1}{m^{iH+\frac{1}{2}}}+\frac{1}{m^{iH+1/2}}\sum_{l=c+2}^{d-1}\frac{(l-c)^{(i+1)H}-1}{(l-c)^{H}-1}\right].
\end{aligned}
$$

We have $\sum_{l=c+2}^{d-1}\frac{(l-c)^{(i+1)H}-1}{(l-c)^{H}-1}=\sum_{l=2}^{d-c-1}\frac{l^{(i+1)H}-1}{l^{H}-1}$ and since $\frac{l^{(i+1)H}-1}{l^{H}-1}\sim_{l\to\infty} l^{iH}$, the comparison of diverging series criterion yields $\sum_{l=c+2}^{d-1}\frac{(l-c)^{(i+1)H}-1}{(l-c)^{H}-1}\sim(t-s)^{iH+1}(d-c)^{iH+1}\sim(t-s)^{iH+1}m^{iH+1}$. Thus, the upper bound $C(H,\alpha,\beta)p^{\frac{i+1}{2}}\sum_{l=c}^{d-1}\sum_{k=0}^{i}(t_l-t_c)^{H(i-k)}\frac{1}{m^{kH+\frac{1}{2}}}$ diverges in $m^{1/2}$. We do not end up with the desired 8.2.

The only step where we could have too crude in our bounding is before applying the triangle inequality. Indeed, one way to improve our bounds from the previous subsection is to avoid applying the triangular inequality to $\sum_{l=c}^{d-1}$, where the number of terms is destined to go to infinity.

9.2. **Reflections arising from this blind alley.**

Remark 9.3. *The F,G decomposition like in Proof of Proposition 3.1 of* [17] *only gives the order of the Wiener chaos, that allows to use hypercontractivity (Lemma 15.20 of* [15]*), then Itô isometry only once and then Gaussian integrability condition 8.1. Thus, we will not get a better estimate writing $X^{(i)}$ as an F-G decomposition.*

As we have seen in the above remark, using the F,G decomposition like in Proof of Proposition 3.1 of [17] should not be helpful. Let us try and upper bound the quantity of interest $\|\tilde{X}_{st}^{(i),ab}-\tilde{X}_{st}^{D,(i),ab}\|_p=\|\sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\left(\hat{X}_{t_c,t_l}^{a}\right)^{i-k}\left(\tilde{X}_{t_l,t_{l+1}}^{(k),ab}-\frac{1}{(k+1)!}\left(\hat{X}_{t_l,t_{l+1}}^{a}\right)^{k}X_{t_l,t_{l+1}}^{b}\right)\|_{L^p}$ in without applying the triangular inequality over the $\sum_{l=c}^{d-1}$.

We already know the Hölder regularity of $\|\tilde{X}_{st}^{(i),ab}\|_p$ from 7.3 so we can focus on $\|\tilde{X}_{st}^{D,(i),ab}\|_p$ if we start by $\|\tilde{X}_{st}^{(i),ab}-\tilde{X}_{st}^{D,(i),ab}\|_p\le\|\tilde{X}_{st}^{(i),ab}\|_p+\|\tilde{X}_{st}^{D,(i),ab}\|_p$. We will use hypercontractivity (Lemma 15.20 of [15]) so can focus on

$$
\|\tilde{X}_{st}^{D,(i),ab}\|_2=\|\sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!}\left(\hat{X}_{t_c,t_l}^{a}\right)^{i-k}\left(\tilde{X}_{t_l,t_{l+1}}^{D,(k),ab}\right)\|_2.
$$

Remember that $\tilde{X}_{t_l,t_{l+1}}^{D,(k),ab}=\frac{1}{(k+1)!}\left(\hat{X}_{t_l,t_{l+1}}^{a}\right)^{k}X_{t_l,t_{l+1}}^{b}$. Therefore, we shall focus on:

$$
\begin{aligned}
&|\sum_{l=c}^{d-1}\sum_{k=0}^{i}\frac{1}{(i-k)!(k+1)!}\left(\hat{X}_{t_c,t_l}^{a}\right)^{i-k}\left(\hat{X}_{t_l,t_{l+1}}^{a}\right)^{k}X_{t_l,t_{l+1}}^{b}|\\
&=|\sum_{l=c}^{d-1}X_{t_l,t_{l+1}}^{b}\sum_{k=0}^{i}\frac{1}{(i-k)!(k+1)!}\left(\hat{X}_{t_c,t_l}^{a}\right)^{i-k}\left(\hat{X}_{t_l,t_{l+1}}^{a}\right)^{k}|
\end{aligned}
$$

Remark 9.4. *We do not know what the sign of $X_{t_l,t_{l+1}}^{b}$ is so we cannot upper bound*

$$
|\sum_{l=c}^{d-1}X_{t_l,t_{l+1}}^{b}\sum_{k=0}^{i}\frac{1}{(i-k)!(k+1)!}\left(\hat{X}_{t_c,t_l}^{a}\right)^{i-k}\left(\hat{X}_{t_l,t_{l+1}}^{a}\right)^{k}|
$$

by $|\sum_{l=c}^{d-1}X_{t_l,t_{l+1}}^{b}\sum_{k=0}^{i}\frac{1}{(i-k)!k!}\left(\hat{X}_{t_c,t_l}^{a}\right)^{i-k}\left(\hat{X}_{t_l,t_{l+1}}^{a}\right)^{k}|$ *and cannot apply binomial theorem.*

We do not yet see how we could upper bound $\| \sum_{l=c}^{d-1} X^b_{t_l,t_{l+1}} \sum_{k=0}^{i} \frac{1}{(i-k)!(k+1)!} \left(\hat{X}^a_{t_c,t_l}\right)^{i-k} \left(\hat{X}^a_{t_l,t_{l+1}}\right)^{k} \|_{L^2}$ by $C(\alpha,\beta,H)|D|^\theta (t-s)^\epsilon$ with ϵ, θ big enough, as mentioned in Remark 8.4.

10. Following the proof of Theorem 15.42 of [15] in the case of (α,β) rough paths

In Section 9 we needed an explicit expression of $\tilde{X}^{(i)} - \tilde{X}^{D,(i)}$ as a function of increments "adapted to the dissection" to control $\|\tilde{X}^{(i)} - \tilde{X}^{D,(i)}\|_{L^2}$.

In Proposition 15.30 of [15], they control $\|\pi_i(\mathbf{X}_{st}) - \pi_i(\mathbf{Y}_{st})\|_{L^2}$ by ω the control on the covariance of (X,Y), provided that X,Y are of finite variation sample paths. In our case $Y = (X,\hat{X})^D$, From 14.1 and Proposition 15.13 of [15], we can Hölder dominate the covariances R_{X,X^D} and $R_{\hat{X},\hat{X}^D}$ by ω_D and $\hat{\omega}_D$.

Like in the beginning of the proof of Theorem 15.42, by interpolation, we can dominate $|R_{X-X^D}|_{\rho'-var;[s,t]^2}$ by ω_D and $|R_{\hat{X}-\hat{X}^D}|_{\rho'-var;[s,t]^2}$ by $\hat{\omega}_D$.

In the beginning of Section 15.3.3 of [15], they say that, for any d-dimensional continuous Gaussian process X, whose sample paths are not of bounded variation, but whose covariance has finite ρ-variation, $\rho \in [1,2)$, we may consider suitably smooth approximations (X^n) for which

$$\sup_{n,m} \omega^{n,m}([0,1]^2) \leq K$$

where $\omega^{n,m}$ is a 2D control which controls the ρ-variation of $R_{(X^n,X^m)}$. We can apply this to both X and $\hat{X}$. That allows to transpose an analog of Proposition 15.30 of [15] into Theorem 15.37, which is relieved from the bounded variation sample paths assumption.

Proposition 15.30 is based on 2D Young estimates whereas the (α,β) rough path lift is defined with Itô integrals. We are in a simpler case because we do not *define* our lifting as a limit of liftings of smooth approximations.

From this Hölder domination, let us follow proof of (b) of Proposition 15.30 of [15] to Hölder dominate our i levels of an (α,β) rough path. We can proceed as

$$\begin{aligned}
\|\tilde{X}^{(i)}_{st} - \tilde{X}^{D,(i)}_{st}\|_2^2 &\leq 2\|\tilde{X}^{(i)}_{st} - \int_s^t (\hat{X}_{sr})^i dX^D_r\|_2^2 + 2\|\int_s^t (\hat{X}_{sr})^i dX^D_r - \tilde{X}^{D,(i)}_{st}\|_2^2 \\
&= 2\|\int_s^t (\hat{X}_{sr})^i d(X - X^D)_r\|_2^2 + 2\|\int_s^t \Big((\hat{X}_{sr})^i - (\hat{X}^D_{sr})^i\Big) dX^D_r\|_2^2 \\
&= 2\epsilon\|\int_s^t (\hat{X}_{sr})^i d(\frac{X - X^D}{\sqrt{\epsilon}})_r\|_2^2 + 2\epsilon\|\int_s^t \Big(\frac{(\hat{X}_{sr})^i - (\hat{X}^D_{sr})^i}{\sqrt{\epsilon}}\Big) dX^D_r\|_2^2,
\end{aligned}$$

or as

$$\begin{aligned}
\|\tilde{X}^{(i)}_{st} - \tilde{X}^{D,(i)}_{st}\|_2^2 &\leq 2\|\tilde{X}^{(i)}_{st} - \int_s^t (\hat{X}^D_{sr})^i dX_r\|_2^2 + 2\|\int_s^t (\hat{X}^D_{sr})^i dX_r - \tilde{X}^{D,(i)}_{st}\|_2^2 \\
&= 2\|\int_s^t \Big((\hat{X}_{sr})^i - (\hat{X}^D_{sr})^i\Big) dX_r\|_2^2 + 2\|\int_s^t (\hat{X}^D_{sr})^i d(X - X^D)_r\|_2^2 \\
&= 2\epsilon\|\int_s^t \Big(\frac{(\hat{X}_{sr})^i - (\hat{X}^D_{sr})^i}{\sqrt{\epsilon}}\Big) dX_r\|_2^2 + 2\epsilon\|\int_s^t (\hat{X}^D_{sr})^i d(\frac{X - X^D}{\sqrt{\epsilon}})_r\|_2^2,
\end{aligned}$$

with ϵ destined to be $\frac{1}{m} = |D|$ to a positive power, as we can see in the proof of Theorem 15.42 of [15] where $\epsilon^2 = c_3 \max_{t_i \in D} \omega_D([t_i,t_{i+1}]^2)^{\frac{1}{\rho}-\frac{1}{\rho'}}$. Afterwards, what [15] do in this proof (of (b) of Proposition 15.30) is they make use of Proposition 15.28 for a 2-dimensional Gaussian process of the form $\left(\frac{(\hat{X}_{sr})^i - (\hat{X}^D_{sr})^i}{\sqrt{\epsilon}}, X_r\right)$, for example in our (α,β) case. We want to prove an analog of Proposition 15.28 of [15] in the case of (α,β) rough paths. To do so, we will first prove the statements of Proposition 15.28 for piecewise linear (or mollifier) approximations, which allows to make the bounded variation assumption, and therefore consider random Riemann-Stieltjes integrals instead of Itô integrals. Then, we will make the passage to the limit as they do in 17.4.

In fact, Fukasawa [17] defines his lift via Itô integrals but I can try to define it as a limit of lifts of piecewise linear approximations or mollifiers. Then, as with the Brownian motion in [15] 16, we could see if it coincides with the (α,β) lift defined with Itô integration.

I could try to follow the proof of dominance of the $1/2H$ variation of the fBM (17.1) to dominate the ρ-variation of the power i of the fBM. I can then use Young's 2D estimates (Theorem 6.18 of [15]), which are a generalisation of Theorem 6.8 of [15] to 2D functions. Yet, as we see in the following subsection, the nature of the processes $(\hat{X})^i$ are not naturally compatible with a reasoning with their covariance, which is intimately linked with the theory of controlling 2-dimensional variations of functions.

10.1. Blind alley not allowing to use the Young 2D estimates.

Let $\hat{X}$ be a bounded variation Gaussian process (that will be a mollifier or piecewise linear approximation to the fBM $\hat{X}$) and X, a bounded variation Gaussian process (that will be a mollifier or piecewise linear approximation to the BM X). To treat the simpler case first, suppose $\hat{X}$ independent of X. Following the proof of (b) of Proposition 15.28, we upper bound as follows:

$$\begin{aligned}
\mathbb{E}\left[\left(\int_s^t (\hat{X}_{sr})^i dX_r\right)\left(\int_{s'}^{t'} (\hat{X}_{s'r})^i dX_r\right)\right] &= \mathbb{E}\left[\int_s^t \int_{s'}^{t'} (\hat{X}_{su})^i (\hat{X}_{s'v})^i dX_u d(X)_v\right] \\
= \int_s^t \int_{s'}^{t'} \mathbb{E}\left[(\hat{X}_{su})^i (\hat{X}_{s'v})^i dX_u dX_v\right] &= \int_s^t \int_{s'}^{t'} \mathbb{E}[(\hat{X}_{su})^i (\hat{X}_{s'v})^i] d\mathbb{E}[X_u X_v]
\end{aligned}$$

Remark 10.1. *We want to be able to bound* $|\mathbb{E}[(\hat{X}_{su})^i(\hat{X}_{s'v})^i]|_{\rho-var;[0,1]^2}$. *To find the right* ρ, *we know we will apply the 2D Young estimate (Theorem 6.18 of* [15]*) to* $\int_s^t\int_{s'}^{t'}\mathbb{E}[(\hat{X}_{su})^i(\hat{X}_{s'v})^i]d\mathbb{E}[X_uX_v]$ *with* $\mathbb{E}[X_uX_v]=R_X(u,v)=u\vee v$ *which is of bounded variation. Thus, by Theorem 6.18 of* [15]*, we just need to control the 2D* ρ*-variation of* $\mathbb{E}[(\hat{X}_{s\cdot})^i(\hat{X}_{s'\cdot})^i]$ *for any* $\rho>0$. *That way, we would have* $1+\frac{1}{\rho}>1$ *and will be able to apply Theorem 6.18 of* [15].

Recall that $R_X\begin{pmatrix}s,u\\s',v\end{pmatrix}=\mathbb{E}[X_{su}X_{s'v}]=\mathbb{E}[X_uX_v]-\mathbb{E}[X_sX_v]-\mathbb{E}[X_uX_{s'}]+\mathbb{E}[X_sX_{s'}]=R_X\begin{pmatrix}s\\s'\end{pmatrix}+R_X\begin{pmatrix}u\\v\end{pmatrix}-R_X\begin{pmatrix}s\\v\end{pmatrix}-R_X\begin{pmatrix}u\\s'\end{pmatrix}$.

Therefore, we have to transform $\mathbb{E}[(\hat{X}_{su})^i(\hat{X}_{s'v})^i]$ into a 2D increment as defined in Section 5.5.1 of [15].

But, even with binomial theorem, we would only have:

$$\mathbb{E}[(\hat{X}_{su})^i(\hat{X}_{s'v})^i]=\sum_{k=0}^{i}\sum_{l=0}^{i}\binom{i}{k}\binom{i}{l}\mathbb{E}[(-\hat{X}_s)^k(-\hat{X}_{s'})^l\hat{X}_u^{i-k}\hat{X}_v^{i-l}]$$

As a conclusion, it is very unlikely that we would be able to transform $\mathbb{E}[(\hat{X}_{su})^i(\hat{X}_{s'v})^i]$ into a 2D increment as defined in Section 5.5.1 of [15] in order to use Theorem 6.18 of [15]. Indeed, we need would need $\mathbb{E}[(\hat{X}_{su})^i(\hat{X}_{s'v})^i]$ to equal the sum of four terms of a 2D function like in $R_X\begin{pmatrix}s,u\\s',v\end{pmatrix}=R_X\begin{pmatrix}s\\s'\end{pmatrix}+R_X\begin{pmatrix}u\\v\end{pmatrix}-R_X\begin{pmatrix}s\\v\end{pmatrix}-R_X\begin{pmatrix}u\\s'\end{pmatrix}$.

11. Convergence of (α,β) lifting of piecewise approximation for nested dissections

In the face of the previous unsuccessful attempts to show a strong approximation lemma as 8.2 for any sequence of dissections (we even considered the case of uniform dissections but it did not suffice), we try to show an approximation for nested piecewise linear approximatons by naively following the reasoning of Section 13.3.2 of [15]. Following this reasoning also fails because the fBM is neither a Markov process nor a semimartingale. Let us briefly depict why this property is core to the reasoning of Section 13.3.2 of [15]. To do so, we will sum up the reasoning of Section 13.3.2 of [15], which was for the Brownian motion and stress the key points of the proof that fail in our (α,β) rough path framework.

Consider a sequence (D_n) of nested dissections, that is $D_n\subseteq D_{n+1}$ for all n, such that $|D_n|\to 0$. Define $\mathcal{F}_n:=\sigma(B_t:t\in D_n)$. It forms a family of σ-algebras increasing in n, that is $\mathcal{F}_n$ is a filtration. Define $B^n=B^{D_n}(\omega)$ as the piecewise linear approximation of B along D_n.

Proposition 11.1 (Proposition 13.17 of [15])**.** *For fixed* $t\in[0,T]$, *the convergence* $\boldsymbol{B}_t^n\to\boldsymbol{B}_t$ *holds almost surely and in* $L^2(\mathbb{P})$. *This is equivalent to*

$$|B_t^n-B_t|\to 0 \text{ and}$$
$$|A_t^n-A_t|\to 0$$

Proof. (a) Using the Markov property of B, [15] are able to show that $\mathbb{E}[B_t|\mathcal{F}_n]=B_t^n$. Indeed, the Markov property yields $\mathbb{E}[B_t|\mathcal{F}_n]=\mathbb{E}[B_t|B_{t_i},B_{t_{i+1}}]$ with t_i,t_{i+1} neighbours in D_n with $t\in[t_i,t_{i+1}]$. By Gaussian conditioning, they obtain

$$\mathbb{E}[B_t|B_{t_i},B_{t_{i+1}}]=\frac{t_{i+1}-t}{t_{i+1}-t_i}B_{t_i}+\frac{t-t_i}{t_{i+1}-t_i}B_{t_{i+1}}$$

and this is precisely equal to B_t^n.

Therefore, they can show that, for fixed t, $(B_t^n)_n$ is a discrete martingale, that is

$$\mathbb{E}[B_t^{n+1}|\mathcal{F}_n]=\mathbb{E}[\mathbb{E}[B_t|\mathcal{F}_{n+1}]|\mathcal{F}_n]=\mathbb{E}[B_t|\mathcal{F}_n]=B_t^n.$$

On the other hand, we have:

$$\sup_n\mathbb{E}[|B_t^n|^2]=\sup_n\mathbb{E}[|\mathbb{E}[B_t|\mathcal{F}_n]|^2]\leq\mathbb{E}[\mathbb{E}[|B_t|^2||\mathcal{F}_n]]=\mathbb{E}[|B_t|^2]<\infty.$$

So, by L^2 martingale convergence theorem, we have $B_t^n\to B_t$ a.s. and in L^2.

□

The fBM is neither a Markov process nor a semimartingale. In (a) of the proof above, they use the Markov property of the Brownian motion to say that $\mathbb{E}[B_t|\mathcal{F}_n]=\mathbb{E}[B_t|B_{t_i},B_{t_{i+1}}]$. We cannot say that $\mathbb{E}[\hat{X}_t|\mathcal{F}_n]=\mathbb{E}[\hat{X}_t|\hat{X}_{t_i},\hat{X}_{t_{i+1}}]$ with t_i,t_{i+1} neighbours in D_n with $t\in[t_i,t_{i+1}]$.

12. Conclusion

12.1. **Summary of our contribution.** The most important established contribution of this report is the generalisation of the (α,β) lifting to N dimensions in 7.1 and the proof of that our lifting 7.3 yields an N-dimensional (α,β) rough path.

We have not been able to show the strong approximation lemma depicted in 8.2 but have explored various paths to show such a property. The most promising method, which was to follow the proof of 3.19, that is the case of Brownian rough paths, was natural in the sense that the (α,β) lifting is also defined as Itô integrals of Gaussian processes, for which we can use 8.1. But, we were not able to make use of the covariance of the increments $\hat{X}_{t_l,t_{l+1}}$ for $[t_l,t_{l+1}]$ elementary intervals of the dissection.

The second main track, which was to follow the stochastic integration of Gaussian processes depicted in Section 15 of [15], also failed because the powers $(\hat{X}_{s,r})^i$ contained in the $\tilde{X}^{(i)}$ levels do not seem naturally associated to the control of their covariance function.

If we are able to show such a lemma, we have most of the analogs to the tools used in [7] to prove an enhanced Sanov theorem for (α, β) rough paths. Indeed, except from 6.9, the lemmas that follow in [7] should be proven in the same way provided we have a strong approximation lemma analog to 3.19.1 or 3.26.

12.2. **Further open problems.** Alternative phrasing: Assuming an enhanced Sanov-type LDP for N-dimensional (α, β) rough paths, a natural next step is to define an integral, like in Theorem 2.5 of [17], of N-dimensional (α, β)rough paths against themselves. The stability properties of such a rough integral, paired with the contraction principle, would allow us to deduce a Sanov-type LDP for the price or log-price process under rough volatility 2.

One should also explore to what extent we could also embed the N-dimensional (α, β) rough paths in the framework of branched rough paths, as mentioned in Remark 2.2 of [17].

The second main track, where we tried to follow the stochastic integration of Gaussian processes of [15], might have failed because the framework of Gaussian processes might not be the best to treat powers of a Gaussian process like $(\hat{X})^i$. Indeed, just like a χ^2 law is the law of a sum of squares of Gaussian laws, one should find the analog for the power $i \in I$ of a Gaussian process to be able to treat $(\hat{X})^i$.

Part 3. Appendix

13. Controls and control distances and norms

Denote the simplex

$$\Delta := \Delta_T = \{(s,t) : 0 \le s \le t \le T\}.$$

Definition 13.1 (Definition 1.6 of [15]). *A map $\omega : \Delta_T \to [0, \infty)$ is called superadditive if for all $s \le t \le u$ in $[0,T]$,*

$$\omega(s,t) + \omega(t,u) \le \omega(s,u).$$

If, in addition, ω is continuous and zero on the diagonal, i.e. $\omega(s,s) = 0$ for $0 \le s \le T$ we call ω a control or, more precisely, a control function on $[0,T]$.

Definition 13.2 (Homogenous $p-\omega$ distance, Definition 8.2 of [15]). *(i) Given a control function ω on $[0,T]$ and $\mathbf{x} \in C([0,T], G^N(\mathbb{R}^d))$ we define the homogenous $p-\omega$ norm*

$$\|\mathbf{x}\|_{p-\omega;[0,T]} := \sup_{0 \le s < t \le T} \frac{d(\mathbf{x}_s, \mathbf{x}_t)}{\omega(s,t)^{1/p}} = \sup_{0 \le s < t \le T} \frac{\|\mathbf{x}_{s,t}\|}{\omega(s,t)^{1/p}}$$

and

$$C^{p-\omega}([0,T], G^N(\mathbb{R}^d)) = \{\mathbf{x} \in C([0,T], G^N(\mathbb{R}^d)) : \|\mathbf{x}\|_{p-\omega;[0,T]} < \infty\}.$$

(ii) Given $\mathbf{x}, \mathbf{y} \in C^{p-\omega}([0,T], G^N(\mathbb{R}^d))$ and a control function ω on $[0,T]$, we define

$$d_{p-\omega;[0,T]}(\mathbf{x}, \mathbf{y}) := \sup_{0 \le s < t \le T} \frac{d(\mathbf{x}_{s,t}, \mathbf{y}_{s,t})}{\omega(s,t)^{1/p}}.$$

14. 2-dimensional functions and controls

Lemma 14.1. *like in the proof of Proposition 15.11 of* [15]*: if $E \simeq \mathbb{R}^d$ and $f : [0,,T]^2 \to E$, then for all $[s,t] \times [u,v] \subseteq [0,,T]^2$ we have:*

$$|f|^p_{p-var;[s,t]\times[u,v]} \le \sum_{k=1}^d |f_k|^p_{p-var;[s,t]\times[u,v]}$$

Proof.

$$|f|^p_{p-var;[s,t]\times[u,v]} := \sup_{\substack{(t_i) \in \mathcal{D}([s,t]) \\ (t'_i) \in \mathcal{D}([u,v])}} \sum_{i,j} \left| f \begin{pmatrix} t_i, t_{i+1} \\ t'_j, t'_{j+1} \end{pmatrix} \right|^p_E \le \sup_{\substack{(t_i) \in \mathcal{D}([s,t]) \\ (t'_i) \in \mathcal{D}([u,v])}} \sum_{i,j} \sum_{k=1}^d \left| f_k \begin{pmatrix} t_i, t_{i+1} \\ t'_j, t'_{j+1} \end{pmatrix} \right|^p_{\mathbb{R}}$$
$$\le \sum_{k=1}^d \sup_{\substack{(t_i) \in \mathcal{D}([s,t]) \\ (t'_i) \in \mathcal{D}([u,v])}} \sum_{i,j} \left| f_k \begin{pmatrix} t_i, t_{i+1} \\ t'_j, t'_{j+1} \end{pmatrix} \right|^p_{\mathbb{R}} = \sum_{k=1}^d |f_k|^p_{p-var;[s,t]\times[u,v]}$$

□

15. Proof of equation 10.13 of [15]

Proposition 8.12 of [15] gives us the statement but with convergence in d_∞ instead of $d_{0;[0,T]}$. Then, we use the rhs inequality of Proposition 8.15 of [15], called the d_0/d_∞ estimate, that is (denoting $S_{[p]}(x_n) = \mathbf{x}^n$):

$$d_0(\mathbf{x}^n, \mathbf{x}) \le C \max\{d_\infty(\mathbf{x}^n, \mathbf{x}), d_\infty(\mathbf{x}^n, \mathbf{x})^{1/N}(\|\mathbf{x}\|_\infty + \|\mathbf{x}^n\|_\infty)^{1-1/N}\}$$

(Moreover, we have the uniform p-variation bounds $\sup_n \|S_{[p]}(x_n)\|_{p-var;[0,T]} < \infty$). Recall the definition $\|\mathbf{x}\|_\infty := d_\infty(\mathbf{x}, o) = \sup_{t \in [0,T]} d(\mathbf{x}_t, o)$ (where d is still the Carnot-Caratheodory metric): since $\lim_{n\to\infty} d_{\infty;[0,T]}(\mathbf{x}_n, \mathbf{x}) = 0$, we deduce $\sup_n \|\mathbf{x}_n\|_\infty < \infty$. Therefore, we get

$$\lim_{n\to\infty} d_{0;[0,T]}(\mathbf{x}_n, \mathbf{x}) = 0$$

Remark 15.1. *When two distances are locally β-Hölder equivalent, like d_0 and d_∞ above, the notion of convergence of x_n is the same (because we get "local" for n large enough).*

16. Itô vs Young lifting of Brownian motion in [15]'s book

The rough path lifting of a standard Brownian motion is defined as an Itô integral 3.14. But it is a particular instance of a more general lifting for Gaussian processes in 3.22.

Right after 3.22, [15] say that "Theorem 15.33 asserts in particular that d-dimensional Brownian motion can be naturally lifted to an enhanced Gaussian process, easily identified as enhanced Brownian motion (in view of (iv) of Theorem 15.33 and the results of Section 13.3.3)". Section 15.3.3 is precisely the section with Corollary 3.19.1. On the other hand, (iv) of 3.22 states "the lift $\mathbf{X}$ is natural in the sense that it is the limit of $S_3(X^n)$ where X^n is any sequence of piecewise linear or mollifier approximations to X such that $d_\infty(X^n, X)$ converges to 0 almost surely.". Let X be a standard Brownian motion. Denote any piecewise linear approximation $X^n := X^{D_n}$. Denote $\tilde{S}_2(X)$ the enhanced Brownian motion defined in 3.14 and $S_2(X)$ the enhanced Gaussian process built from X and 3.22.

$$d_{\alpha-Hol}(\tilde{S}_2(X), S_2(X)) \leq d_{\alpha-Hol}(\tilde{S}_2(X), S_2(X^n)) + d_{\alpha-Hol}(S_2(X^n), S_2(X))$$

The left term in the rhs goes to 0 as $n \to \infty$ by Exercise 13.22 of [15] and the right term in the rhs goes to 0 as $n \to \infty$ by 3.22 (in the proof in [15], the sentence above equation (15.21) of the proof).

17. Proofs of Section 15 of [15]

17.1. Controlling the covariance of a fractional Brownian motion.

Let us state a quick fact:

Proposition 17.1. *Let X be a real-valued Gaussian process with covariance R and $[s,t]\times[u,v]$ a rectangle (in $[0,1]$ say). The increment $R\begin{pmatrix} s,t \\ u,v \end{pmatrix} = \mathbb{E}[X_{s,t}X_{u,v}]$*

Proposition 17.2 (Proposition 15.5 of [15])**.** *A fractional Brownian motion β^H with $H \in (0,1/2]$ has covariance of finite $1/2H$-variation, controlled by*

$$\omega_H(\cdot,\cdot) := |R^H|_{1/(2H)-var;[\cdot,\cdot]\times[\cdot,\cdot]}$$

Moreover, there exists a constant $C = C(H)$ such that, for all $s < t \in [0,1]$,

$$|R^H|_{1/(2H)-var;[s,t]^2} \leq C|t-s|^{\frac{1}{2H}}$$

so that ω_H is a Hölder-dominated control.

Proof. From common properties of the fBM, we have that $\mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{t_j,t_{j+1}}] \leq 0$ for $i \neq j$ since $H \leq 1/2$.

From Proposition 19.4, we have that, for any real numbers $(x_i)_{i=1}^n$, we have

$$\Big|\sum_{i=1}^n x_i\Big|^p \leq \begin{cases} \sum_{i=1}^n |x_i|^p & p \in (0,1] \\ n^{p-1}(\sum_{i=1}^n |x_i|^p) & p > 1 \end{cases}$$

Thus $1/(2H) \geq 1$. We use the inequality above with $n = 2$ and $p = 1/(2H)$ as $\sum_{j\neq i}\mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{t_j,t_{j+1}}] = \sum_j \mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{t_j,t_{j+1}}] - \mathbb{E}[|\beta^H_{t_i,t_{i+1}}|^2]$. This yields:

$$\Big|\sum_{j\neq i}\mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{t_j,t_{j+1}}]\Big|^{\frac{1}{2H}} \leq 2^{\frac{1}{2H}-1}\Big|\sum_j \mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{t_j,t_{j+1}}]\Big|^{\frac{1}{2H}} + 2^{\frac{1}{2H}-1}\Big|\mathbb{E}[|\beta^H_{t_i,t_{i+1}}|^2]\Big|^{\frac{1}{2H}}$$

In the end of this sequence of upper bounds, they use $\sum_j \beta^H_{t_j,t_{j+1}} = \beta^H_{s,t}$.

We use that $\mathbb{E}[|\beta^H_t|^2] = |t|^{2H}$ and $\mathbb{E}[|\beta^H_{st}|^2] = |t-s|^{2H}$. Thus

$$\sum_i (\mathbb{E}[|\beta^H_{t_i,t_{i+1}}|^2])^{1/2H} = \sum_i (|t_{i+1}-t_i|^{2H})^{1/2H} = t-s = |t-s|$$

Hence the upper bound

$$\sum_{i,j}|\mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{t_j,t_{j+1}}]|^{\frac{1}{2H}} \leq C_H|t-s| + C_H\sum_i |\mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{s,t}]|^{\frac{1}{2H}} \tag{40}$$

We now prove that, for $[u,v] \subseteq [s,t]$ we have $\mathbb{E}[\beta^H_{u,v}\beta^H_{s,t}] \leq C_H|v-u|^{2H}$. We use that $|\sum_i x_i|^{2H} \leq \sum_i |x_i|^{2H}$. From essential properties of the fBM, we have that $\mathbb{E}[\beta^H_{u,v}\beta^H_{s,t}] = \frac{1}{2}\Big(|t-u|^{2H} + |s-v|^{2H} - |s-u|^{2H} - |t-v|^{2H}\Big)$. By triangle inequality we have, in our case where $s < u < t < v$ that

$$\begin{aligned}
|\mathbb{E}[\beta^H_{u,v}\beta^H_{s,t}]| &= \frac{1}{2}\Big|(t-u)^{2H} + (v-s)^{2H} - (u-s)^{2H} - (t-v)^{2H}\Big| \\
&\leq \frac{1}{2}\Big|(t-u)^{2H} - (t-v)^{2H}\Big| + \frac{1}{2}\Big|(v-s)^{2H} - (u-s)^{2H}\Big| \\
&= \frac{1}{2}\Big((t-u)^{2H} - (t-v)^{2H}\Big) + \frac{1}{2}\Big((v-s)^{2H} - (u-s)^{2H}\Big)
\end{aligned}$$

Then, using a special case of $|\sum_{i=1}^n x_i|^{2H} \leq \sum_i |x_i|^{2H}$ for $n = 2$ and positive x_1, x_2 with $t - u = (t-v) + (v-u)$ and $v - s = (v-u) + (u-s)$, we get

$$|\mathbb{E}[\beta^H_{u,v}\beta^H_{s,t}]| \leq 2\frac{1}{2}(v-u)^{2H} = (v-u)^{2H}$$

In particular, this yields

$$|\mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{s,t}]|^{\frac{1}{2H}} \leq t_{i+1} - t_i$$

Therefore, for any dissection $D=(t_i)$ of $[s,t]$, (40) yields:

$$\sum_{i,j}\left|R^H\begin{pmatrix}t_i,t_{i+1}\\ t_j,t_{j+1}\end{pmatrix}\right|^{\frac{1}{2H}}=\sum_{i,j}|\mathbb{E}[\beta^H_{t_i,t_{i+1}}\beta^H_{t_j,t_{j+1}}]|^{\frac{1}{2H}}\leq 2C_H|t-s|$$

This gives:

$$|R^H|^{1/(2H)}_{\frac{1}{2H}-var;[s,t]^2}\leq 2C_H|t-s|$$

By Lemma 5.54 of [15], the $1/(2H)$-variation of R^H is finite (taking $s=0,t=1$ in the above inequality). Moreover, the above inequality gives the Hölder domination of the control $|R^H|_{\frac{1}{2H}-var;[s,t]^2}$. □

If $H=1/2$, we look at a standard Brownian motion, with covariance $(s,t)\mapsto t\wedge s$ which is of finite 1-variation (constant a.s.).

17.2. **Proof of Proposition 15.28 of** [15]**.** In the proof of (b) of Proposition 15.28, $\int_s^t\int_{s'}^{t'}X^i_{s,u}X^i_{s',v}dX^j_u dX^j_v$ is a 2D young integral

Since, in Proposition 15.28, $X\in C^{1-var}([0,T],E)$, we can, by Proposition 6.16, define the 2D random Young integral $\int_s^t\int_{s'}^{t'}X^i_{s,u}X^i_{s',v}dX^j_u dX^j_v$ "omega by omega" as a sum of 4 1D Young integrals, which are themselves defined in Proposition 6.4 as Riemann-Stieltjes integrals, defined in Proposition 2.2 or 2.4.

By definition of the 2D Young integral of finite variation continuous paths:

$$\int_s^t\int_{s'}^{t'}X^i_{s,u}X^i_{s',v}dX^j_u dX^j_v=$$
$$\int_{[s,t]\times[s',t']}X^i_{s,u}X^i_{s',v}dX^j\begin{pmatrix}u\\ v\end{pmatrix}$$

Since, in the Definition of the Riemman-Stieltjes integral (Definition 2.1 of [15]), the limit must not depend on the sequence of dissections, we can take the same sequence $(D_n)_n$ of dissection of $[s\wedge s',t\vee t']$ for each integral in the product below:

$$\left(\int_s^t X^i_{s,u}dX^j_u\right)\left(\int_{s'}^{t'}X^i_{s',v}dX^j_v\right)=\left(\lim_{n\to\infty}\sum_{t_k\in D_n\cap[s,t]}X^i_{s,t_k}X^j_{t_k,t_{k+1}}\right)\left(\lim_{n\to\infty}\sum_{t_k\in D_n\cap[s',t']}X^i_{s',t_k}X^j_{t_k,t_{k+1}}\right)$$
$$=\lim_{n\to\infty}\sum_{t_k\in D_n\cap[s,t]}\sum_{t_l\in D_n\cap[s',t']}X^i_{s,t_k}X^i_{s',t_l}X^j_{t_k,t_{k+1}}X^j_{t_l,t_{l+1}}$$

and by dominated convergence (because X has Gaussian integrability) or monotone convergence of series

$$\mathbb{E}[\sum_{t_k\in D_n\cap[s,t]}\sum_{t_l\in D_n\cap[s',t']}X^i_{s,t_k}X^i_{s',t_l}X^j_{t_k,t_{k+1}}X^j_{t_l,t_{l+1}}]=\sum_{t_k\in D_n\cap[s,t]}\sum_{t_l\in D_n\cap[s',t']}\mathbb{E}[X^i_{s,t_k}X^i_{s',t_l}X^j_{t_k,t_{k+1}}X^j_{t_l,t_{l+1}}]$$
$$=\sum_{t_k\in D_n\cap[s,t]}\sum_{t_l\in D_n\cap[s',t']}\mathbb{E}[X^i_{s,t_k}X^i_{s',t_l}]\mathbb{E}[X^j_{t_k,t_{k+1}}X^j_{t_l,t_{l+1}}]$$

The computation above allows [15] to write, by passage to the limit in n:

$$\mathbb{E}\left[\left(\int_s^t X^i_{s,u}dX^j_u\right)\left(\int_{s'}^{t'}X^i_{s',v}dX^j_v\right)\right]=\int_s^t\int_{s'}^{t'}\mathbb{E}[X^i_{s,u}X^i_{s',v}]d\mathbb{E}[X^j_uX^j_v]$$

Since R_X is a 2D function of finite ρ-variation, we have, by Theorem 6.18:

$$\mathbb{E}\left[\left(\int_s^t X^i_{s,u}dX^j_u\right)\left(\int_{s'}^{t'}X^i_{s',v}dX^j_v\right)\right]=\sum_{t_k\in D_n\cap[s,t]}\sum_{t_l\in D_n\cap[s',t']}R_{X^i}\begin{pmatrix}s,t_k\\ s',t_l\end{pmatrix}R_{X^j}\begin{pmatrix}t_k,t_{k+1}\\ t_l,t_{l+1}\end{pmatrix}$$

17.3. **Proof of Theorem 15.33 of** [15]**.** We call (15.20 bis) the property below: There exists $\theta>0$ and $c_2=c_2(p,q,K,\theta)$ so that:

$$\|d_{p-var;[0,1]}(S_3(X_n),S_3(X_m))\|_{L^q}\leq c_2\|R_{X_n-X_m}\|^\theta_\infty$$

It follows, by Markov that

$$\mathbb{P}\left(d_{p-var;[0,1]}(S_3(X_n),S_3(X_m))>\epsilon\right)\leq\frac{\mathbb{E}[d_{p-var;[0,1]}(S_3(X_n),S_3(X_m))^q]}{\epsilon^q}\leq\frac{(c_2\|R_{X_n-X_m}\|^\theta_\infty)^q}{\epsilon^q}\to_{m,n\to\infty}0$$

So $(S_3(X_n))_n$ is Cauchy-in-probability. By the Cauchy criterion for convergence in probability of random variables with values in a Polish space, we get that there exists $\mathbf{X}\in C^{0,p-var}_o([0,1],G^3(\mathbb{R}^d))$ such that $d_{p-var;[0,1]}(S_3(X_n),\mathbf{X})\to 0$ in probability.

One should also remark that the proof using piecewise linear approximations is the same except minor modifications. Indeed, instead of using Proposition 15.14 of [15], we can use Proposition 15.11, which allows to prove

$$\sup_{n,m}|R(X_n,X_m)|^\rho_{\rho-var;[0,1]^2}\leq c_1(\rho)|R|^\rho_{\rho-var;[0,1]^2}$$

with $c_1(\rho)=4\times 9^{1-\frac{1}{\rho}}$. The rest of the proof follows the same way.

17.4. **Proof of Theorem 15.37 of** [15]**.** It is the same as the proof of Corollary 15.31. Corollary 15.31 is Proposition 15.30 combined with hypercontractivity (Lemma 15.20 of [15]) and definition of the p-variation norm. So, to prove Theorem 15.37, we shall follow the proof of Proposition 15.30 and do a passage to the limit of the sequence of piecewise linear or mollifier approximations that allow to define the enhanced Gaussian process (Theorem 15.33).

Let us start by the first statement of the proof of Proposition 15.28.

$$\begin{aligned}\mathbb{E}[\mathbf{X}^{i,j}_{s,t}\mathbf{X}^{i,j}_{s',t'}] &\leq |\mathbb{E}[\mathbf{X}^{i,j}_{s,t}\mathbf{X}^{i,j}_{s',t'}] - \mathbb{E}[S_3(X^n)^{i,j}_{s,t}S_3(X^n)^{i,j}_{s',t'}]| + |\mathbb{E}[S_3(X^n)^{i,j}_{s,t}S_3(X^n)^{i,j}_{s',t'}| \\ &\leq |\mathbb{E}[\mathbf{X}^{i,j}_{s,t}\mathbf{X}^{i,j}_{s',t'}] - \mathbb{E}[S_3(X^n)^{i,j}_{s,t}S_3(X^n)^{i,j}_{s',t'}]| + C\omega([s,t]\times[s',t'])^{2/\rho}\end{aligned} \tag{41}$$

and since $d_{p-var;[0,1]}(S_3(X_n),\mathbf{X}) \to 0$ in probability, $d_{p-var;[0,1]}(S_3(X^n)^{i,j}_{s,t}S_3(X^n)^{i,j}_{s',t'}, \mathbf{X}^{i,j}_{s,t}\mathbf{X}^{i,j}_{s',t'}) \to 0$ in distribution by Slutsky's lemma. So, in particular, $|\mathbb{E}[\mathbf{X}^{i,j}_{s,t}\mathbf{X}^{i,j}_{s',t'}] - \mathbb{E}[S_3(X^n)^{i,j}_{s,t}S_3(X^n)^{i,j}_{s',t'}]| \to 0$. Thus, by passage to the limit in (41) we get:

$$\mathbb{E}[\mathbf{X}^{i,j}_{s,t}\mathbf{X}^{i,j}_{s',t'}] \leq C\omega([s,t]\times[s',t'])^{2/\rho}$$

Doing the same along the rest of the proof of Proposition 15.28, Proposition 15.30 and Corollary 15.31, we prove their exact analogs but without the bounded variation sample paths assumption.

18. Using Kolmogorov's theorem to deduce Hölder norm estimate on the lifted path from moment estimates on the levels

For this section, we will make extensive use of the following lemma, without mentioning it.

Lemma 18.1. *On a probability space $(\Omega,\mathcal{F},\mathbb{P})$, we have $\|\cdot\|_{L^{p'}} \leq \|\cdot\|_{L^p}$ for $1 \leq p' < p \leq \infty$.*

Proof. Let $1 \leq p < q \leq \infty$, since $\frac{q}{q-p}$. Since $\frac{q}{p}$ are Hölder conjuguates, we have

$$\|1f^p\|_1 \leq \|1\|_{\frac{q}{q-p}}\|f^p\|_{\frac{q}{p}}.$$

And since $\|1\|_{\frac{q}{q-p}} = 1$ we get, taking the power $\frac{1}{p}$ of the above expression, we get

$$\|f\|_p \leq \|f\|_q.$$

□

Let us state the famous Kolmogorov's continuity theorem.

Proposition 18.2 (Proposition A.2 of [17])**.** *Let X be a process on $[0,T]$ and assume there exists $p \in \mathbb{N}^*, \epsilon > 0$ and a constant $c = c(p,\epsilon,T)$ such that, for all $0 \leq s < t \leq T$:*

$$\|X_{st}\|_p \leq c(t-s)^{\epsilon+2/p}$$

Then, for all $\gamma \in [0,\epsilon)$,

$$\left\|\sup_{0\leq s<t\leq T}\frac{|X_{st}|}{(t-s)^\gamma}\right\|_p \leq \frac{2c}{2^{-\gamma}-2^{-\epsilon}}(6\sqrt{2})^{\epsilon+2/p}.$$

We can consider $\frac{2}{2^{-\gamma}-2^{-\epsilon}}(6\sqrt{2})^{\epsilon+2/p}$ as a constant depending on (ϵ,p,γ).

Theorem 18.3. *Fix $H \in (0,1/2)$. Let $\alpha \in \left(\frac{1}{3},\frac{1}{2}\right), \beta \in (0,H)$.*

Let $\alpha' \in (\alpha,1/2)$ and suppose $\|X^{(i)}_{st}\|_p \leq C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{i+1}(t-s)^{i\beta+\alpha'}$ for all $s < t \in [0,1]$ and $p \in [1,\infty)$ (in practice M will be $|D|^\eta$). Let us fix $\gamma \in [0, i\beta+\alpha')$ (recall that $I = \{i \in \mathbb{N} | i\beta+\alpha < 1\}$ by 8.2 so $i\beta+\alpha \leq i\beta+\alpha' < 1$ so we take $\gamma = i\beta+\alpha$ in practice). Then, for all $p \in [1,\infty)$ we have:

$$\left\||X^{(i)}|_{\gamma-Hld;[0,T]}\right\|_p = \left\|\sup_{0\leq s<t\leq T}\frac{|X^{(i)}_{st}|}{(t-s)^\gamma}\right\|_p \leq (Mp^{\frac{1}{2}})^{i+1}C(\gamma,\alpha,\beta,H).$$

Similarly, from the proof of Proposition 3.1 of [17]*, Suppose $\|X^{(jk)}_{st}\|_p \leq C(H,\alpha,\beta)(Mp^{1/2})^{j+k+2}(t-s)^{(j+k)H+1}$ for all $s < t \in [0,1]$ and $p \in [1,\infty)$. Let us fix $\gamma \in [0,(j+k)\beta+2\alpha')$ (recall that $J = \{(j,k) \in \mathbb{N}^2 | (j+k)\beta+2\alpha < 1\}$ by 8.2 so $(j+k)\beta+2\alpha \leq (j+k)\beta+2\alpha' < 1$), then for all $p \in [1,\infty)$ we have, for $\gamma' := (j+k)\beta+2\alpha$,*

$$\left\||X^{(jk)}|_{\gamma'-Hld;[0,T]}\right\|_p = \left\|\sup_{0\leq s<t\leq T}\frac{|X^{(jk)}_{st}|}{(t-s)^{\gamma'}}\right\|_p \leq (Mp^{\frac{1}{2}})^{j+k+2}C(\gamma',\alpha,\beta,H).$$

Proof. We have $\|X^{(i)}_{st}\|_p \leq C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{i+1}(t-s)^{i\beta+\alpha'}$ for all $s < t \in [0,1]$ with M independent of p (in practice M will be $|D|^\eta$). so $c = C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{i+1} = C(H,\alpha,\beta,p,i)$.

Applying proposition 18.2 in this case yields $\epsilon+2/p = i\beta+\alpha'$ so $\epsilon = i\beta+\alpha'-\frac{2}{p}$. Therefore, Proposition A.2 of [17] yields, for all $\gamma \in [0,\epsilon) = [0, i\beta+\alpha'-\frac{2}{p})$ (in practice we will want $\gamma = i\beta+\alpha$):

$$\left\|\sup_{0\leq s<t\leq T}\frac{|X^{(i)}_{st}|}{(t-s)^\gamma}\right\|_p \leq \frac{2C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{i+1}}{2^{-\gamma}-2^{-\epsilon}}(6\sqrt{2})^{\epsilon+2/p} = \frac{2C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{i+1}}{2^{-\gamma}-2^{-i\beta-\alpha'+\frac{2}{p}}}(6\sqrt{2})^{i\beta+\alpha'} \tag{42}$$

Lemma 18.4. $\frac{2C(H,\alpha,\beta)}{2^{-\gamma}-2^{-i\beta-\alpha'+\frac{2}{p}}}(6\sqrt{2})^{i\beta+\alpha'}$ *decreases with p so is lower than or equal to $C(\gamma,\alpha,\alpha',\beta,H) :=$ $\max_{i\in i(\alpha,\beta)}\frac{2C(H,\alpha,\beta)}{2^{-\gamma}-2^{-i\beta-\alpha'+2}}(6\sqrt{2})^{i\beta+\alpha'}$, which does not depend on p anymore but only on $(\gamma,\alpha,\alpha',\beta,H)$. In fact, taking the max over $\alpha' \in (\alpha,\frac{1}{2})$ we can write $C(\gamma,\alpha,\beta,H)$ instead of $C(\gamma,\alpha,\alpha',\beta,H)$.*

Proof. As p increases to infinity, $\frac{2}{p}$ decreases to zero, so $-i\beta-\alpha'+\frac{2}{p}$ decreases to $-i\beta-\alpha'$ and $2^{-i\beta-\alpha'+\frac{2}{p}}$ decreases to $2^{-i\beta-\alpha'}$. Thus, $2^{-\gamma}-2^{-i\beta-\alpha'+\frac{2}{p}}$ increases to $2^{-\gamma}-2^{-i\beta-\alpha'}$ and $\frac{2C(H,\alpha,\beta)}{2^{-\gamma}-2^{-i\beta-\alpha'+\frac{2}{p}}}(6\sqrt{2})^{i\beta+\alpha'}$ decreases to $\frac{2C(H,\alpha,\beta)}{2^{-\gamma}-2^{-i\beta-\alpha'}}(6\sqrt{2})^{i\beta+\alpha'}$ □

Thus, (42) becomes, for all $\gamma\in[0,\epsilon)=[0,i\beta+\alpha'-\frac{2}{p})$:

$$\left\|\sup_{0\le s<t\le T}\frac{|X^{(i)}_{st}|}{(t-s)^\gamma}\right\|_p\le (Mp^{\frac{1}{2}})^{i+1}C(\gamma,\alpha,\alpha',\beta,H) \tag{43}$$

In (43) we had fixed p and let $\gamma\in[0,\epsilon)=[0,i\beta+\alpha'-\frac{2}{p})$. Now, let us do the converse:

Since $H<\frac{1}{2}$, we have $i\beta+\alpha<i\beta+\alpha'<iH+\frac{1}{2}$. For $p>\frac{2}{i\beta+\alpha'-\gamma}$, we have $i\beta+\alpha'-\frac{2}{p}>i\beta+\alpha$. For each of such values of p, choose $\gamma=i\beta+\alpha$ to apply (43).

Then, for all $p>\frac{2}{i\beta+\alpha'-\gamma}:=p_i(H,\gamma,\alpha',\beta)$ ($\gamma=i\beta+\alpha$ so $p_i(H,\gamma)=\frac{1}{\alpha'-\alpha}$ in practice)(following the notation q_0 of Theorem A.13 and Remark A.14 of [15]) we have:

$$\left\||X^{(i)}|_{\gamma-Hld;[0,T]}\right\|_p=\left\|\sup_{0\le s<t\le T}\frac{|X^{(i)}_{st}|}{(t-s)^\gamma}\right\|_p\le (Mp^{\frac{1}{2}})^{i+1}C(\gamma,\alpha,\alpha',\beta,H)$$

A fortiori, the above equation is true for $p\ge p(H,\gamma,\alpha',\beta):=\max_{i\in I(\alpha,\beta)}p_i(H,\gamma,\alpha',\beta)+1$ (independent of i).

Using 18.1, we have that for all $p\le p(H,\gamma,\alpha',\beta)$,

$$\begin{aligned}\left\||X^{(i)}|_{\gamma-Hld;[0,T]}\right\|_p&\le\left\||X^{(i)}|_{\gamma-Hld;[0,T]}\right\|_{p(H,\gamma,\alpha',\beta)}\le\Big(Mp(H,\gamma,\alpha',\beta)^{1/2}\Big)^{i+1}C(\gamma,\alpha,\alpha',\beta,H)\\ &=\left(M\Big(\frac{p(H,\gamma,\alpha',\beta)}{p}p\Big)^{1/2}\right)^{i+1}C(\gamma,\alpha,\alpha',\beta,H)\le\max_i p(H,\gamma,\alpha',\beta)^{\frac{i+1}{2}}C(\gamma,\alpha,\alpha',\beta,H)(Mp^{\frac{1}{2}})^{i+1}\\ &=C'(\gamma,\alpha,\alpha',\beta,H)(Mp^{\frac{1}{2}})^{i+1}\end{aligned}$$

Thus, for all $p\in[1,\infty)$ we have:

$$\left\||X^{(i)}|_{\gamma-Hld;[0,T]}\right\|_p=\left\|\sup_{0\le s<t\le T}\frac{|X^{(i)}_{st}|}{(t-s)^\gamma}\right\|_p\le (Mp^{\frac{1}{2}})^{i+1}C(\gamma,\alpha,\alpha',\beta,H)$$

□

Theorem 18.5. *Fix $H\in(0,1/2)$. Let $\alpha\in\left(\frac{1}{3},\frac{1}{2}\right),\beta\in(0,H)$ ($H<\frac{1}{4}$ is the case we are interested in).*

Let $\alpha'\in(\alpha,1/2)$ and suppose we are able to show that for all $i\in I,(j,k)\in J,s<t\in[0,1]$ and $p\in[1,\infty)$ (it suffices to show it for p big enough and then and use 18.1 and increase the constant):

$$\|X^{(i)}_{st}-Y^{(i)}_{st}\|_{L^p}\le C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{i+1}(t-s)^{i\beta+\alpha'}$$
$$\|X^{(jk)}_{st}-Y^{(jk)}_{st}\|_{L^p}\le C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{j+k+2}(t-s)^{(j+k)\beta+2\alpha'}$$

Then, for all $p\ge1$ we have:

$$\|d_{(\alpha,\beta)-Hol}(\boldsymbol{X},\boldsymbol{Y})\|_{L^p}\le C'(\alpha,\alpha',\beta,H)Mp^{\frac{1}{2}}$$

with $C'(\alpha,\beta,H)=C_{00}(\beta,H)+C_{01}(\alpha,H)+\sum_{i\in I^*}C(\alpha,\beta,i,H)+\sum_{(j,k)\in J}C(\alpha,\beta,j,k,H)$ each of the terms in the sum being a constant defined in the proof of 18.3.

Remark 18.6. *In the proof of Corollary 13.21 of* [15]*, they use the result of Theorem A.13 of* [15] *applying it with $\epsilon=c_1|D|^\eta$ and $M=\sqrt{q}$ which yields $|d_{\alpha-Hol}(S_2(B^D),\boldsymbol{B})|_{L^q}\le c_6q^{1/2}|D|^{\eta/2}$. Indeed, $\max(\epsilon,\epsilon^{1/N})=\max\left(c_1|D|,(c_1|D|)^{1/2}\right)$ because, for $|D|$ small enough, $c_1|D|\le1$.*

This is why they call this constant ϵ in the statement of Theorem A.13, it is because it is supposed to be small, that is it is supposed to "contain" $|D^\eta|$. In [15]*, they always apply their Theorem A.13 with $\epsilon=C(\alpha,\eta,T)$ that is $C_1(H,\alpha,\alpha',\beta,T)$ for us and with $M=\sqrt{q}$.*

We could have done the same as we just need a strictly positive power on $|D|$ but the way the above theorem 18.5 is stated makes us need upper bounds in $|D|^{\eta(i+1)}$ and $|D|^{\eta(j+k+2)}$.

Proof. suppose we are able to show that for all $i\in I,(j,k)\in J,s<t\in[0,1]$ and $p\in[1,\infty)$:

$$\|X^{(i)}_{st}-Y^{(i)}_{st}\|_{L^p}\le C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{i+1}(t-s)^{i\beta+\alpha'}$$
$$\|X^{(jk)}_{st}-Y^{(jk)}_{st}\|_{L^p}\le C(H,\alpha,\beta)(Mp^{\frac{1}{2}})^{j+k+2}(t-s)^{(j+k)\beta+2\alpha'}$$

Then, by Theorem 18.3 applied for each $i\in I(\alpha,\beta)$ with $\gamma(i)=i\beta+\alpha<i\beta+\alpha'<1$ we get for all $i\in I(\alpha,\beta)$ and all $p\in[1,\infty)$:

$$\left\||X^{(i)}|_{\gamma-Hld;[0,T]}\right\|_p=\left\|\sup_{0\le s<t\le T}\frac{|X^{(i)}_{st}|}{(t-s)^\gamma}\right\|_p\le (Mp^{\frac{1}{2}})^{i+1}C(\gamma(i),\alpha,\beta,H)=(Mp^{\frac{1}{2}})^{i+1}C(\alpha,\beta,H)$$

So, for all p such that $\frac{p}{i+1}\ge1$, that is, for all $p\ge i+1$ and a fortiori for all $p\ge n(\alpha,\beta)+1$ we have for all $i\in I(\alpha,\beta)$:

$$\left\|\sup_{0\le s<t\le T}\frac{|X^{(i)}_{st}-Y^{(i)}_{st}|}{(t-s)^{i\beta+\alpha}}\right\|_{L^{\frac{p}{i+1}}}\le M^{i+1}\Big(\frac{p}{i+1}\Big)^{\frac{i+1}{2}}C(\alpha,\beta,H)\le M^{i+1}p^{\frac{i+1}{2}}C_2(\alpha,\beta,H)$$

Similarly, let us apply Theorem 18.3 for each $(j,k)\in J(\alpha,\beta)$ with $\gamma(j,k)=(j+k)\beta+2\alpha<(j+k)\beta+2\alpha'<1$ we get for all $(j,k)\in J(\alpha,\beta)$ and all $p\in[1,\infty)$:

$$\Big\| |X^{(jk)}|_{\gamma-Hld;[0,T]}\Big\|_p = \Big\| \sup_{0\le s<t\le T}\frac{|X^{(jk)}_{st}|}{(t-s)^{\gamma}}\Big\|_p \le (Mp^{\frac12})^{j+k+2}C(\gamma(j,k),\alpha,\beta,H) = (Mp^{\frac12})^{j+k+2}C(\alpha,\beta,H)$$

So, for all p such that $\frac{p}{j+k+2}\ge 1$, that is, for all $p\ge j+k+2$ and a fortiori for $p\ge m(\alpha,\beta)+2$ we have for all $(j,k)\in J(\alpha,\beta)$:

$$\Big\|\sup_{0\le s<t\le T}\frac{|X^{(jk)}_{st}-Y^{(jk)}_{st}|}{(t-s)^{(j+k)\beta+2\alpha}}\Big\|_{L^{\frac{p}{j+k+2}}} \le M^{j+k+2}\Big(\frac{p}{j+k+2}\Big)^{\frac{j+k+2}{2}}C(\gamma,\alpha,\beta,H)\le M^{j+k+2}p^{\frac{j+k+2}{2}}C_3(\alpha,\beta,H)$$

Define $p(\alpha,\beta):=\big(m(\alpha,\beta)+2\big)\vee\big(n(\alpha,\beta)+1\big)$. We use the same trick as in the end of the proof of 18.3: to treat the case of $p\le p(\alpha,\beta)$, we use 18.1 and increase the constants. This yields that, for all $p\in[1,\infty)$, for all $i\in I(\alpha,\beta)$ and for all $(j,k)\in J(\alpha,\beta)$:

$$\begin{aligned}\Big\|\sup_{0\le s<t\le T}\frac{|X^{(i)}_{st}-Y^{(i)}_{st}|}{(t-s)^{i\beta+\alpha}}\Big\|_{L^{\frac{p}{i+1}}} &\le M^{i+1}p^{\frac{i+1}{2}}C_4(\alpha,\beta,H),\\ \Big\|\sup_{0\le s<t\le T}\frac{|X^{(jk)}_{st}-Y^{(jk)}_{st}|}{(t-s)^{(j+k)\beta+2\alpha}}\Big\|_{L^{\frac{p}{j+k+2}}} &\le M^{j+k+2}p^{\frac{j+k+2}{2}}C_4(\alpha,\beta,H)\end{aligned}\tag{44}$$

Recall the definition (33)

$$\begin{aligned}&d_{(\alpha,\beta)-Hol}(\tilde S^N(X,\hat X),\tilde S^N(Y,\hat Y))=\\ &\quad\|\hat X-\hat Y\|_\beta+\|X-Y\|_\alpha+\sum_{i\in I^*}\Big(\|\tilde X^{(i)}-\tilde Y^{(i)}\|_{i\beta+\alpha}\Big)^{1/(i+1)}+\sum_{(j,k)\in J}\Big(\|\tilde X^{(jk)}-\tilde Y^{(jk)}\|_{(j+k)\beta+2\alpha}\Big)^{1/(j+k+2)}\end{aligned}$$

We can upper bound the L^p-norm of this distance as follows:

$$\begin{aligned}&\|d_{(\alpha,\beta)-Hol}(\mathbf{X},\mathbf{Y})\|_{L^p}\\ &=\Big\|\|\hat X-\hat Y\|_\beta+\|X-Y\|_\alpha+\sum_{i\in I^*}\Big(\|\tilde X^{(i)}-\tilde Y^{(i)}\|_{i\beta+\alpha}\Big)^{\frac{1}{i+1}}+\sum_{(j,k)\in J}\Big(\|\tilde X^{(jk)}-\tilde Y^{(jk)}\|_{(j+k)\beta+2\alpha}\Big)^{\frac{1}{j+k+2}}\Big\|_{L^p}\\ &\le\Big\|\|\hat X-\hat Y\|_\beta\Big\|_{L^p}+\Big\|\|X-Y\|_\alpha\Big\|_{L^p}+\sum_{i\in I}\Big\|\Big(\|\tilde X^{(i)}-\tilde Y^{(i)}\|_{i\beta+\alpha}\Big)^{\frac{1}{i+1}}\Big\|_{L^p}\\ &\quad+\sum_{(j,k)\in J}\Big\|\Big(\|\tilde X^{(jk)}-\tilde Y^{(jk)}\|_{(j+k)\beta+2\alpha}\Big)^{1/(j+k+2)}\Big\|_{L^p}\\ &=\Big\|\|\hat X-\hat Y\|_\beta\Big\|_{L^p}+\Big\|\|X-Y\|_\alpha\Big\|_{L^p}+\sum_{i\in I}\Big\|\|\tilde X^{(i)}-\tilde Y^{(i)}\|_{i\beta+\alpha}\Big\|^{\frac{1}{i+1}}_{L^{\frac{p}{i+1}}}+\sum_{(j,k)\in J}\Big\|\|\tilde X^{(jk)}-\tilde Y^{(jk)}\|_{(j+k)\beta+2\alpha}\Big\|^{\frac{1}{j+k+2}}_{L^{\frac{p}{j+k+2}}}\end{aligned}$$

Applying (44) yields:

$$\begin{aligned}&\|d_{(\alpha,\beta)-Hol}(\mathbf{X},\mathbf{Y})\|_{L^p}\\ &\le Mp^{\frac12}C(\beta,H)+Mp^{\frac12}C(\alpha)+\sum_{i\in I}\Big(M^{i+1}p^{\frac{i+1}{2}}C(\alpha,\beta,H)\Big)^{\frac{1}{i+1}}+\sum_{(j,k)\in J}\Big(M^{j+k+2}p^{\frac{j+k+2}{2}}C(\alpha,\beta,H)\Big)^{\frac{1}{j+k+2}}\\ &=Mp^{\frac12}C(\beta,H)+Mp^{\frac12}C(\alpha)+\sum_{i\in I}Mp^{\frac12}C(\alpha,\beta,H)^{\frac{1}{i+1}}+\sum_{(j,k)\in J}Mp^{\frac12}C(\alpha,\beta,H)^{\frac{1}{j+k+2}}\\ &\le C'(\alpha,\beta,H)Mp^{\frac12}\end{aligned}$$

□

Corollary 18.6.1. *Coming from the conclusion of the small interval case* (39).
Fix $\alpha\in(0,\frac12)$ and $\beta\in(0,H)$. Then fix $\alpha'\in(\alpha,\frac12)$. Suppose we are able to show that for all $i\in I,(j,k)\in J,s<t\in[0,1]$ and $p\in[1,\infty)$:

$$\begin{aligned}\|X^{(i)}_{st}-X^{D,(i)}_{st}\|_{L^p}&\le p^{\frac{i+1}{2}}C(\alpha,\beta,H)|D|^{\frac12-\alpha'}(t-s)^{\frac{2i\beta+2\alpha'}{2}}\\ \|X^{(jk)}_{st}-X^{D,(jk)}_{st}\|_{L^p}&\le p^{\frac{j+k+2}{2}}|D|^{\frac12-\alpha'}C(\alpha,\beta,H)(t-s)^{(j+k)\beta+2\alpha'}\end{aligned}$$

Then, for all $p\ge1$ we have:

$$\|d_{(\alpha,\beta)-Hol}(\boldsymbol{X},\tilde S^N(X^D,\hat X^D))\|_{L^p}\le C'(\alpha,\beta,H)|D|^{\frac12-\alpha'}p^{\frac12}$$

with $C'(\alpha,\beta,H)=C_{00}(\beta,H)+C_{01}(\alpha,H)+\sum_{i\in I^*}C(\alpha,\beta,i,H)+\sum_{(j,k)\in J}C(\alpha,\beta,j,k,H)$ each of the terms in the sum being a constant defined in the proof of 18.3.

This is the analog to the following upper bound of Remark A.14 of [15]:

$$\|\|\mathbf{X}-\mathbf{Y}\|_{(\alpha,\beta)}\|_{L^p}\le C\max(\epsilon,\epsilon^{1/H})q^{\frac12}.$$

Indeed, in Remark A.14 of [15], the constant $\tilde C(r,\gamma,T,N)$ does not depend on q, the dependence in q of $M=Cq^{1/2}$ is fully contained in the $q^{1/2}$.

In [17], it is the same, the dependence in p is fully contained in the $p^{\frac{i+1}{2}}$ and, for 18.2, $c = C_H p^{\frac{i+1}{2}} = C(H,p,i)$.

19. Analysis reminders

Proposition 19.1. *Let $S \neq \varnothing$ be any non-empty set of real numbers.*

(i) Assume f is a continuous function whose domain contains S and $\sup S$. Then $f(\sup S)$ is an adherent point of the set $f(S) \stackrel{def}{=} \{f(s) : s \in S\}$.

(ii) If in addition to what has been assumed, the continuous function f is also an increasing or non-decreasing function, then it is even possible to conclude that $\sup f(S) = f(\sup S)$.

Proof. If $S \neq \varnothing$ is any non-empty set of real numbers then there always exists a non-decreasing sequence $s_1 \leq s_2 \leq \cdots$ in S such that $\lim_{n\to\infty} s_n = \sup S$. Similarly, there will exist a (possibly different) non-increasing sequence $s_1 \geq s_2 \geq \cdots$ in S such that $\lim_{n\to\infty} s_n = \inf S$.

(i) Expressing the infimum and supremum as a limit of a such a sequence allows theorems from various branches of mathematics to be applied. Consider for example the well-known fact from topology that if f is a continuous function and $s_1, s_2, \ldots$ is a sequence of points in its domain that converges to a point p, then $f(s_1), f(s_2), \ldots$ necessarily converges to $f(p)$. It implies that if $\lim_{n\to\infty} s_n = \sup S$ is a real number (where all $s_1, s_2, \ldots$ are in S) and if f is a continuous function whose domain contains S and $\sup S$,

$$f(\sup S) = f\left(\lim_{n\to\infty} s_n\right) = \lim_{n\to\infty} f(s_n) \tag{45}$$

(ii) For each n, $s_n \in S$ so $f(s_n) \in f(S)$ and so $f(s_n) \leq \sup f(S)$ by definition of the supremum. Thus, by (45), $f(\sup S) \leq \sup f(S)$. Conversely, for each $s \in S$, $s \leq \sup S$ and, since f is non-decreasing, $f(s) \leq f(\sup S)$ and by taking the supremum over S we get $\sup f(S) \leq f(\sup S)$. □

This may be applied, for instance, to conclude that whenever g is a real (or complex) valued function with domain $\Omega \neq \varnothing$ whose sup norm $\|g\|_\infty \stackrel{\text{def}}{=} \sup_{x\in\Omega} |g(x)|$ is finite, then for every non-negative real number $\|g\|_\infty^q \stackrel{\text{def}}{=} \left(\sup_{x\in\Omega} |g(x)|\right)^q = \sup_{x\in\Omega} (|g(x)|^q)$ since the map $f : [0,\infty) \to \mathbb{R}$ defined by $f(x) = x^q$ is a continuous non-decreasing function whose domain $[0,\infty)$ always contains $S := \{|g(x)| : x \in \Omega\}$ and $\sup S \stackrel{\text{def}}{=} \|g\|_\infty$. Although this discussion focused on sup, similar conclusions can be reached for inf with appropriate changes (such as requiring that f be non-increasing rather than non-decreasing).

Proposition 19.2. *We have* $\sup_{0\leq s<t\leq T} \max_{i=1,\ldots,N} f_i(s,t) = \max_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$

Proof. Let $(s_n^i, t_n^i) \subseteq \{0 \leq s < t \leq T\}^{\mathbb{N}}$ be i sequences such that $f_i(s_n^i, t_n^i) \to \sup_{0\leq s<t\leq T} f_i(s,t)$ for all i. Then, by continuity of $f := \max_{i=1,\ldots,N}$ on $\mathbb{R}^n$, we get

$$\max_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t) = \max_{i=1,\ldots,N} \lim_n f_i(s_n^i, t_n^i) = \lim_n \max_{i=1,\ldots,N} f_i(s_n^i, t_n^i)$$

For each n, $\max_{i=1,\ldots,N} f_i(s_n^i, t_n^i) \leq \sup_{0\leq s<t\leq T} \max_{i=1,\ldots,N} f_i(s,t)$ so $\max_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t) \leq \sup_{0\leq s<t\leq T} \max_{i=1,\ldots,N} f_i(s,t)$.

Conversely, $f_i(s,t) \leq \sup_{0\leq s<t\leq T} f_i(s,t)$ for all i. Thus, $\max_{i=1,\ldots,N} f_i(s,t) \leq \max_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$. By taking the supremum we get $\sup_{0\leq s<t\leq T} \max_{i=1,\ldots,N} f_i(s,t) \leq \max_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$. □

Proposition 19.3. *The supremum of a finite sum is equivalent to the finite sum of the suprema. More precisely, for fixed $N \in \mathbb{N}$,*

$$\frac{1}{N} \sum_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t) \leq \sup_{0\leq s<t\leq T} \sum_{i=1,\ldots,N} f_i(s,t) \leq \sum_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$$

Proof. We first have $\sup_{0\leq s<t\leq T} \sum_{i=1,\ldots,N} f_i(s,t) \leq \sum_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$.

On the other hand, we have

$$\sup_{0\leq s<t\leq T} \sum_{i=1,\ldots,N} f_i(s,t) \geq \sup_{0\leq s<t\leq T} \max_{i=1,\ldots,N} f_i(s,t).$$

Moreover, by 19.2, $\sup_{0\leq s<t\leq T} \max_{i=1,\ldots,N} f_i(s,t) = \max_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$. Then, since $\max_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t) \geq \frac{1}{N} \sum_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$, we get:

$$\sup_{0\leq s<t\leq T} \sum_{i=1,\ldots,N} f_i(s,t) \geq \frac{1}{N} \sum_{i=1,\ldots,N} \sup_{0\leq s<t\leq T} f_i(s,t)$$

□

Proposition 19.4. *For any real numbers $(x_i)_{i=1}^n$, we have*

$$\Big|\sum_{i=1}^n x_i\Big|^p \leq \begin{cases} \sum_{i=1}^n |x_i|^p & p \in (0,1] \\ n^{p-1}(\sum_{i=1}^n |x_i|^p) & p > 1 \end{cases}$$

Proof. For the first case, by triangle inequality and lemma 19.5, we have

$$\left|\sum_i x_i\right| \leq \sum_i |x_i| = \|x\|_1 \leq \|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

For the second case, recall, that, by Jensen, if f is a convex function on $\mathbb{R}$, then for all $\alpha_i \geq 0$ such that $\alpha_1 + \cdots + \alpha_n = 1$ and for all x_i we have:

$$\alpha_1 f(x_1) + \cdots + \alpha_n f(x_n) \geq f(\alpha_1 x_1 + \cdots + \alpha_n x_n).$$

Thus, since $|\cdot|^p$ is convex for $p > 1$, we have by triangle inequality:

$$\Big|\frac{\sum_{i=1}^n x_i}{n}\Big|^p \leq \Big(\frac{\sum_{i=1}^n |x_i|}{n}\Big)^p \leq \frac{\sum_{i=1}^n |x_i|^p}{n}$$

□

Lemma 19.5. *For any sequence $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $p > 0 \mapsto \|x\|_p$ is non-increasing. In particular, for $0 < p \leq q < \infty$,*

$$(\sum_i |x_i|^q)^{1/q} = \|x\|_q \leq \|x\|_p = (\sum_i |x_i|^p)^{1/p}.$$

Proof. To see why, one can easily prove that if $\|x\|_p = 1$, then $\|x\|_q^q \leq 1$ (bounding each term $|x_i|^q \leq |x_i|^p$), and therefore $\|x\|_q \leq 1 = \|x\|_p$. Next, for the general case, apply this to $y = x/\|x\|_p$, which has unit l^p norm, and conclude by homogeneity of the norm. □

20. Frequently used notations

$x = (x_t : t \in [0, T])$ a generic path with values in some metric space
D a dissection (tj) of $[0, T]$
$|D|$ the mesh of D, i.e. $\max_j |t_{j+1} - t_j|$
$\mathcal{D}([0, T])$ the set of all dissections of $[0, T]$
$C^{p-var}([0, T], E)$ continuous paths of finite p-variation, see 2.2
$C^{\alpha-Hol}([0, T], E)$ Hölder continuous path with exponent α, see 2.2

References

[1] E. Alos, J. A. Leon, and J. Vives, *On the short-time behavior of the implied volatility for jump-diffusion models with stochastic volatility - finance and stochastics*, Aug 2007.
[2] C. Bayer, P. Friz, and J. Gatheral, *Pricing under rough volatility*, Quantitative Finance, 16 (2016), pp. 887–904.
[3] C. Bayer, P. K. Friz, P. Gassiat, J. Martin, and B. Stemper, *A regularity structure for rough volatility*, Mathematical Finance, 30 (2020), pp. 782–832.
[4] C. Bayer, P. K. Friz, A. Gulisashvili, B. Horvath, and B. Stemper, *Short-time near-the-money skew in rough fractional volatility models*, 2018.
[5] M. H. A. Davis and V. Mataix-Pastor, *Negative libor rates in the swap market model*, Finance and Stochastics, 11(2):181-193, (Apr 2007).
[6] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Springer Berlin Heidelberg, 2010.
[7] J.-D. Deuschel, P. K. Friz, M. Maurelli, and M. Slowik, *The enhanced sanov theorem and propagation of chaos*, Stochastic Processes and their Applications, 128 (2018), pp. 2228–2269.
[8] J.-D. Deuschel and D. W. Stroock, *Large deviations*, Academic Press, 1989.
[9] B. Dupire, *Pricing with a smile*, Risk, Vol. 7, No. 1, 1994, pp. 18-20., (1994).
[10] O. El Euch, *Quantitative Finance under rough volatility*, PhD thesis, 2018. Thèse de doctorat dirigée par Rosenbaum, Mathieu Mathématiques financières à Sorbonne université.
[11] M. Forde and H. Zhang, *Asymptotics for rough stochastic volatility models*, SIAM Journal on Financial Mathematics, 8 (2017), pp. 114–145.
[12] P. Friz and M. Hairer, *A Course on Rough Paths: With an Introduction to Regularity Structures*, Universitext, Springer International Publishing, 2020.
[13] P. K. Friz, P. Gassiat, and P. Pigato, *Precise asymptotics: Robust stochastic volatility models*, The Annals of Applied Probability, 31 (2021).
[14] P. K. Friz, P. Gassiat, and P. Pigato, *Short-dated smile under rough volatility: asymptotics and numerics*, Quantitative Finance, 22 (2022), pp. 463–480.
[15] P. K. Friz and N. B. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2010.
[16] M. Fukasawa, *Volatility has to be rough*, Quantitative Finance, 21 (2021), pp. 1–8.
[17] M. Fukasawa and R. Takano, *A partial rough path space for rough volatility*, 2022.
[18] J. Gatheral, T. Jaisson, and M. Rosenbaum, *Volatility is rough*, Quantitative Finance, 18 (2014), pp. 933–949.
[19] S. Gerhold, A. Jacquier, M. Pakkanen, H. Stone, and T. Wagenhofer, *Pathwise large deviations for the rough bergomi model: Corrigendum*, Journal of Applied Probability, 58 (2021), p. 849–850.
[20] M. Gubinelli, *Controlling rough paths*, Journal of Functional Analysis, 216 (2004), pp. 86–140.
[21] M. Gubinelli, *Ramification of rough paths*, Journal of Differential Equations, 248 (2010), pp. 693–721.
[22] A. Gulisashvili, *Large deviation principle for volterra type fractional stochastic volatility models*, SIAM Journal on Financial Mathematics, 9 (2018), pp. 1102–1136.
[23] A. Gulisashvili, F. Viens, and X. Zhang, *Extreme-strike asymptotics for general gaussian stochastic volatility models*, Annals of Finance, 15 (2019), pp. 59–101.
[24] ———, *Small-time asymptotics for gaussian self-similar stochastic volatility models*, Appl. Math. Optim., 82 (2020), p. 183–223.
[25] P. S. Hagan, D. M. S. Kumary, A. S. Lesniewski, and D. E. Woodward, *Managing smile risk*, 2002.
[26] M. Hairer, *A theory of regularity structures.*, Invent. Math. 198, no. 2, 269-504., (2014).
[27] S. L. Heston, *A closed-form solution for options with stochastic volatility with applications to bond and currency options*, The Review of Financial Studies, 6 (1993), pp. 327–343.
[28] J. Hull and A. White, *One-factor interest-rate models and the valuation of interest-rate derivative securities*, Journal of Financial and Quantitative Analysis, 28 (1993), p. 235–254.
[29] A. Jacquier, M. S. Pakkanen, and H. Stone, *Pathwise large deviations for the rough bergomi model*, Journal of Applied Probability, 55 (2018), pp. 1078–1092.
[30] A. Jacquier and A. Pannier, *Large and moderate deviations for stochastic volterra systems*, Stochastic Processes and their Applications, 149 (2022), pp. 142–187.
[31] L. T. John, *On the non-existence of path integrals*, Proc. R. Soc. Lond. A432281–290, (1991).
[32] I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus*, Springer, 2019.
[33] T. Lyons and Z. Qian, *System Control and Rough Paths*, Oxford University Press, December 2002.
[34] T. J. Lyons, *Differential equations driven by rough signals*, Rev. Mat. Iberoam. 14, no. 2, pp. 215-310, (1998).

[35] B. B. Mandelbrot and J. W. V. Ness, *Fractional brownian motions, fractional noises and applications*, SIAM Review, 10 (1968), pp. 422–437.

[36] S. Ninomiya and N. Victoir, *Weak approximation of stochastic differential equations and application to derivative pricing*, Applied Mathematical Finance, 15(2):107-121, (2008).

[37] Y. Y. Nobuyuki Ikeda, Shintaro Nakao, *A class of approximations of brownian motion.*, Publ. Res. Inst. Math. Sci. 13, no. 1, pp. 285-300, (1977).

[38] G. Shevchenko, *Fractional brownian motion in a nutshell*, arXiv: Probability, (2014).

[39] T. L. Terry J. Lyons, Michael Caruana, *Differential Equations Driven by Rough Paths, Ecole d'Eté de Probabilités de Saint-Flour XXXIV-2004*, Springer, 2007.