



university of
 groningen

faculty of science
and engineering

mathematics and applied
mathematics

Differential equations driven by irregular signals

Bachelor's Project Mathematics

June 2018

Student: N.L. Overmars

First supervisor: dr. D. Rodrigues-Valesin

Second assessor: prof. dr. A.C.D. van Enter

Differential equations driven by irregular signals

Nigel Overmars

August 29, 2018

Abstract

In this thesis we will develop a theory that allows us to solve differential equations driven by irregular signals. With fractional Brownian motions in our mind, we use the Young integration theory to determine when we can expect existence and/or uniqueness of such equations. We will also solve some equations, both numerically and explicit. Finally, we discuss an extension of the standard Black-Scholes model and show how it not suitable for praxis in the basic form.

Contents

1	Introduction	1
2	Preliminaries	3
2.1	Analysis of the space of functions of bounded p -variation . . .	3
2.2	Probability theory	13
2.3	Tensors on finite dimensional vector spaces	16
2.3.1	Tensors as homogeneous non-commuting polynomials	18
2.3.2	Taylor's theorem for multivariate functions	18
3	Fractional Brownian motion	20
3.1	Calculating the Hurst parameter for financial time series data	26
4	Young's integral	28
4.1	Controls	28
4.2	Constructing Young's integral	29
5	Differential equations driven by irregular signals	33
5.1	Existence	34
5.2	Uniqueness	35
5.3	Some remarks and further questions	42
6	Computing solutions	44
6.1	Explicit solutions	44
6.2	Numerically approximating solutions	45
6.2.1	Euler-Maruyama scheme	45
6.2.2	Crank-Nicholson	46
6.3	Numerical results for the Euler-Maruyama scheme	48
7	The fractional Black-Scholes model	51
7.1	A brief introduction to mathematical finance	51
7.2	Arbitrage in the fractional Black-Scholes model	53
8	Summary	55
A	R code for estimating convergence rate of Euler-Marayama scheme	56

1 Introduction

Differential equations are one of the most important concepts in mathematics with applications in nearly every field, from physics to economics to sociology. Using differential equations, we can explain the world around us in a concise but precise manner.

The most well-known class of differential equations is the class of Cauchy problems, usually written as $y' = f(x, y)$, $y(0) = \xi_0$. The theory around this equation is vast but well-known. It is known when one can guarantee existence of a solution, which follows from a theorem of Peano, and when one knows that there is one and only one solution, due to a theorem from Picard and Lindelöf.

In this thesis we will consider a generalization of the aforementioned concept, the stochastic differential equation (SDE). These are usually written as

$$dY_t = f(Y_t)dX_t, \quad Y_0 = \xi \quad (1)$$

Where Y_t is the unknown function, which depends on time (t). The above notation is technically an abuse of notation. The stochastic process X_t is usually far from differentiable, so we do not immediately know how we can define the differential of X_t , dX_t . It should be noted that Paul Malliavin developed a theory in which taking the derivative of a stochastic process makes sense, the framework of Malliavin Calculus. We refer the reader to [44] for a nice introduction to this theory.

In what follows, we will understand (1) as the following integral equation

$$Y_t = \int_0^t f(Y_t)dX_t \quad (2)$$

There are two theories developed to evaluate such integrals. The most well known is the theory by K. Itô. It assumes that X_t is a martingale¹ and then he showed that the Riemann-Stieltjes sums converge in probability to the proper quantity. Even though the assumption of being a martingale is usually not too restrictive, as the most widely used stochastic process, the Brownian motion, satisfies this property, there is still a large class of stochastic processes that are not martingales. The approach we take allows us to consider these stochastic processes.

The main question that we will try to answer is how we can define (2) and use it to solve (1). It should be clear that the ordinary Riemann

¹It is sufficient for X_t to be a semi-martingale, but since we don't need this we will limit ourself to martingales.

and Lebesgue integrals offer no answers, as those only allow us to integrate against t . Since X_t can be seen as a function, we will use the Riemann-Stieltjes integral as a starting point. The classical variant is only defined when X_t is of bounded variation, since most stochastic processes, for example the Brownian motion, have unbounded variation, this is only of limited use to us.

Young developed an extension of the Riemann-Stieltjes integral in 1936, which we will call the Young integral. He proved that if X has finite p -variation, which is a generalization of bounded variation, Y has finite q -variation, then the integral $\int_0^t Y_s dX_s$ exists if $1/p + 1/q > 1$. It should be noted that this implies that both p and q must be less than 2. Even though this is still a significant restriction, it does allow us to integrate against certain stochastic processes. We will consider the generalization of a Brownian motion, the fractional Brownian motion. For this stochastic process, the regularity depends on a parameter, so we can choose for what values it should be of bounded p -variation.

As noted above, we are still limited to functions of bounded p -variation for $p < 2$. We will show that this is not a shortcoming of the Young integral, but has a deep significance. The case $p \geq 2$ is completely different and needs a more sophisticated theory.

In the 1990s, Terry Lyons provided this theory, which is now known as the theory of rough paths. A rough path is defined as an element of the completion of the space of continuous paths with bounded p -variation and some other technical properties. The theory of rough paths has been fruitful. In 2014, Martin Hairer received the Fields medal for his construction of a robust solution theory of the Kardar–Parisi–Zhang (KPZ) equation using rough paths and even proposed a further generalization. If the reader is interested in this theory, we refer to [17] for an detailed study and to [28] for a more gentle introduction.

In this thesis, we will answer three questions: First we will show how you can define (2) in the sense of Young. We will also state and proof the conditions that imply the existence and uniqueness of (1).

We will numerically approximate solutions to (1) and compare different algorithms. Lastly, we shall look at the fractional Black-Scholes model, which is an extension of the standard Black-Scholes option pricing model.

Before we start with the thesis itself, I would like to express my gratitude to my supervisor Daniel Valesin for many helpful comments, discussions and because he stuck with me, despite the bumpy process that was this project.

2 Preliminaries

In this section we will define some terminology and state the necessary knowledge for this thesis. We will look at some real and functional analysis, measure theoretic probability and an introduction to tensors.

2.1 Analysis of the space of functions of bounded p -variation

In this thesis we will deal with a class of mappings that cannot be integrated using the standard Riemann-Stieltjes integration theory. As a first step we look at how we can define integrals of the form $\int Y dX$ for a certain class of functions X and Y .

Let V and W be two Banach spaces, we denote the set of continuous (i.e. bounded) linear mappings from V to W by $L(V, W)$. We take $J = [0, T]$ to be the compact interval on which we will be working, where T can be seen as some final time.

Definition 1 ((Continuous) path). *Let $X : J \rightarrow E$ be a Banach space valued continuous function, we will call this a continuous path. Since we will assume continuity for all cases, we will mostly refer to simply a path. Instead of $X(t)$ we will write X_t for the evaluation of our continuous path at t . If it holds for some positive α that $|X_t - X_s| < C|t - s|^\alpha$ for all s, t , then we say that X is α -Hölder continuous.*

The reason why we use path instead of function will become clear later, when we have covered our introduction to stochastic processes. We now define one of the key ingredients of this thesis, the p -variation of a continuous path.

Definition 2 (p -variation). *Let $X : J \rightarrow E$ be a continuous path, we define the total p -variation of X by*

$$\|X\|_{p,J} = \left(\sup_{\Pi} \sum_{j=0}^{r-1} |X_{t_j} - X_{t_{j+1}}|^p \right)^{1/p}, \quad (3)$$

where $\Pi = \{0 = t_0, t_1, \dots, t_n = T\}$ is a partition of the interval J and the supremum is taken over all possible partitions, so not just the ones for which the size of mesh, $|\Pi| := \max_{0 \leq i \leq r-1} |t_i - t_{i+1}|$, goes to zero. If the total p -variation is finite for some function X , we say that it is of bounded p -variation. For the case $p = 1$, we simply say that X is of bounded variation and write $X \in BV(J)$.

The notation $\|\cdot\|_{p,J}$ is a bit misleading, since it is not actually a norm but only a semi-norm. It can be easily seen that for a constant nonzero path Y , we have $\|Y\|_{p,J} = 0$. We shall define the proper norm later on.

Sometimes we have to mention an underlying partition explicitly, in that case, where we have a partition Π , we will denote this by

$$\|X\|_{p,\Pi} := \left(\sum_{i=0}^{r-1} |X_{t_i} - X_{t_{i+1}}|^p \right)^{1/p}$$

We stress that we take the supremum over all possible partitions, not only the ones for which the size of the mesh goes to zero. When $p = 1$, this distinction does not matter, but if $p > 1$, it is not necessarily the same, as we will see in the next proposition.

Proposition 3. *For every continuous path, that might not be of bounded variation, if $p > 1$, then we can find a partition such that the total p -variation of that path corresponding to that partition goes to zero.*

Proof. Let $J = [a, b]$ be an interval. Given a partition Λ of J and points $x, y \in \Lambda$, we will write $x \sim y$ if there are no points of Λ between x and y .

Let $X : J \rightarrow \mathbb{R}$ be continuous. We will construct a partition Λ_0 of J satisfying

$$|X(x) - X(y)| \leq 1 \text{ for each } x, y \in \Lambda \text{ with } x \sim y.$$

To do so, first use uniform continuity of X to find $\delta > 0$ such that

$$x, y \in J, |x - y| \leq \delta \implies |X(x) - X(y)| < 1.$$

Let $x_0 = a$. In case $|X(x) - X(a)| < 1$ for all $x \in [a, b]$, then let $x_1 = b$ and $\Lambda_0 = \{x_0, x_1\} = \{a, b\}$; in this case, (2.1) is satisfied. Otherwise, let

$$x_1 = \min\{x \geq a : |X(x) - X(a)| = 1\};$$

then, in case $|X(x) - X(x_1)| < 1$ for all $x \geq x_1$, we can set $x_2 = b$ and $\Lambda_0 = \{a, x_1, b\}$, so that (2.1) is satisfied. Otherwise, let

$$x_2 = \min\{x \geq x_1 : |X(x) - X(x_1)| = 1\},$$

and so on. Note that, due to (2.1), we have $|x_{i+1} - x_i| \geq \delta$ for each i , so the procedure has to eventually end with some n so that $x_n = b$. Let K be the number of intervals in Λ (that is, Λ has $K + 1$ points).

Now that we have Λ_0 , we will define a sequence of partitions Λ_n , $n \in \{0, 1, \dots\}$, such that, for each n , Λ_n has $K2^n$ intervals and

$$x, y \in \Lambda_n, x \sim y \implies |X(x) - X(y)| \leq 2^{-n}$$

Assume that Λ_n is already defined. For each $x, y \in \Lambda_n$ with $x < y$ and $x \sim y$, since $|X(x) - X(y)| \leq 2^{-n}$, there exists some intermediate point $z \in (x, y)$ such that $|X(x) - X(z)| \leq 2^{-(n+1)}$ and $|X(y) - X(z)| \leq 2^{-(n+1)}$. We then define Λ_{n+1} by including all the points of Λ_n , together with intermediate points chosen as we just explained. It is then clear that Λ_{n+1} has twice as many intervals as Λ_n , and satisfied (2.1) with n replaced by $n+1$. Moreover,

$$\|X\|_{p, \Lambda_n} \leq K2^n \cdot (2^{-n})^p \xrightarrow{n \rightarrow \infty} 0$$

since $p > 1$. □

We will limit our study to the case $p \geq 1$. The next proposition shows why the case $0 < p < 1$ is not interesting.

Proposition 4. *Let $X : J = [0, T] \rightarrow E$ be a continuous path with bounded p -variation with $p < 1$. Then X is constant, i.e. $X_t = X_0$ for all t .*

Proof. Let $0 \leq u \leq T$ and let Π be a partition of J of size r , then

$$\begin{aligned} |X_u - X_0| &\leq \sum_{i=0}^{r-1} |X_{t_i} - X_{t_{i+1}}|^p \\ &\leq (\max |X_{t_i} - X_{t_{i+1}}|^{1-p}) \left(\sum_{i=0}^{r-1} |X_{t_i} - X_{t_{i+1}}|^p \right) \\ &\leq (\max |X_{t_i} - X_{t_{i+1}}|^{1-p}) \|X\|_{p, J}^p \end{aligned}$$

Since X is continuous, it is also uniformly continuous on $J = [0, T]$, since J is compact. This means that we can make $|X_{t_i} - X_{t_{i+1}}|$ as small as we want. Since, by assumption, X has bounded p -variation, it follows that

$$|X_u - X_0| = 0$$

So we can conclude that $X_t = X_0$ for all $t \in J$. □

Calculating the p -variation is in general quite cumbersome, as there are uncountably many possible partitions, but for the case $p = 1$ and when X is differentiable and its derivative is integrable, we have a standard results that $\|X\|_{1, J} = \int_J |X'_t| dt$. So we can interpret this as the vertical component of the arc-length. We shall now prove that a path that is α -Hölder continuous for $\alpha \in (0, 1)$ has finite $\frac{1}{\alpha}$ -variation.

Proposition 5. *A path that is α -Hölder continuous for $\alpha \in (0, 1)$ has finite $\frac{1}{\alpha}$ -variation.*

Proof. Let $X : J \rightarrow E$ be a path that is $\frac{1}{\alpha}$ -Hölder continuous and assume that J is bounded. Then,

$$\begin{aligned} \|X\|_{\alpha, J} &= \left(\sup_{\Pi} \sum_{j=0}^{r-1} |X_{t_j} - X_{t_{j+1}}|^\alpha \right)^{1/\alpha} \\ &\leq C \left(\sup_{\Pi} \sum_{j=0}^{r-1} |t_j - t_{j+1}| \right)^{1/\alpha} \\ &\leq C \times |J|^{1/\alpha} < \infty \end{aligned}$$

Hence, X is of bounded α -variation. \square

This proposition will be useful to us later, when we define fractional Brownian motion, for which it is much simpler to determine the Hölder continuity than the p -variation directly. We will also prove a partial converse to this proposition in section 4.

For some of the proofs that will follow we need the following lemma:

Lemma 6. *Let $(a_i)_{i=0}^n$ be a sequence of positive real numbers and $p \geq 1$, then*

1. $(\sum_{i=1}^n a_i^p)^{1/p}$ is decreasing in p ;
2. $\ln \sum_{i=1}^n a_i^p$ is convex in p

Proof. 1. Let $q \geq p$ be given. Without loss of generality, we can assume that $(\sum_{i=1}^n a_i^p)^{1/p} = 1$ by scaling. It follows that $\sum_{i=1}^n a_i^p = 1$. Hence $0 \leq a_i \leq 1$ which implies $a_i^p \geq a_i^q$. Summing and taking the sum to the power of $1/q$ yields the result.

2. The most straight forward method would be using Hölder's inequality, this proof can be found here [40]. We will reproduce the novel (probabilistic) proof from [15]. Let $f(i) = a_i$ and let $\mu(\{i\}) = 1$ be the counting measure so that

$$\sum_{i=1}^n a_i^p = \mu(f^p), \quad \text{and} \quad \varphi(p) = \ln \mu(f^p)$$

We have that $\frac{d}{dp}f^p = f^p \ln(f)$. Hence

$$\begin{aligned}\varphi'(p) &= \frac{\mu(f^p \ln(f))}{\mu(f^p)} \\ \varphi''(p) &= \frac{\mu(f^p \ln^2(f))}{\mu(f^p)} - \left(\frac{\mu(f^p \ln(f))}{\mu(f^p)} \right)^2\end{aligned}$$

Let $\mathbb{E}X := \mu(f^p \ln f)/\mu(f^p)$. Then $\varphi'(p) = \mathbb{E} \ln f$ and hence

$$\varphi''(p) = \mathbb{E} \ln^2 f - (\mathbb{E} \ln f)^2 = \text{Var} \ln f \geq$$

Hence $\varphi'' \geq 0$ and it follows that φ is convex. □

Now we state and prove some properties of $\|\cdot\|_{p,J}$.

Proposition 7. *Let $X : J \rightarrow E$ be a continuous path.*

1. *Let $\varphi : J \rightarrow J$ be a non-decreasing surjection. Then, for all $p \geq 1$ $\|X\|_{p,J} = \|X \circ \varphi\|_{p,J}$;*
2. *The function $p \mapsto \|X\|_{p,J}$ from $[1, \infty)$ to $[0, \infty]$ is non-increasing;*
3. *The function $p \mapsto \ln \|X\|_{p,J}^p$ is convex, and continuous on any interval where it is finite;*
4. *For all $p \geq 1$, $\|X\|_{p,J} \geq \sup_{t,s \in J} |X_t - X_s|$;*
5. *The p -variation is lower semi-continuous, that means: Let $(X(n))$ be a sequence of elements of $C^0(J, E)$, i.e. the linear space of continuous paths, which converges in the topology of pointwise convergence to a continuous path X . Then,*

$$\liminf_{n \rightarrow \infty} \|X(n)\|_{p,J} \geq \|X\|_{p,J}$$

Proof. 1. We will show this by showing that both quantities are greater or equal to each other and hence they can only be equal. Let $T_1 = \{\tau_i\}$ be a partition of J and then let $T = \{t_i\}$, with $t_i = \varphi(\tau_i)$. Since φ is nondecreasing and a surjection, T is also a partition of J . Therefore

(all is to the power of p , for notational convenience),

$$\begin{aligned}\|X \circ \varphi\|_{p, T_1}^p &= \sum_{i=0}^{r-1} |(X \circ \varphi)_{\tau_i} - (X \circ \varphi)_{\tau_{i+1}}| \\ &= \sum_{i=0}^{r-1} |X_{t_i} - X_{t_{i+1}}|^p \\ &= \|X\|_{p, T}^p \leq \|X\|_{p, J}^p\end{aligned}$$

For the other inequality, let T_1 be a partition of J such that $t_i < t_{i+1}$ for all i , then there exists τ_i such that $t_i = \varphi(\tau_i)$, since φ is monotonic and surjective. We also have that $T = \{\tau_i\}$ is a partition of J . Hence

$$\|X\|_{p, T}^p = \|X \circ \varphi\|_{p, T_1}^p \leq \|X \circ \varphi\|_{p, J}^p$$

2. Let $q > p$. From the first item Lemma 6, it follows for any partition Π that

$$\|X\|_{q, \Pi} \leq \|X\|_{p, \Pi} \leq \|X\|_{p, J}$$

and since this holds for any partition, it follows that

$$\|X\|_{q, J} = \sup_{\Pi} \|X\|_{q, \Pi} \leq \|X\|_{p, J}$$

And hence $p \mapsto \|X\|_{p, J}$ is decreasing.

3. Consider the function $\varphi_{\Pi}(p) := \ln \|X\|_{p, \Pi}^p$. By the second item of Lemma 6, we have that φ_{Π} is convex. Let $p_0, p_1 \in [1, \infty)$ and let $\lambda \in [0, 1]$, then from the definition of convexity it follows

$$\varphi_{\Pi}(\lambda p_0 + (1 - \lambda)p_1) \leq \lambda \varphi_{\Pi}(p_0) + (1 - \lambda) \varphi_{\Pi}(p_1)$$

Since the logarithm is an increasing function, we can conclude that $\varphi_{\Pi}(p) \leq \varphi(p) := \ln \|X\|_{p, J}^p$ and hence we have that

$$\varphi_{\Pi}(\lambda p_0 + (1 - \lambda)p_1) \leq \lambda \varphi(p_0) + (1 - \lambda) \varphi(p_1)$$

By definition, $\sup_{\Pi} \varphi_{\Pi}(p) = \varphi(p)$, which allows us to conclude that

$$\begin{aligned}\varphi(\lambda p_0 + (1 - \lambda)p_1) &= \sup_{\Pi} \varphi_{\Pi}(\lambda p_0 + (1 - \lambda)p_1) \\ &\leq \lambda \varphi(p_0) + (1 - \lambda) \varphi(p_1)\end{aligned}$$

4. Let t_s and u_s be the points where the supremum is reached, where we assume without loss of generality that $t_s < u_s$ and let $\Pi = \{t_0, t_s, u_s, t_T\}$. Then we have

$$\begin{aligned}\|X\|_{p,J} &\geq \left(\sum_{\Pi} |X_{t_i} - X_{t_{i+1}}|^p \right)^{1/p} \\ &\geq (|X_{t_s} - X_{u_s}|^p)^{1/p} = \sup_{t,s \in J} |X_t - X_s|\end{aligned}$$

5. Let $\varepsilon > 0$ and let \mathcal{D}_ε be a partition of J such that

$$\left[\sum_{\mathcal{D}_\varepsilon} |X_{t_j} - X_{t_{j+1}}|^p \right]^{1/p} \geq \|X\|_{p,J} - \varepsilon$$

Since \mathcal{D}_ε is finite, we have that

$$\lim_{n \rightarrow \infty} \left[\sum_{\mathcal{D}_\varepsilon} |X_{t_j} - X_{t_{j+1}}|^p \right]^{1/p} \geq \|X\|_{p,J} - \varepsilon$$

The result follows from letting ε tend to 0. □

For $p \geq 1$, we will define a norm on the space of continuous functions with finite p -variation, $\mathcal{V}^p(J, E)$ which we may abbreviate to just \mathcal{V}^p . On this space we will define the norm ($X \in \mathcal{V}^p$)

$$\|X\|_{\mathcal{V}^p} = \|X\|_{p,J} + \sup_{t \in J} |X_t|$$

We will now state and prove some properties of the resulting normed space

Proposition 8. *For $p \geq 1$, the normed space $\mathcal{V}^p(J, E)$ is a linear subspace of $C^0(J, E)$, the space of continuous paths. Furthermore, $(\mathcal{V}^p(J, E), \|\cdot\|_{\mathcal{V}^p(J, E)})$ is a Banach space. Lastly, if $1 \leq p \leq q$, then the following inclusions hold*

$$\mathcal{V}^1(J, E) \subset \mathcal{V}^p(J, E) \subset \mathcal{V}^q(J, E) \subset C^0(J, E)$$

Proof. Let $X, Y \in \mathcal{V}^p(J, E)$, it should be clear that $\|X\|_{\mathcal{V}^p(J, E)} \geq 0$, $\|\lambda X\|_{\mathcal{V}^p(J, E)} = |\lambda| \|X\|_{\mathcal{V}^p(J, E)}$ and that $\|X\|_{\mathcal{V}^p(J, E)} = 0$ if and only if $X = 0$. The triangle inequality follows from $|X_{t_j} + Y_{t_j} - X_{t_{i+1}} - Y_{t_i}| \leq |X_{t_j} - X_{t_{i+1}}| + |Y_{t_j} - Y_{t_{i+1}}|$ and the fact that $\sup(X+Y) \leq \sup(X) + \sup(Y)$. This shows that $\|\cdot\|_{\mathcal{V}^p(J, E)}$ is a norm and that $\mathcal{V}^p(J, E)$ is a linear subspace of $C^0(J, E)$.

Now we show that it is a Banach space, let $(X_t(n))$ be a Cauchy sequence, it follows from

$$\|X_t(n) - X_t(m)\|_\infty \leq \|X(n) - X(m)\|_{j,E} + \sup_t |X_t(n) - X_t(m)| = \|X(n) - X(m)\|_{\mathcal{V}^p}$$

So $X_t(n)$ converges uniformly, and hence the limit function, say X_t , is continuous. So we only need to show that it has bounded variation. Since X_t is continuous and J is compact, it follows that $\|X\|_\infty < M$ for some $M > 0$. So we only need to show that $\|X_t\|_{p,J}$ is bounded, for which we proceed as follows: Let $\Pi = \{0 = t_0, t_1, \dots, t_{r-1} = T\}$ be a partition of J . By uniform continuity, there is a $m \geq 0$ such that $\|X - X(m)\|_\infty \leq \frac{1}{2n}$, hence

$$\begin{aligned} \sum_{j=0}^{r-1} |X_{t_{j+1}} - X_{t_{j+1}}(m)| &\leq \sum_{j=0}^{r-1} |X_{t_j} - X_{t_j}(m)| + \sum_{j=0}^{r-1} |X_{t_{j+1}} - X_{t_{j+1}}(m)| + \|X(m)\|_{p,J} \\ &\leq \frac{1}{2} + \frac{1}{2} + \sup_n \|X(n)\|_{p,J} \end{aligned}$$

This implies,

$$\|X\|_{p,J} \leq 1 + \sup_n \|X(n)\|_{p,J} < \infty$$

So we conclude that $X \in \mathcal{V}^p(J, E)$ and hence $\mathcal{V}^p(J, E)$ is a Banach space.

The inclusions follow Proposition 7. \square

One of the questions one could ask about \mathcal{V}^p is whether it is separable or not. The answer to this question is negative if J has more than one element or $E = \{0\}$. We will only consider the case $\mathcal{V}^p([0, T], \mathbb{R})$. The following proof comes from [18] and is slightly adapted to our notation.

Theorem 9. $\mathcal{V}^p(J, \mathbb{R})$ is not separable

Proof. We will construct an uncountable family of functions such that the distance between any two such functions stays bounded from below by a constant. Without loss of generality, let $T = 1$. We consider the following uncountable subset of $C([0, T], \mathbb{R})$,

$$f_\varepsilon(t) = \sum_{k \geq 1} \varepsilon_k 2^{-k/p} \sin(2^k \pi t)$$

Where $\varepsilon = (\varepsilon_k) \in \{1, -1\}^\mathbb{N}$. We will prove two things, first that f_ε is $\frac{1}{p}$ -Hölder continuous and then that $\|f_\varepsilon - f_{\varepsilon'}\|_{p,[0,1]} > 2$.

For $0 \leq s < t \leq 1$, we have

$$\begin{aligned} |f_\varepsilon(t) - f_\varepsilon(s)| &\leq \sum_{1 \leq k \leq |\log_2(t-s)|} \varepsilon_k 2^{-k/p} (\sin(2^k \pi t) - \sin(2^k \pi s)) \\ &\quad + \sum_{k > |\log_2(t-s)|} \varepsilon_k 2^{-k/p} (\sin(2^k \pi t) - \sin(2^k \pi s)) \end{aligned}$$

Since $\|\varepsilon\|_{\ell^\infty} \leq 1$, we have that $|\sin(2^k \pi t) - \sin(2^k \pi s)| \leq 2^k \pi |t - s|$ for the first sum and we use that $|\sin(x)| \leq 1$, hence

$$\begin{aligned} |f_\varepsilon(t) - f_\varepsilon(s)| &\leq \pi |t - s| \sum_{1 \leq k \leq |\log_2(t-s)|} 2^{-k/p} 2^k + \sum_{k > |\log_2(t-s)|} 2 \cdot 2^{-k/p} \\ &\leq c_1(p) |t - s|^{1/p} \end{aligned}$$

And hence $f_\varepsilon \in \mathcal{V}^p([0, 1], \mathbb{R})$. Now we show that can bound the distances between two elements of our set from below. Assume $\varepsilon \neq \varepsilon'$ and let $j \geq 1$ be the first index for which $\varepsilon_j \neq \varepsilon'_j$. Define the following partition of $[0, 1]$: $\mathcal{D} = \{t_i = i2^{-j-1}\}$ for $i = 0, \dots, 2^{j+1}$. Then it holds that

$$|\sin(2^j \pi t_{i+1}) - \sin(2^j \pi t_i)| = 1$$

Hence we have that $\|\sin(2^j \pi \cdot)\|_{p, [0, 1]} \geq 2^{j/p}$. Furthermore,

$$\begin{aligned} |(f_\varepsilon - f_{\varepsilon'})(t_{i+1}) - (f_\varepsilon - f_{\varepsilon'})(t_i)| &= |\varepsilon_j - \varepsilon'_j| 2^{-j/p} |\sin(2^j \pi t_{i+1}) - \sin(2^j \pi t_i)| \\ &= 2 \cdot 2^{-j/p} \end{aligned}$$

Hence it follows that $\|f_\varepsilon - f_{\varepsilon'}\|_{p, [0, 1]} \geq 2$, so we conclude that $\mathcal{V}^p([0, T], \mathbb{R})$ is not separable. \square

This has a far reaching consequence, as this limits our ability to approximate paths of bounded p -variation in the p -variation norm. Later we will see a result that for $q > p$, we can approximate in the q -variation norm. Before we can state that result, we first need to do some ground work. We consider approximations by piecewise linear paths. Let $X \in C^0(J, E)$ be a path and let \mathcal{D} be a partition of J . We denote by $X^\mathcal{D}$ the continuous path which coincides with X on the points of \mathcal{D} and is affine on the sub-intervals of J delimited by \mathcal{D} . Since there is only one way to linearly connect two adjacent points of \mathcal{D} , it follows that $X^\mathcal{D}$ is unique.

Proposition 10. *Let $X \in \mathcal{V}^p(J, E)$ and let \mathcal{D} be a partition of J , then*

$$\|X^\mathcal{D}\|_{p, J} \leq \|X\|_{p, J}$$

Proof. Let $\varepsilon > 0$ be given and let \mathcal{D}_ε be a partition of J such that $\|X^\mathcal{D}\|_{p,\mathcal{D}_\varepsilon} \geq \|X^\mathcal{D}\|_{p,J} - \varepsilon$. We will show that \mathcal{D}_ε can be chosen such that $\mathcal{D}_\varepsilon \subset \mathcal{D}$.

We proceed by contradiction, assume that the aforementioned inclusion does not hold. If \mathcal{D}_ε does not contain the endpoints of J , we add them. We note that this only can increase $\|X^\mathcal{D}\|_{p,\mathcal{D}_\varepsilon}$. Now assume that there is a time in \mathcal{D}_ε that is not in \mathcal{D} . We consider the smallest of such times, which we denote by u . Let t_i be the last time in \mathcal{D}_ε before u , t_j the last time in \mathcal{D} before u and v the first time after u in $\mathcal{D} \cup \mathcal{D}_\varepsilon$. Since $s \mapsto X_s^\mathcal{D}$ is affine on $[t_j, v]$, the function $s \mapsto |X_s^\mathcal{D} - X_{t_i}^\mathcal{D}|^p + |X_v^\mathcal{D} - X_s^\mathcal{D}|^p$ is convex on $[t_j, v]$ and must attain its maximum at one of the points t_j or v . If we then remove u from \mathcal{D}_ε and making sure that t_j or v , depending on where the function reaches its maximum, belongs to \mathcal{D}_ε , we do not decrease $\|X^\mathcal{D}\|_{p,\mathcal{D}_\varepsilon}$, but we do decrease the number of points in \mathcal{D}_ε , which are not in \mathcal{D} by one. If we repeat this procedure enough times, we can make sure that $\mathcal{D}_\varepsilon \subset \mathcal{D}$. Since X and $X^\mathcal{D}$ coincide on \mathcal{D}_ε , we can now conclude that

$$\|X^\mathcal{D}\|_{p,J} - \varepsilon \leq \|X^\mathcal{D}\|_{p,\mathcal{D}_\varepsilon} = \|X\|_{p,\mathcal{D}_\varepsilon} \leq \|X\|_{p,J}$$

Letting ε go to zero results into

$$\|X^\mathcal{D}\|_{p,J} \leq \|X\|_{p,J}$$

As required. \square

The following lemma is a straightforward estimation of the distance of two paths $X, Y \in \mathcal{V}^p$ in q -variation.

Lemma 11. *Let p, q such that $1 \leq p < q$ and let $X, Y \in \mathcal{V}^p(J, E)$. Then,*

$$\|X - Y\|_{\mathcal{V}^q(J,E)} \leq \left(\sup_{u \in J} |X_u - Y_u| \right)^{1 - \frac{p}{q}} \|X - Y\|_{p,J}^{\frac{p}{q}} + \sup_{u \in J} |X_u - Y_u|$$

Proof. This follows from $a^p = a^q a^{p-q}$, which gives us

$$\begin{aligned} \|X - Y\|_{\mathcal{V}^q(J,E)} &= \|X - Y\|_{q,J} + \sup_{u \in J} |X_u - Y_u| \\ &= \left(\sup_{\Pi} \sum_{j=0}^{r-1} |X_{t_j} - X_{t_{j+1}}|^q \right)^{1/q} + \sup_{u \in J} |X_u - Y_u| \\ &= \left(\sup_{\Pi} \sum_{j=0}^{r-1} |X_{t_j} - X_{t_{j+1}}|^p |X_{t_j} - X_{t_{j+1}}|^{q-p} \right)^{1/q} + \sup_{u \in J} |X_u - Y_u| \\ &\leq \left(\sup_{u \in J} |X_u - Y_u| \right)^{1 - \frac{p}{q}} \|X - Y\|_{p,J}^{\frac{p}{q}} + \sup_{u \in J} |X_u - Y_u| \quad \square \end{aligned}$$

We are now in the position to state and prove our approximation result. This result will be used a lot in the section on the Young integral in this thesis. Since \mathcal{V}^p is not separable, we expect that this is the best we can do. We denote by $|\mathcal{D}|$ the maximum size of the mesh.

Theorem 12. *Let p and q be such that $1 \leq p < q$ and let $X \in \mathcal{V}^p(J, E)$. Then the paths $X^\mathcal{D}$ converge to X in the q -variation norm as the mesh of \mathcal{D} goes to zero. In other words, if for all $\varepsilon > 0$ there exists a $\delta > 0$ such that, if \mathcal{D} is a partition of J with $|\mathcal{D}| < \delta$, then $\|X^\mathcal{D} - X\|_{\mathcal{V}^q(J, E)} < \varepsilon$*

Proof. Let \mathcal{D} be a partition of J . By Lemma 11, we have

$$\|X^\mathcal{D} - X\|_{\mathcal{V}^q(J, E)} \leq \left(\sup_{u \in J} |X_u^\mathcal{D} - X_u| \right)^{1 - \frac{p}{q}} \|X^\mathcal{D} - X\|_{p, J}^{\frac{p}{q}} + \sup_{u \in J} |X_u^\mathcal{D} - X_u|$$

Since X is uniformly continuous on J , we can make $\sup_{u \in J} |X_u^\mathcal{D} - X_u|$ as small as we want, by taking the mesh of \mathcal{D} small enough. So we only need to show that we can uniformly bound the p -variation norm. By Proposition 10 and the fact that $\|X^\mathcal{D} - X\|_{p, J}^p \leq 2^{p-1}(\|X^\mathcal{D}\|_{p, J}^p + \|X\|_{p, J}^p) \leq 2^p \|X\|_{p, J}^p$, we can indeed uniformly bound this quantity and the result follows. \square

For finite dimensional spaces, this has the following important corollary

Corollary 13. *Assume that E is finite dimensional and let p, q be such that $1 \leq p < q$. Let $\mathcal{X} \subset \mathcal{V}^p(J, E)$ be bounded. If \mathcal{X} is uniformly equicontinuous, then it is relatively compact in $\mathcal{V}^q(J, E)$.*

Proof. Since \mathcal{X} is bounded in \mathcal{V}^p and equicontinuous, it is also relatively compact in the uniform topology. Hence, from every sequence in \mathcal{X} one can extract a uniformly convergent sequence, which converges in \mathcal{V}^q . \square

2.2 Probability theory

Developed in the first half of the previous century, measure theoretic probability can be seen as a mathematical formalization of probability theory using the language of measure theory and functional analysis. In all that follows, assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space when not otherwise mentioned explicitly. This means that Ω is a set, \mathcal{F} is a σ -algebra and \mathbb{P} is a measure for which it holds that $\mathbb{P}(\Omega) = 1$.

We will provide only a short and high level introduction to the theory of measure theoretic probability. For an excellent book-length treatment we refer the reader to [41]. We will not try to make this section fully rigorous

and a significant part of the mathematical theory has been omitted, as we expect the reader to be comfortable with standard probability. This includes for example (conditional) expectation, but the reader can simply substitute his knowledge of standard probability theory.

The definition of a random variable is as follows:

Definition 14 (Random variable). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let (E, \mathcal{E}) be a measurable space. An (E, \mathcal{E}) -valued random variable is a function $X : \Omega \rightarrow E$ which is measurable. If $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}(\mathbb{R})$, which will be usually the case, then we call X simply a random variable.*

The main object that we will study is a stochastic process, which is a special kind of random variable, as we will see in the next definition.

Definition 15 (Stochastic process). *Let \mathbb{T} be a set. A stochastic process X_t , $t \in \mathbb{T}$, defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a family of random variables. This means that for every $t \in \mathbb{T}$, the function $X_t(\cdot)$ is a random variable. For a single $\omega \in \Omega$, we say that $t \mapsto X_t(\omega)$ is a realization or path of the stochastic process.*

We call \mathbb{T} the time-indexing set. We will mostly use $\mathbb{T} = [0, T]$, but other possibilities are for example \mathbb{N}, \mathbb{Z} and $[-T, T]$. An example of a stochastic process, consider the following: Take $\mathbb{T} = \mathbb{N}$, and define X_t as the process of flipping a fair coin at every $t \in T$. This can be mathematically modeled as a $X_t \sim \text{Bernoulli}(\frac{1}{2})$ for all $t \in \mathbb{T}$. This is one of the simplest examples of a stochastic process and has a couple of nice properties that we will define later in this thesis. Later we will see much more complicated examples of stochastic processes.

Given a stochastic process X_t , as times progresses, more ‘information’ becomes known about the stochastic process

Definition 16 (Filtration). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A filtration $(\mathcal{F}_t, t \in \mathbb{R})$ is defined to be a collection of sub- σ -algebras of \mathcal{F} such that $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$. Further, if \mathcal{F}_t satisfies*

1. $\mathcal{F}_t = \bigcap_{s > t} \mathcal{F}_s$. This is called the right-continuity criterion.
2. \mathcal{F}_0 contains all \mathbb{P} -null sets.

We say that \mathcal{F}_t is a standard filtration. We call the quadruplet $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ a filtered probability space.

Associated to a stochastic process X_t is the natural filtration, which is $\mathcal{F}_t = \sigma(X_s : 0 \leq s \leq t)$. This filtration holds all the information of the past of the stochastic process, but nothing more.

An important class of stochastic processes is the class of martingales.

Definition 17 (Martingale). *A stochastic process X_t with filtration \mathcal{F}_t a martingale if it holds that*

1. X_t is \mathcal{F}_t measurable for all admissible t ;
2. $\mathbb{E}|X_t| < \infty$;
3. $\mathbb{E}(X_t|\mathcal{F}_s) = X_s$ for $0 \leq s \leq t$.

If the last item holds with \geq (\leq), X_t is a supermartingale (submartingale).

Intuitively, a martingale is a process where the current state X_t is always the best prediction for its further states. In this sense, martingales describe fair games. Moreover, a martingale has the remarkable property that its expectation as a function of t is constant. This follows from

$$\mathbb{E}X_s = \mathbb{E}[\mathbb{E}(X_t|\mathcal{F}_s)] = \mathbb{E}X_t,$$

which holds for all s, t .

We note that there exists a generalization of a martingale, a so called semi-martingale. For the application of the Itô theory, it is enough to be a semi-martingale. As our focus on the Itô theory is limited, we will not develop this notion any further.

Even though a realization of a stochastic path might not be continuous, sometimes we can change this realization a little bit (we will define exactly in what sense in the definition) so that the resulting path is continuous.

Definition 18 (Modification). *Let $X, Y : [0, T] \rightarrow \Omega$ be two stochastic processes, we will say that X is a modification of Y if it holds that for all $t \in [0, T]$,*

$$\mathbb{P}(X_t = Y_t) = 1$$

We need this definition for the following theorem:

Theorem 19 (Kolmogorov continuity theorem). *Let X_t be a stochastic process. Suppose that there exists positive constants α, β, K such that*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leq K|t - s|^{1+\beta}$$

for all s, t . Then there exists a modification \tilde{X}_t of X_t that is continuous and furthermore it holds that these paths are γ -Hölder continuous for $0 < \gamma < \frac{\beta}{\alpha}$.

2.3 Tensors on finite dimensional vector spaces

When we will prove our main result, the uniqueness and existence of a differential equation driven by an irregular signal, we shall come across objects called tensors and tensors products. For completeness, we will provide an introduction to these objects. For most proofs we refer the reader to the literature.

For all that follows, assume that V and W are finite dimensional vector spaces over the field \mathbb{R} . These assumptions can be relaxed, but we won't need this. Let $\{e_1, \dots, e_n\}$ and $\{f_1, \dots, f_m\}$ be a basis of V and W , respectively.

A familiar operation on V and W is the direct sum, $V \oplus W$. A natural question to ask is: can we also take the product of two vector spaces in a way that is natural? The answer to this question is positive and is known as the tensor product. The reader might be already familiar with a tensor product without knowing it. In the case $V = W = \mathbb{R}^n$, the outer product $vw^T \in \mathbb{R}^{n \times n}$ for $v, w \in \mathbb{R}^n$ is a tensor as we will see later.

The tensor product $V \otimes W$ is defined to be the vector space with a basis of formal symbols $e_i \otimes f_j$, where we define these quantities as linearly independent. This means that an element of $V \otimes W$ can be written as the (formal) sum $\sum_{ij} c_{ij} e_i \otimes f_j$, where $c_{ij} \in \mathbb{R}$. Moreover, for any $v \in V$ and $w \in W$ we define $v \otimes w$ to be the element of $V \otimes W$ obtained by writing v and w in terms of the original bases of V and W and then expanding out $v \otimes w$ as if it were a non-commutative product (allowing any scalars to be pulled out).

As an example, take $V = W = \mathbb{R}^2$, with basis $\{e_1, e_2\}$. Then $\mathbb{R} \otimes \mathbb{R}^2$ is a four dimensional space with basis $\{e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2\}$. Also, let $v = e_1 - e_2$ and $w = e_1 + 2e_2$, then

$$\begin{aligned} v \otimes w &= (e_1 - e_2) \otimes (e_1 + 2e_2) \\ &= e_1 \otimes e_1 + 2e_1 \otimes e_2 - e_2 \otimes e_1 - 2e_2 \otimes e_2 \end{aligned}$$

Notice how explicit the basis of V and W are in this calculation. One could wonder, if we would have another basis for V and W , would this change anything? In other words, is the tensor product basis dependent? The answer to this question is negative. We will not provide a proof of this statement, but the reader is invited to try a different base on the previous multiplication and see that after changing back to the original basis, the result is the same.

We now list some properties of tensor products. Since we are working on finite dimensional vector spaces with a basis, the proofs are mostly trivial and omitted.

Proposition 20. *Let V and W be vector spaces of dimension n and m respectively, with basis $\{e_i\}$ and $\{f_j\}$ and let $v, v' \in V$ and $w, w' \in W$, then*

1. $V \otimes W$ is a vector space with basis $\{e_i \otimes f_j\}$;
2. $\dim(V \otimes W) = nm$;
3. $V \otimes W \cong W \otimes V$, in other words, they are isomorphic as vector spaces. This also means that the tensor product is symmetric;
4. $\otimes : V \times W \rightarrow V \otimes W$ is bilinear. In the case that $W = V$, this bilinear product is symmetric, see the property above;
5. $(w + w') \otimes v = w \otimes v + w' \otimes v$, $w \otimes (v + v') = w \otimes v + w \otimes v'$;
6. For $r \in \mathbb{R}$, $r(w \otimes v) = (rw) \otimes v = w \otimes (rv)$

We can also give meaning to higher order tensor products, for example $V \otimes V \otimes V$. For now, we will define this as $V \otimes (V \otimes V)$. Since it can be shown that there exists an isomorphism between $V \otimes (V \otimes V)$ and $(V \otimes V) \otimes V$, we will just write $V \otimes V \otimes V$.

Since even higher orders of tensor products will quickly become a notational burden, we will use the notation $V^{\otimes j}$ for the j -fold tensor product of V . Later on we will use this notation extensively. We will also note that there exists a j -linear symmetric map on $V^{\otimes j}$, which can be built from composing the lower order linear mappings. Lastly, observe that $\dim V^{\otimes j} = (\dim V)^j$, so in higher dimensions or tensor powers, thing can become quite unwieldy quickly. Hence we will sometimes use the Einstein notation. Instead of writing $\sum_i a_{ik} a_{ij}$, we will just simply write $a_{ik} a_{ij}$ and understand that we sum the repeated indices.

Normally we have that V is equipped with a norm $\|\cdot\|$. For further reference, we will state the properties of the norm on tensors product which we assume to be true

Definition 21. *Assume that V is a finite dimensional normed vector space. We say that its tensor powers are endowed with admissible norms if the following conditions hold:*

1. For all $n \geq 1$, the group of symmetric permutations S_n acts by isometries on $V^{\otimes n}$, i.e.

$$\|\sigma v\| = \|v\|, \quad v \in V^{\otimes n}, \quad \sigma \in S_n$$

2. The tensor product has norm 1, i.e. for all $n, m \geq 1$,

$$\|v \otimes w\| \leq \|v\| \|w\|, \quad v \in V^{\otimes n}, \quad w \in V^{\otimes m}$$

2.3.1 Tensors as homogeneous non-commuting polynomials

The above notions are quite abstract. We will use this section to give the reader a more intuitive and concrete exposition. We will think of the tensor powers of V as spaces of homogeneous non-commuting polynomials in a family of variables indexed by a basis of V . Let $\{v_1, \dots, v_n\}$ be a basis of V . Then a basis of $V^{\otimes j}$ is given by the set of tensors $v_I = v^{i_1} \otimes \dots \otimes v^{i_j}$, where $I = \{i_1, \dots, i_j\}$ spans $\{1, \dots, n\}^j$.

Hence, if $(a_I)_{I \in \{1, \dots, n\}^j}$ is a set of real numbers, then the tensor $\sum_I a_I v_I$ can be identified with the polynomial $\sum_I a_I X_I$ in the indeterminates X_1, \dots, X_n and $X_I = X_{i_1} \dots X_{i_j}$. It should be noted that all the terms in this polynomial have the same degree, namely j . If we have a sum of such polynomials with varying degrees but at most k , then we will have the truncated free algebra on V of order k , $T^k(V)$. In symbols this would be

$$T^k(V) = \bigoplus_{i=0}^k V^{\otimes i}$$

This object is very important in the study of the signature of a (rough) path and the corresponding rough path theory, which is an extension of the theory we will develop in this thesis.

2.3.2 Taylor's theorem for multivariate functions

In most undergraduate vector calculus classes, an extension of the standard Taylor expansion to multivariate functions is presented. The following presentation is usually used: Let $f \in C^\infty(\mathbb{R}^n, \mathbb{R})$ be a smooth function, let Df denote the Jacobian matrix and D^2f be the Hessian matrix, then we have for a fixed $h \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$,

$$f(x+h) = f(x) + Df(x)(h) + \frac{1}{2}D^2f(x)(h, h) + \dots$$

The form that we have written above is not entirely standard in undergraduate courses, but makes it explicit that the first derivative is a linear form, the second derivative is a linear 2-form and so on. We will use this form later on again.

Almost all the literature stops after the second term in the expansion. This makes sense, because writing higher order terms becomes a notational nightmare and is usually not necessary. Fully written out, the k^{th} -term has n^k terms. So the third order term of a function defined on \mathbb{R}^3 already has

$3^3 = 27$ elements. Even though some terms will be equal due to symmetry, one can imagine that this will become a mess quickly.

Tensors and tensor notation allow us to write it much more succinctly. Using Einstein notation, we can write for the k^{th} -derivative of f

$$D^k f = f_{I_k} dx^{\otimes I_k}$$

Where I_k spans $\{1, \dots, n\}^k$, hence $f_{i_k} = \partial_{i_1} \dots \partial_{i_k} f$ and $dx^{\otimes I_k} = dx^{i_1} \otimes \dots \otimes dx^{i_k}$. Using this notation and writing \mathbf{h}_i for the i -fold tuple consisting of h , we can now write the full Taylor series of f :

$$\begin{aligned} f(x+h) &= \sum_{i=0}^{\infty} \frac{1}{i!} D^i f(x)(\mathbf{h}_i) \\ &= f(x) + Df(x)(h) + \frac{1}{2!} D^2 f(x)(h, h) + \frac{1}{3!} D^3 f(x)(h, h, h) + \dots \end{aligned}$$

3 Fractional Brownian motion

Before we consider the two main parts of this thesis, namely the Young integral and differential equations driven by irregular signals, we shall make clear why the extension we provide is useful and not an academic exercise. We assume that the reader is well known with standard Brownian motions, if this is not the case, we refer the reader to [34] for a thorough introduction.

One of the most important properties of a Brownian motion is the independence of increments, meaning that past behavior has no influence on future behavior. This properties (with the zero mean) is fundamental to the fact that it is a martingale. And since it is a martingale, we can apply the Itô theory.

But there are many processes that can't be modeled as having independent increments. For example, take human behaviour. If an action gives positive utility, one is keen to keep repeating this behavior and possible do it more. In this case we have nonindependent and positively correlated increments.

Such behaviour pops up all over physics, biology and finance. Since in this case processes are described by differential equations, which we want to solve or even just know whether there exists (unique) solutions or not. Since we cannot apply the Itô theory, but we still want to deal with such equations, we need a new theory. In the next sections we develop this theory, but first we discuss a special class of stochastic processes, which will serve as an example in what follows.

In this section we will discuss a generalization of the Brownian motion, the factional Brownian motion, which was first mentioned by Kolmogorov in 1940, but he called it a Wiener spiral. The name fractional Brownian motion was proposed by Mandelbrot and Van Ness, which used a fractional integral to represent it. We will now define this class of stochastic processes.

Definition 22 (Fractional Brownian motion). *A fractional Brownian motion (fBm) with Hurst index $H \in (0, 1)$, B_t^H , is a continuous time (centered Gaussian) stochastic processes that starts at zero, has zero expectation and has covariance function $\mathbb{E}[B_t^H B_s^H] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H})$*

The parameter H decides how the increments are correlated. There are three possibilities, which we will list

- For $H = \frac{1}{2}$, the increments are uncorrelated.
- For $H < \frac{1}{2}$, the increments are negatively correlated.

- For $H > \frac{1}{2}$, the increments are positively correlated.

If H gets closer to 0 or 1, the stronger the negative resp. positive correlation is.

We shall now list some properties of fractional Brownian motions, for which the proofs can be found in the literature.

Proposition 23. *A fBm:*

1. *is self-similar, that is, $B_{at}^H \sim a^H B_t^H$ in the sense of probability distributions;*
2. *has stationary increments, that is, $B_t^H - B_s^H \sim B_{t-s}^H$;*
3. *exhibits long-range dependence if $H > 0.5$, meaning that*

$$\sum_{n=1}^{\infty} \mathbb{E}[B_1^H (B_{n+1}^H - B_n^H)] = \infty;$$

4. *has with probability one a Hausdorff and box dimension of $2 - H$.*

For a proof of its existence, we refer the reader to [34]. We note that if $H = \frac{1}{2}$ we have that $\mathbb{E}[B_t^H B_s^H] = s \wedge t$ which is a standard Brownian motion.

Proposition 24. *The fractional Brownian motion B^H has a continuous modification whose trajectories are γ -Hölder continuous for any $\gamma < H$*

Proof. For any $\alpha > 0$ we have

$$\mathbb{E}|B_t^H - B_s^H|^\alpha = \mathbb{E}|B_1^H|^\alpha |t - s|^{\alpha H} = K|t - s|^{1+\alpha H-1}$$

We can therefore apply the Kolmogorov continuity theorem, where the result follows if we let $\alpha \rightarrow \infty$. \square

From this result we recover the most important result in this section, namely that the paths of a fBm with Hurst parameter H have bounded $1/H$ -variation.

Proposition 25. *Let B_t^H be a fractional Brownian motion, then $B^H \in \mathcal{V}^{H+\varepsilon}(J, E)$ for $\varepsilon > 0$.*

Proof. This is a straight forward application of Proposition 24 and Proposition 5. \square

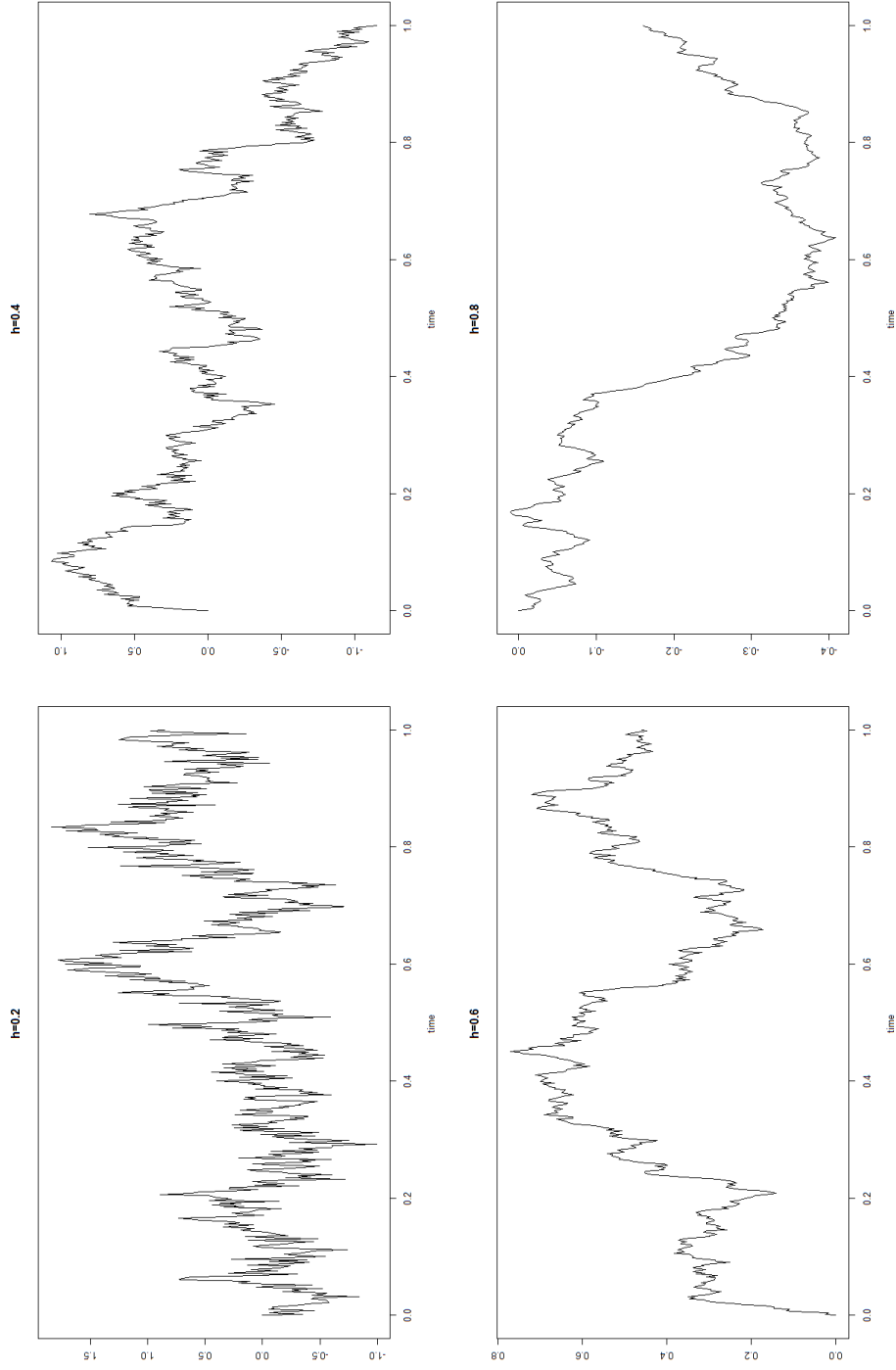


Figure 1: Four sample paths of fractional Brownian motions with different Hurst parameters.

The previous two results show that if H is lower, realizations of the process become more irregular. This makes sense, since if $H > \frac{1}{2}$, increments tend to keep doing what they were doing previously. So we don't expect a lot of jumping around. On the contrary, if $H < \frac{1}{2}$, the process becomes very stubborn. If it went down last increment, it wants to go up next time. The more H decreases, the more this behavior becomes apparent.

In this thesis most results only apply for the case $H > \frac{1}{2}$. We will also take briefly about the case $H < \frac{1}{2}$, but one could write another full length thesis on dealing with this case. The sample paths become so irregular that you need another fully new theory to deal with this case.

For applications, we need to know the Hurst parameter. It is essentially all we need to know about the process to describe it. We will now describe how one can estimate this parameter from a sample. Consider the set of observations $\{B_i^H\}_{i=1}^N$, where N is suitable large. We want to estimate H . First we will use a filter to reduce the dependence of the data. A filter of order r is a polynomial $a(x) = \sum_{k=0}^q a_k x^k$ such that $a^{(i)}(1) = 0$ for $0 \leq i \leq r \leq q$. Then we define the filtered observations as

$$B_n^a = \sum_{k=0}^q a_k B_{n+k}^H, \quad n = 1, 2, \dots, N - q$$

Popular filters are: $a(x) = x - 1$, $a(x) = \frac{1}{4}(x - 1)(x^2(1 - \sqrt{3}) - 2x)$ and $a(x) = (x - 1)^2$. The first two filters are of order one, the last is of order two. Consider now the covariance of a process which is filtered by a filter of

order r :

$$\begin{aligned}
\mathbb{E}[B_n^a B_m^a] &= \sum_{k=0}^q \sum_{j=0}^q a_k a_j \mathbb{E}[B_{n+k}^H B_{m+j}^H] \\
&= \frac{1}{2} \sum_{k=0}^q \sum_{j=0}^q a_k a_j ((n+k)^{2H} + (m+j)^{2H} - |m+k-n-j|^{2H}) \\
&= \frac{1}{2} \left(\sum_{k=0}^q a_k (n+k)^{2H} \sum_{j=0}^q a_j + \sum_{j=0}^q a_k (m+j)^{2H} \sum_{k=0}^q a_k \right. \\
&\quad \left. - \sum_{k=0}^q \sum_{j=0}^q a_k a_j |m+k-n-j|^{2H} \right) \\
&= -\frac{1}{2} \sum_{k=0}^q \sum_{j=0}^q a_k a_j |m-n+k-j|^{2H} \\
&=: \rho_H^a(m-n)
\end{aligned}$$

Where we used the fact that $\sum_{k=0}^q a_k = a(1) = 0$ (i.e. the polynomial evaluated at 1 is equal to the sum of coefficients which we have set to zero). Hence the filtered data $\{B_i^a\}_{i=1}^{N-q}$ is a stationary process.

We shall now define an estimator for the Hurst parameter. For $m \geq 1$, consider the dilated filter $a(x) = a(x^m) = \sum_{k=0}^q a_k x^{km}$. It follows that $\rho_H^{a^m} = m^{2H} \rho_H^a(0)$, or equivalently:

$$\log \rho_H^{a^m} = 2H \log m + \log \rho_H^a(0)$$

From this equation one can estimate H by using standard linear regression techniques, by regressing $\log \rho_H^{a^m}$ on $\log m$. Obviously we want a consistent estimator. It turns out that the empiric moments are suitable

Theorem 26. *The empiric variance*

$$V_N^{a^m} = \frac{1}{N - mq} \sum_{k=1}^{N-mq} (B_k^{a^m})^2$$

is a strongly consistent estimator of $\rho_H^{a^m}(0)$. This means that $V_N^{a^m} \xrightarrow{a.s.} \rho_H^{a^m}(0)$.

Even though the proof of this theorem is very short, it includes a theorem we have not covered, so for a proof we refer the reader to [36], where we warn the reader that there is a significant number of typos, so a careful reading is advisable.

Corollary 27. *Let a set $M \subset \mathbb{N}$ contain at least two elements and let $\hat{k}_N^{a,M}$ be the coefficient of linear regression of $\{\log V_N^{a^m} : m \in M\}$ on $\{\log m : m \in M\}$. Then the statistic $\hat{H}_N^{a,M} := \hat{k}_N^{a,M}/2$ is a strongly consistent estimator of H .*

Consider the following case, let $M = \{1, 2\} \subset \mathbb{N}$ and take $a(x) = x - 1$ as a filter. Then we have the following strongly consistent estimator

$$\hat{H}_N = \frac{1}{2 \log 2} \left(\log V_N^{d^2} - \log V_N^d \right) = \frac{1}{2} \log_2 \frac{V_N^{d^2}}{V_N^d}$$

with

$$V_n^{d^i} = \frac{1}{N-i} \sum_{k=1}^{N-i} (B_{k+i}^H - B_k^H)^2$$

It is up to the statistician to choose a proper filter and a suitable set M .

The procedure described above can also be used in the case our observations are scaled by an unknown constant c , so we have $\{cB_i^H\}$. In this case our regression equation will have an extra term $\log c$, which has no effect on the estimation process. Even more useful is that we don't need to have an integer valued grid. All we need is an equidistant grid, as the self-similarity property allows us to scale it back to an integer grid. But it should be noted that the aforementioned proof then only gives normal consistency, not strong consistency. Lastly, we note that $\hat{H}_N^{a,M}$ is an asymptotically normal estimator of H . For the details we refer once more to [36].

For applications one might also want to simulate some realizations of a fBm. We now describe the easiest way to do this. Even though it is easy to describe and implement, it is rather slow. In practice other algorithms are used, but those take too much time to develop. We refer the reader to [36].

The easiest method to simulate a fBm with Hurst parameter $H \in (0, 1)$ is the Cholesky decomposition method. We will consider a sample on the interval $[a, b]$ of size n . First we will define the covariance matrix, $\Gamma(n)$. Let $\gamma(t_1, t_2) = \frac{1}{2}(|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H})$. Then, if $\{t_i\}$ is a (uniform) partition of $[a, b]$ of size n , the covariance matrix $\Gamma(n)$ is defined by:

$$\Gamma(n) = \begin{pmatrix} \gamma(t_1, t_1) & \gamma(t_1, t_2) & \dots & \gamma(t_1, t_n) \\ \gamma(t_2, t_1) & \gamma(t_2, t_2) & \dots & \gamma(t_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(t_n, t_1) & \gamma(t_n, t_2) & \dots & \gamma(t_n, t_n) \end{pmatrix}$$

It can be proven that this matrix is positive definite and hence we decompose it in a lower triangular matrix $L(n)$, such that $L(n)L(n)^T = \Gamma(n)$ using the

Cholesky decomposition. This step takes $\mathcal{O}(n^3)$ flops and is quite expensive as n grows. Then, let $V(n)$ be a sample of n standard normal random variables. When the matrix $L(n)$ has been found, we can compute our the sample of our fBm as

$$U(n) = L(n)V(n)$$

Note that for each n , we only have to find the Cholesky decomposition once if we want multiple samples. If we just take an a.s. different set of draws from n standard normal random variables, say $V'(n)$, we just take $U'(n) = L(n)V'(n)$. Since vector-matrix multiplications have a complexity of $\mathcal{O}(n^2)$, this is significantly faster than applying the whole procedure again.

3.1 Calculating the Hurst parameter for financial time series data

As an example, we will now compute the Hurst parameter for three different time series. The data we took is from two stock markets, namely the Dutch AEX-500 and the NASDAQ, as well as the currency exchange market for USD and EUR. The first date of the data is 01/01/2010 and it ends at 06/06/2018. The series look as follows



Figure 2: We estimate a Hurst parameter of 0.5504.



Figure 3: We estimate a Hurst parameter of 0.5116.



Figure 4: We estimate a Hurst parameter of 0.5020.

4 Young's integral

In this section we will proof our first main result, the existence of the Young integral. Before we do so, we first consider the concept of controls, which allows us to control the p -variation of a path. After that we will construct the Young integral and state an existence theorem.

4.1 Controls

Let $X \in \mathcal{V}^p(J, E)$ for some $p \geq 1$ and let $s, t \in J$ such that $s \leq t$, define the function $\omega : \Delta_J \rightarrow \mathbb{R}$ by

$$\omega_X(s, t) = \|X\|_{p, [s, t]}^p$$

Where $\Delta_J = \{(s, t) \in J \times J : s \leq t\}$. It is obvious that this function $(s, t) \mapsto \omega_X(s, t)$ is non-negative and that $\omega_X(s, s) = 0$, i.e., it vanishes on the diagonal. It is non-decreasing in t and non-increasing in s . A key property of ω_X is the fact that it satisfies the following inequality, for $s \leq u \leq t$

$$\omega_X(s, u) + \omega_X(u, t) \leq \omega_X(s, t)$$

If a function posses this property, we will call it superadditive. Since X is a continuous path, ω_X is continuous in both s and t . We care about this function, as it is reparametrization of X that comes quite naturally. To see this, assume that X is not constant on a sub-interval of $J = [0, T]$. Then the function $t \mapsto \frac{\omega_X(0, t)}{\omega_X(0, T)}T$ is an increasing, continuous, bijection on J and hence has an inverse. Let $t \mapsto \tau(t)$ be this inverse. Then for all $(s, t) \in \Delta_J$,

$$\begin{aligned} |X_{\tau(s)} - X_{\tau(t)}|^p &\leq \omega_X(\tau(s), \tau(t)) \\ &\leq \omega_X(0, \tau(t)) - \omega_X(0, \tau(s)) \\ &= \frac{\omega_X(0, T)}{T}(t - s) \end{aligned}$$

We have proved a sort of converse to proposition 5, namely that every continuous path of bounded p -variation can be reparametrized to a Hölder continuous path with exponent $1/p$. This has as a consequence that sets of paths of uniformly bounded p -variation can be reparametrized to become uniformly equicontinuous.

This discussion is the motivation of a control function, which we define now.

Definition 28 (Control function). *A control function, or control, on $J = [0, T]$ is a continuous, non-negative function ω on Δ_J that is superadditive and vanishes on the diagonal.*

The function ω_X from above is of course a control function.

The following lemma is a direct result of Theorem 12 and Corollary 13:

Lemma 29. *For a control ω , the set of paths controlled in p -variation by ω is relatively compact in $\mathcal{V}^q(J, E)$*

We will need the following lemma in the next section.

Lemma 30. *Let ω be a control and let $X \in \mathcal{V}_{J,E}^p$. Assume that for some $p \geq 1$ and for all $(s, t) \in \Delta_J$, one has $|X_s - X_t|^p \leq \omega(s, t)$. Then, for all $(s, t) \in \Delta_J$, $\|X\|_{p,[s,t]} \leq \omega(s, t)^{\frac{1}{p}}$.*

Proof. For notational convenience, we will show that $\|X\|_{p,[s,t]}^p \leq \omega(s, t)$. Let X_t be as in the lemma and Π a partition of $[s, t]$, then

$$\begin{aligned} \|X\|_{p,[s,t]} &:= \sup_{\Pi} \sum_{\Pi} |X_{t_i} - X_{t_{i+1}}|^p \\ &\leq \sup_{\Pi} \sum_{\Pi} \omega(t_i, t_{i+1}) \\ &\leq \omega(s, t) \end{aligned}$$

Where in the last step we made use of the super additivity of ω . \square

If ω satisfies the conclusion of this lemma, we say that it controls the p -variation of X . This will be a useful concept in the next section, where we will introduce and develop the Young integral.

4.2 Constructing Young's integral

In this section we will prove the following theorem:

Theorem 31 (Young integral, 1936). *Let V and W be Banach spaces and let $p, q \geq 1$ be such that $1/p + 1/q > 1$. Let T be a positive real number. Consider $X \in \mathcal{V}^p([0, T], V)$ and $Y \in \mathcal{V}^q([0, T], L(V, W))$. then, for all $t \in [0, T]$, the limit*

$$\int_0^t Y_s dX_s = \lim_{|\mathcal{D}| \rightarrow 0, \mathcal{D} \subset [0, t]} \int_{\mathcal{D}} Y dX \quad (4)$$

exists and as a function of t belongs to $\mathcal{V}^p([0, T], W)$. Furthermore, there exists a constant $C_{p,q}$ that only depends on p and q such that the following inequality holds:

$$\left\| \int_0^\bullet (Y_s - Y_0) dX_s \right\|_{p,[0,T]} \leq C_{p,q} \|Y\|_{q,[0,T]} \|X\|_{p,[0,T]} \quad (5)$$

Before we prove this, we first provide some background. Since the Young integral is an extension of the Riemann-Stieltjes integral, we use similar terminology. Let X and Y be two paths defined, as usual, on the compact interval $J = [0, T]$. Let $\mathcal{D} = \{0 = t_0, \dots, t_r = T\}$ be a partition of J . We will use the following notation in this section.

$$\int_{\mathcal{D}} Y dX = \sum_{i=0}^{r-1} Y_{t_i} (X_{t_{i+1}} - X_{t_i})$$

We will consider only partitions that include both endpoints.

First of all, we define what we mean by the limit exists in (4): We say that the integral $\int_0^T Y_s dX_s$ exists if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that, if \mathcal{D} and \mathcal{D}' are two partitions of J such that $|\mathcal{D}| < \delta$ and $|\mathcal{D}'| < \delta$, then

$$\left| \int_{\mathcal{D}} Y dX - \int_{\mathcal{D}'} Y dX \right| < \varepsilon$$

It is well known that the Riemann-Stieltjes integral exists $\int X_t dX_t$ exists if X_t is of bounded variation. Theorem 4 allows us to consider the case where X_t may be only of bounded p -variation for $p < 2$. The price we have to pay for this is that we require more regularity of Y_t than just continuity. We shall now state the proof of theorem (4).

Proof. The proof is inspired by the proof of Proposition 10. We will find a maximal inequality that bounds $|\int_{\mathcal{D}} Y dX|$ uniformly with respect to \mathcal{D} , i.e., independent of \mathcal{D} . Like in the proof of the lemma, we will repeatedly remove points from \mathcal{D} until we reach this bound. We proceed as follows.

We replace X and Y by \tilde{X} and \tilde{Y} , where these are defined for all $u \in J$ by

$$\tilde{X}_u = \frac{X_u}{\|X\|_{p,J}}, \quad \tilde{Y}_u = \frac{Y_u}{\|Y\|_{q,J}}$$

I.e., we normalize the variation. Now, let ω be the following control function for all $(u, v) \in \Delta_J$

$$\omega(u, v) = \frac{\|X\|_{p,[u,v]}^p}{\|X\|_{p,J}^p} + \frac{\|Y\|_{q,[u,v]}^q}{\|Y\|_{q,J}^q}$$

It should be clear that $\omega \leq 2$ and that it controls both the p -variation of \tilde{X} and the q -variation of \tilde{Y} on J .

Let $\mathcal{D} = \{0 = t_0, \dots, t_r = T\}$ be a partition of $J = [0, T]$. We will successively remove points from \mathcal{D} in such a way that the variation of X and Y remains as evenly spread out as possible between the different sub-intervals induced by the partition. If \mathcal{D} has only three points, so when $r = 2$, we choose the middle point, t_1 . If $r > 2$, we choose t_i , where i is between 1 and $r - 1$ and such that

$$\omega(t_{i-1}, t_{i+1}) \leq \frac{2}{r-1} \omega(0, T) \quad (6)$$

It is guaranteed that such i exists. To see this, assume that the inequality is false and assume that r is even, then, using superadditivity

$$2\omega(0, T) < \omega(t_0, t_r) + \omega(t_1, t_r) \leq 2\omega(0, T)$$

A similar argument can be made if r is odd. Furthermore, observe that (6) holds even for $r = 2$. Now we consider the partition $\mathcal{D} \setminus \{t_i\}$. We introduce the following notation, $\tilde{X}_{u,v}^1 = \tilde{X}_v - \tilde{X}_u$. Then we have, using the telescoping property of the sums,

$$\begin{aligned} \left| \int_{\mathcal{D}} \tilde{Y} d\tilde{X} - \int_{\mathcal{D} \setminus \{t_i\}} \tilde{Y} d\tilde{X} \right| &= \left| \sum_{j=0}^{r-1} \tilde{Y}_{t_j} (\tilde{X}_{t_{j+1}} - \tilde{X}_{t_j}) - \sum_{\substack{j=0 \\ t_j \neq t_i}}^{r-1} \tilde{Y}_{t_j} (\tilde{X}_{t_{j+1}} - \tilde{X}_{t_j}) \right| \\ &\leq |\tilde{Y}_{t_{i-1}} \tilde{X}_{t_{i-1}, t_i}^1 + \tilde{Y}_{t_i} \tilde{X}_{t_i, t_{i+1}}^1 - \tilde{Y}_{t_{i-1}} \tilde{X}_{t_{i-1}, t_{i+1}}^1| \\ &= |(\tilde{Y}_{t_i} - \tilde{Y}_{t_{i-1}})(\tilde{X}_{t_{i+1}} - \tilde{X}_{t_i})| \\ &\leq \omega(t_{i-1}, t_i)^{\frac{1}{q}} \omega(t_i, t_{i+1})^{\frac{1}{p}} \\ &\leq \omega(t_{i-1}, t_{i+1})^{\frac{1}{p} + \frac{1}{q}} \\ &\leq \left(\frac{4}{r-1} \right)^{\frac{1}{p} + \frac{1}{q}} \end{aligned}$$

If we apply this procedure iteratively until we are left with a partition consisting of only two points, it follows that

$$\left| \int_{\mathcal{D}} (Y - Y_0) dX \right| \leq 4^{\frac{1}{p} + \frac{1}{q}} \zeta \left(\frac{1}{p} + \frac{1}{q} \right) \|Y\|_{q,J} \|X\|_{p,J}$$

Recall that the Riemann-zeta function ζ is defined as $\zeta(s) = \sum_{n \geq 1} n^{-s}$ and that it converges in the ordinary sense for $s > 1$. This explains why we need

$\frac{1}{p} + \frac{1}{q} > 1$. Otherwise we would not be able to construct our uniform bound. Define $c_{p,q} = 4^{\frac{1}{p} + \frac{1}{q}} \zeta \left(\frac{1}{p} + \frac{1}{q} \right)$. Then our uniform bound is:

$$\sup_{\mathcal{D} \subset [s,t]} \left| \int_{\mathcal{D}} Y dX \right| \leq c_{p,q} \|Y\|_{\mathcal{V}^q(J,E)} \|X\|_{p,J} \quad (7)$$

We now proceed by an approximation argument to show the existence of the integral over $J = [0, T]$. Let $p' > p$ such that $\frac{1}{p'} + \frac{1}{q} > 1$. From proposition Theorem 12 it follows that there exists a sequence $(X(n))_{n \geq 0}$ of piecewise linear paths that converge in p' -variation to X . It should be clear that for all n , $X(n) \in \mathcal{V}^1(J, E)$ and

$$\lim_{|\mathcal{D}| \rightarrow 0, \mathcal{D} \subset [s,t]} \int_{\mathcal{D}} Y dX(n) = \int_s^t Y_u dX_u(n)$$

The right hand side of this equation is a Riemann-Stieltjes integral, which exists since Y is continuous and $X(n)$ is of bounded variation. Our uniform bound (8) now implies that

$$\sup_{\mathcal{D} \subset [s,t]} \left| \int_{\mathcal{D}} Y dX(n) - \int_{\mathcal{D}} Y dX \right| \leq c_{p',q} \|Y\|_{\mathcal{V}^q(J,E)} \|X(n) - X\|_{p',J} \quad (8)$$

It follows that if $n \rightarrow \infty$, the supremum tends to 0. For convenience we introduce the notation $a \vee b = \max(a, b)$.

$$\lim_{\delta \rightarrow 0} \sup_{|\mathcal{D}| \vee |\mathcal{D}'| < \delta} \left| \int_{\mathcal{D}} Y dX - \int_{\mathcal{D}'} Y dX \right| = 0$$

Hence we have proven that $\int_0^T Y_u dX_u$ exists and is well-defined. \square

Sometimes we need the inequality of the next corollary, which is a direct consequence of equation (8)

Corollary 32.

$$\left\| \int_0^\bullet Y_s dX_s \right\|_{\mathcal{V}^p([0,T],W)} \leq 2c_{p,q} \|Y\|_{\mathcal{V}^q([0,T],W)} \|X\|_{p,[0,T]} \quad (9)$$

We have now proved the existence of the Young integral. In the next section we will apply it to prove theorems concerning (ordinary) differential equations driven by irregular signals.

5 Differential equations driven by irregular signals

In this chapter we will extend the results of the classical ODE theory. From now on, we assume that $X \in \mathcal{V}^p(J, W)$, where again $J = [0, T]$ and W is a Banach space, and that we have $p < 2$. For instance, X could be a fractional Brownian motion with Hurst parameter $H \in (\frac{1}{2}, 1)$. We showed in section 3 that this process is an element of $\mathcal{V}^{\frac{1}{H} + \varepsilon}(J, W)$ for all $\varepsilon > 0$. This will be our prime example why it makes sense and would be useful to solve such differential equations. Furthermore, since a fBm is not a martingale when $H \neq \frac{1}{2}$, the usual Itô stochastic integration theory does not apply.

In our current setting, we only have that X is of bounded p -variation for $p < 2$. We have shown that this implies that the solution Y must have (at least) bounded p -variation. For the differential equation and the corresponding integral equation to even make sense, we need that f is of bounded q -variation such that $1/p + 1/q > 1$. Otherwise the Young integral is not well defined. Secondly, we expect that we need more regularity of f for the existence let alone uniqueness of solutions. In this section we will develop and prove the necessary conditions for existence and uniqueness. Even though the proofs are similar to those in the classical case, they are significantly more technical.

First we identify a class of functions f such that $f(Y)$ has finite q -variation for $Y \in \mathcal{V}^p$. This is relatively straightforward. Let V and W be two Banach spaces, then we recall that $L(V, W)$ is the space of bounded linear operators from V to W .

Proposition 33. *Assume that f is Hölder continuous with exponent γ and $0 < \gamma \leq 1$. Also assume that $Y \in \mathcal{V}^p(J, W)$ for some $p \geq 1$. Then $f \circ Y \in \mathcal{V}^{\frac{p}{\gamma}}_\gamma(J, L(V, W))$ and there exists a $K > 0$ such that*

$$\|f \circ Y\|_{\mathcal{V}^{\frac{p}{\gamma}}_\gamma} \leq K \|Y\|_{\mathcal{V}^p}^\gamma$$

Proof. Assume that $K > 0$ is such that for all $w, w' \in W$ it holds that $|f(w) - f(w')| \leq K|w - w'|^\gamma$. Then, for every partition $\mathcal{D} \subset J$,

$$\|f \circ Y\|_{\mathcal{V}^{\frac{p}{\gamma}}_\gamma, \mathcal{D}} \leq K \|Y\|_{p, \mathcal{D}}^\gamma$$

From taking suprema it follows that

$$\|f \circ Y\|_{\mathcal{V}^{\frac{p}{\gamma}}_\gamma} \leq K \|Y\|_{\mathcal{V}^p}^\gamma$$

As required. □

So if $X \in \mathcal{V}^p$, we need that f is γ -Hölder continuous for $\gamma \in (0, 1]$ such that $1/p + \gamma/p > 1$. This amounts to $\gamma > p - 1$. We note that this also implies that $p < 2$. Hence, we have now found the class of suitable functions f for which our ode makes sense. From now on we will define $\text{Lip}(\gamma)$ the set of γ -Hölder continuous functions.

Lemma 34. *Let $(Y(n))_{n \geq 0}$ be a sequence that converges to Y in the p -variation norm. If $f \in \text{Lip}(\gamma)$, then the sequence $(f(Y(n)))_{n \geq 0}$ converges to $f(Y)$ in the $\frac{p'}{\gamma}$ -variation norm for every $p' > p$.*

Proof. From the previous lemma it follows that the sequence $(f(Y(n)))_{n \geq 0}$ is bounded in $\frac{p'}{\gamma}$ -variation. From the definition of convergence in p -variation, it follows that it also converges uniformly. From Theorem 12 it follows that for every $p' > p$,

$$\begin{aligned} \|f(Y) - f(Y(n))\|_{\mathcal{V}^{\frac{p'}{\gamma}}} &\leq \left(\sup_{u \in J} |f(Y_u) - f(Y_u(n))| \right)^{1 - \frac{p\gamma}{p'}} \|f(Y) - f(Y(n))\|_{\frac{p'}{\gamma}, J}^{\frac{p\gamma}{p'}} \\ &\quad + \sup_{u \in J} |f(Y_u) - f(Y_u(n))| \end{aligned}$$

Since $f(Y(n))$ converges to $f(Y)$ uniformly, we have that $\lim_{n \rightarrow \infty} f(Y(n)) = f(Y)$ in the $\frac{p'}{\gamma}$ -variation norm. \square

5.1 Existence

We first state and prove a theorem for existence. The proof is a relative straight forward application of the Schauder fixed point theorem.

Theorem 35 (Existence). *Assume that W is finite-dimensional, let p and γ be such that $1 \leq p < 2$ and $p - 1 < \gamma < 1$. Assume that $X \in \mathcal{V}^p$ and that $f \in \text{Lip}(\gamma)$. Then, for all $\xi \in W$, the ode (1) has a solution.*

Proof. Let $t \in [0, T] = J$ and $\xi \in W$. Take $p'' > p' > p$ such that $\gamma > p'' - 1$. We define the functional $F : \mathcal{V}^{p'}([0, t], W) \rightarrow \mathcal{V}^{p'}([0, t], W)$ by

$$F(Y_\bullet) = \xi + \int_0^\bullet f(Y_s) dX_s \quad (10)$$

First we will prove that this functional is continuous, using the sequential definition. Let $(Y(n))_{n \geq 0}$ be a sequence that converges to Y in p -variation norm. By the previous lemma, we have that $f(Y(n))$ converges to $f(Y)$ in the $\frac{p''}{\gamma}$ -variation norm. By Theorem 12, $F(Y(n))$ converges to $F(Y)$ in the

p -variation norm, which implies that it also does in p' -variation norm, since $p' > p$. This proves that F is continuous.

We have proved that $\mathcal{V}^{p'}([0, t], W)$ is a Banach space in Proposition 8. We will need this fact to apply our fixed point theorem. Let $M = \max(1, 2|\xi|)$, assume that $\|Y\|_{\mathcal{V}^{p'}} \leq M$ and let K be the least Hölder constant of f . Choose t such that $2c_{p, \frac{p'}{\gamma}} K \|X\|_{p, [0, t]} \leq \frac{1}{2}$. It follows now that

$$\begin{aligned} \|F(Y)\|_{\mathcal{V}^{p'}} &\leq |\xi| + \left\| \int_0^\bullet f(Y_s) dX_s \right\|_{\mathcal{V}^{p'}} \\ &\leq |\xi| + 2c_{p, \frac{p'}{\gamma}} K \|Y\|_{\mathcal{V}^{p'}}^\gamma \|X\|_{p, [0, t]} \\ &\leq M \left(\frac{1}{2} + 2c_{p, \frac{p'}{\gamma}} K \|X\|_{p, [0, t]} \right) \\ &\leq M \end{aligned}$$

This shows that if we choose t suitably, then $F(B) \subset B$, where $B = \{Y \in \mathcal{V}^{p'}([0, t], W) : \|Y\|_{\mathcal{V}^{p'}} \leq M\}$. Now we only need to show that $F(B)$ is relatively compact.

To do so, assume that ω is a control such that the p -variation of X is controlled by ω . It follows that the elements of $F(B)$ have p -variation uniformly controlled by $C\omega$ for some $C > 0$. From Corollary 13, it then follows that $F(B)$ is relatively compact in $\mathcal{V}^{p'}([0, t], W)$.

From Schauder's fixed point theorem, it now follows that F has a fixed point in B . This fixed point will be a solution on the interval $[0, t]$. For the solution on $[0, T]$, we can just subdivide the interval into sufficiently small subintervals where the previous argument applies. Stitching these local solution together yields the full solution on $[0, T]$. \square

Thus we now know when there is a solution to our ODE. In the next section we will answer obvious follow up question: is it unique? This is significantly more difficult to answer, as we will see in the next section. It should be clear that this would require more regularity of f . How much regularity exactly will be expanded on in the next section.

5.2 Uniqueness

Before we can state and prove our extension to the Picard-Lindelöf theorem on existence and uniqueness, we first have to develop some theory regarding Lipschitz functions on a Banach space V . We will assume that V is finite dimensional. Most of the theory can be extended to infinite dimensional spaces, but we will not consider this.

The aim for the first part of this section is to better understand the function $Y \mapsto f(Y)$. In the previous section, Proposition 33 and Lemma 34 gave us some information on this mapping. We will now proceed to study it more carefully.

Before we consider the general case, we shall explain the idea first on polynomials. Let $P : V \rightarrow \mathbb{R}$ be a polynomial of degree $k \geq 0$. Since V is finite dimensional, we can write $P(v) = C + av^T + v^T Av + \dots$ for $v \in V$. Let $X : [0, T]$ be a Lipschitz (or equivalently, a 1-Hölder) continuous path. We denote by $P^1 : V \rightarrow L(V, \mathbb{R})$ the derivative of P . It then follows from Taylor's theorem that

$$P(X_t) = P(X_0) + \int_0^t P^1(X_s) dX_s \quad (11)$$

Let $P^2 : V \rightarrow L(V \otimes V, \mathbb{R})$ be the second derivative of P , where $L(V \otimes V, \mathbb{R})$ is the space of bilinear forms on V . Similarly by Taylor's theorem, we have

$$P^1(X_t) = P^1(X_0) + \int_0^t P^2(X_s) dX_s \quad (12)$$

Substituting this into (11) yields

$$P(X_t) = P(X_0) + P^1(X_0) \int_{0 < u_1 < t} dX_{u_1} + \iint_{0 < u_1 < u_2 < t} P^2(X_{u_1}) dX_{u_1} \otimes dX_{u_2} \quad (13)$$

If we keep doing this and use the fact that $P^{k+1} : V \rightarrow L(V^{\otimes k+1}, \mathbb{R})$ is zero everywhere, just like even higher order derivatives (as P is a polynomial), we end up with the following expression for P :

$$\begin{aligned} P(X_t) = & P(X_0) + P^1(X_0) \int_{0 < u_1 < t} dX_{u_1} + \iint_{0 < u_1 < u_2 < t} P^2(X_{u_1}) dX_{u_1} \otimes dX_{u_2} + \dots \\ & + P^k(X_0) \int \dots \int_{0 < u_1 < \dots < u_k < t} dX_{u_1} \otimes \dots \otimes dX_{u_k} \end{aligned} \quad (14)$$

Before we continue, we clarify on some notation. By $L(V^{\otimes j}, \mathbb{R})$ we mean the space of symmetric j -linear forms on V . This means that for $f \in L(V^{\otimes j}, \mathbb{R})$, $f(v_1, \dots, v_n) = f(v_{\sigma(1)}, \dots, v_{\sigma(n)})$ for every $\sigma \in S_n$, where S_n is the group of permutations on $\{1, \dots, n\}$. The fact that it is symmetric follows from Young's theorem on the symmetry of partial derivatives. Since it is symmetric, it follows that we have the following equality:

$$\frac{1}{j!} \sum_{\sigma \in S_j} \int \dots \int_{0 < u_1 < \dots < u_k < t} dX_{u_{\sigma(1)}} \otimes \dots \otimes dX_{u_{\sigma(n)}} = \frac{1}{j!} (X_t - X_0)^{\otimes j} \quad (15)$$

Using this equation we can write our Taylor expansion about zero as:

$$P(X_t) = P(X_0) = \sum_{j=1}^k P^j(X_0) \frac{1}{j!} (X_t - X_0)^{\otimes j} \quad (16)$$

Which resembles an ordinary Taylor series, except we have tensor powers instead of normal powers.

We are now ready for the general case, we will use the ideas from the polynomial case we just saw.

Definition 36. Let V and W be two Banach spaces, let $k \geq 0$ be an integer and let $\gamma \in (k, k+1]$ be a real number. Let $F \subset V$ be closed and let $f : F \rightarrow W$ be a function. For each integer $j = 1, \dots, k$, let $f^j : F \rightarrow L(V^{\otimes j}, W)$ be a function which takes its values in the space of symmetric j -linear mappings from V to W . We say that the collection $(f = f^0, f^1, \dots, f^k)$ is an element of $\text{Lip}(\gamma, F)$ if for all $j = 0, \dots, k$ it holds that f^j is uniformly bounded on F , i.e.,

$$\sup_{x \in F} |f^j(x)| \leq M$$

and there exists a function $R_j : V \times V \rightarrow L(V^{\otimes j}, W)$ such that for all $x, y \in F$ and every $v \in V^{\otimes j}$,

$$f^j(y)(v) = \sum_{l=0}^{k-j} \frac{1}{l!} f^{j+l}(x)(v \otimes (y-x)^{\otimes l}) + R_j(x, y)(v)$$

and furthermore that

$$|R_j(x, y)| \leq M|x-y|^{\gamma-j}$$

If there is no confusion possible, we will say $f \in \text{Lip}(\gamma, F)$ instead of $(f, f^1, \dots, f^k) \in \text{Lip}(\gamma, F)$. The smallest constant M such that all the inequalities hold for all j is called the $\text{Lip}(\gamma, F)$ -norm and is denoted by $\|f\|_{\text{Lip}(\gamma)}$.

Two remarks are in order: First of all, we have not specified the underlying norm. In the case we consider this does not matter, as all norms on finite dimensional are equivalent (i.e. induce the same topology).

Secondly, since the functions f^1, \dots, f^k have as target space the space of symmetric multi-linear functions from V to W by definition, we have for

all Lipschitz continuous paths $X : [0, T] \rightarrow F$, for all $0 \leq s < t \leq T$, each $j = 1, \dots, k$ and each $v \in V^{\otimes j}$ the following equality holds

$$f^j(X_t)(v) = \sum_{l=0}^{k-j} f^{j+l}(X_s) \left(v \otimes \int_{s < u_1 < \dots < u_k < t} \dots \int dX_{u_1} \otimes \dots \otimes dX_{u_l} \right) + R_j(X_s, X_t)(v)$$

The closed set F may be quite weird, for example its interior may be empty. Let $(f, f^1, \dots, f^k) \in \text{Lip}(\gamma, F)$. The functions f^1, \dots, f^k may not be uniquely determined by f . On the interior of F , they are the classical derivatives, but as stated, this interior may be empty. As can be expected from the discussion on polynomials above, the functions f^1, \dots, f^k are the polynomial approximations of f at increasing orders, so just like an ordinary Taylor expansion. If, for example F is contained in a hyperplane, then the functions f^1, \dots, f^k are not determined by f in the directions transverse to this hyperplane.

Theorem 37 (Whitney). *Let V be finite dimensional and let F be a closed subset of V and let $0 < \gamma \leq 1$. Let $f \in \text{Lip}(\gamma, F)$. Then there exists a continuous linear extension operator that extends f from F to V continuously. Furthermore, the operator norm of this operator is independent of F .*

Proof. Since the proof is rather lengthy, we refer the reader to [38] theorem 3 in section VI.2. \square

Unfortunately, this result cannot be extended to the infinite dimensional case.

From this point we will only consider functions f which are elements of $\text{Lip}(\gamma, V)$ for some γ . We will now continue to study the map $Y \mapsto f(Y)$, for which the next proposition is essential.

Proposition 38. *Let $\gamma > 1$. Let $f : V \rightarrow W$ be an element of $\text{Lip}(\gamma)$. Then there exists a function $g : V \times V \rightarrow L(V, W)$ which is $\text{Lip}(\gamma - 1)$ and such that, for all $x, y \in V$, one has*

$$f(x) - f(y) = g(x, y)(x - y)$$

Furthermore, there exists a constant $C = C_{\gamma, V}$ such that

$$\|g\|_{\text{Lip}(\gamma-1)} \leq C \|f\|_{\text{Lip}(\gamma)}$$

We note that this closely resembles the standard mean value theorem from analysis.

Proof. Let k be the integer such that $k < \gamma \leq k + 1$. Let f^1, \dots, f^k be the first k differentials of f and let R_0, \dots, R_k be the error terms in the Taylor expansions of f, f^1, \dots, f^k . Then, define the function $g : V \times V \rightarrow L(V, W)$ by

$$g^0(x, y)(v) = g(x, y) = \int_0^1 f^1(tx + (1 - t)y) dt$$

Similarly, for all $j = 0, \dots, k - 1$ and $u, v \in V$, define

$$g^j(x, y)(u, v)^{\otimes j} = \int_0^1 f^{j+1}(tx + (1 - t)y)(tu + (1 - t)v)^{\otimes j} dt \in L(V, W)$$

And lastly, for all $x', y', u, v \in V$, define

$$\begin{aligned} S_j((x, y), (x', y'))(u, v)^{\otimes j} &= \int_0^1 R_{j+1}((tx + (1 - t)y), (tx' + (1 - t)y')) \\ &\quad \times (tu + (1 - t)v)^{\otimes j} dt \end{aligned}$$

Then (g^0, \dots, g^{k-1}) is a $\text{Lip}(\gamma - 1)$ function on $V \times V$ with corresponding error terms S_0, \dots, S_{k-1} .

Furthermore, we have using the substitution $u = tx + (1 - t)y$, $du = (x - y)dt$, that

$$\begin{aligned} g(x, y) &= \int_0^1 f^1(tx + (1 - t)y) dt \\ &= \frac{1}{x - y} \int_y^x f^1(u) du \\ &= \frac{f(x) - f(y)}{x - y} \end{aligned}$$

And hence it follows that $f(x) - f(y) = g(x, y)(x - y)$.

Lastly, let C_V be an upper bound for the Lipschitz norms of the Lipschitz continuous mappings $(x, y) \mapsto (tx + (1 - t)y)$ from $V \times V$ to V with $t \in [0, 1]$, then

$$\|g\|_{\text{Lip}(\gamma-1)} \leq \max_{j=1, \dots, l} C_V^{\gamma-j} \|f\|_{\text{Lip}(\gamma)}$$

And this concludes the proof of the proposition. \square

We will now present a last preliminary proposition before we are able to prove our main result, the existence of a solution. We have nearly completed our study on the mapping $Y \mapsto f(Y)$ and now we will state and prove our last regularity result.

Proposition 39. *Let W and U be two Banach spaces. Assume that $f : W \rightarrow U$ is a $\text{Lip}(1+\alpha)$ function for some $\alpha \in (0, 1]$. Let $p \geq 1$ be given. For every $K > 0$ there exists a constant $C_{\alpha,p,K} > 0$ such that if $X, Y \in \mathcal{V}^p(J, W)$ with $\|X\|_{\mathcal{V}^p} \leq K$ and $\|Y\|_{\mathcal{V}^p} \leq K$, we have the following inequality*

$$\|f(X) - f(Y)\|_{\mathcal{V}^{\frac{p}{\alpha}}} \leq C_{\alpha,p,K} \|f\|_{\text{Lip}(1+\alpha)} \|X - Y\|_{\mathcal{V}^p}$$

Proof. Using the previous proposition, construct a function $g : W \times W \rightarrow L(W, U)$ such that it is $\text{Lip}(\alpha)$ and that for all $x, y \in W$, $f(x) - f(y) = g(x, y)(x - y)$. Pick $s, t \in J$, then,

$$\begin{aligned} & |(f(X_t) - f(Y_t)) - (f(X_s) - f(Y_s))|_{\alpha}^{\frac{p}{\alpha}} \\ &= |g(X_t, Y_t)(X_t - Y_t) - g(X_s, Y_s)(X_s - Y_s)|_{\alpha}^{\frac{p}{\alpha}} \\ &= |g(X_t, Y_t)((X_t - Y_t) - (X_s - Y_s)) \\ &\quad + (g(X_t, Y_t) - g(X_s, Y_s))(X_s - Y_s)|_{\alpha}^{\frac{p}{\alpha}} \\ &\leq 2^{\frac{p}{\alpha}-1} |g(X_t, Y_t)|_{\alpha}^{\frac{p}{\alpha}} |(X_t - Y_t) - (X_s - Y_s)|_{\alpha}^{\frac{p}{\alpha}} \\ &\quad + 2^{\frac{p}{\alpha}-1} \|g\|_{\text{Lip}(\alpha)}^{\frac{p}{\alpha}} |(X_t, Y_t) - (X_s, Y_s)|^p |X_s - Y_s|_{\alpha}^{\frac{p}{\alpha}} \end{aligned}$$

Since all norms on finite dimensional spaces are equivalent, we can find $C > 0$ such that for all $x, y, x', y' \in W$, $|(x, y) - (x', y')| \leq C(|x - x'| + |y - y'|)$. Using this and taking the supremum in the previous inequality yields

$$\begin{aligned} \|f(X) - f(Y)\|_{\frac{p}{\alpha}, J}^{\frac{p}{\alpha}} &\leq 2^{\frac{p}{\alpha}-1} \sup_{t \in J} |g(X_t, Y_t)|_{\alpha}^{\frac{p}{\alpha}} \|X - Y\|_{\frac{p}{\alpha}, J}^{\frac{p}{\alpha}} \\ &\quad + 2^{\frac{p}{\alpha}-1} \|g\|_{\text{Lip}(\alpha)}^{\frac{p}{\alpha}} C^p 2^{p-1} (\|X\|_{p, J}^p + \|Y\|_{p, J}^p) \|X - Y\|_{\infty}^{\frac{p}{\alpha}} \end{aligned}$$

By collecting terms, applying the inequality in the previous proposition and using the definition of a $\text{Lip}(1 + \alpha)$ function, the result follows. \square

In essence, we have proved in the last proposition that $Y \mapsto f(Y)$ is Lipschitz continuous under some conditions on f . Using our knowledge from the theory of standard ordinary differential equations, we expect that we will use this fact to show that it is a contraction mapping.

Theorem 40 (Uniqueness). *Let p and γ be such that $1 \leq p < 2$ and $p < \gamma$. Assume that $X \in \mathcal{V}^p$ and that $f \in \text{Lip}(\gamma)$. Then, for every $\xi \in W$, the differential equation (1) has a unique solution.*

Let $Y = I_f(X, \xi)$ denote the unique solution to (1) starting at ξ , then the Itô mapping $I_f : \mathcal{V}^p(J, V) \times W \rightarrow \mathcal{V}^p(J, W)$ is continuous.

Just as in the classical case, we require exactly one more degree of regularity on f for uniqueness compared to existence.

Proof. Let $\alpha \in (0, 1]$ be such that $\gamma \geq 1 + \alpha > p$, so that $f \in \text{Lip}(1 + \alpha)$. We note that $p < 1 + \alpha$ if and only if $1/p + \alpha/p > 1$. Let $\xi \in W$, $t \in J = [0, T]$ and just like in the proof of existence, consider the functional $F : \mathcal{V}^p([0, t], W) \rightarrow \mathcal{V}^p([0, t], W)$, which is again defined by

$$F(Y_\bullet) = \xi + \int_0^\bullet f(Y_s) ds$$

We will prove that for suitable t , this functional is a contraction. From which the result will follow. Let $M = 2|\xi|$ and let $Y \in \mathcal{V}^p([0, t], W)$ be such that $\|Y\|_{\mathcal{V}^p} \leq M$. Since we have that $f \in \text{Lip}(\gamma) \subset \text{Lip}(1)$, we have that $f(Y)$ has finite p -variation and it is controlled by $\|f\|_{\text{Lip}(1)}$ times that of Y . Hence, by Theorem 31 and Corollary 32,

$$\begin{aligned} \|F(Y)\|_{\mathcal{V}^p([0, t], W)} &\leq |\xi| + 2 \left\| \int_0^\bullet f(Y_s) dX_s \right\|_{p, [0, t]} \\ &\leq |\xi| + 2C_{p,p} \|f\|_{\text{Lip}(1)} \|Y\|_{p, [0, t]} \|X\|_{p, [0, t]} \\ &\leq M \left(\frac{1}{2} + C_{p,p} \|f\|_{\text{Lip}(1)} \|X\|_{p, [0, t]} \right) \end{aligned}$$

Furthermore, let $Y' \in \mathcal{V}^p([0, t], W)$ such that $\|Y'\|_{\mathcal{V}^p} \leq M$. Then, by Proposition 39,

$$\|f(Y) - f(Y')\|_{\mathcal{V}^{\frac{p}{\alpha}}} \leq C_{\alpha,p,M} \|f\|_{\text{Lip}(1+\alpha)} \|Y - Y'\|_{\mathcal{V}^p}$$

Hence it follows from Corollary 32,

$$\|F(Y) - F(Y')\|_{\mathcal{V}^p} \leq C_{p, \frac{p}{\alpha}} C_{\alpha,p,M} \|f\|_{\text{Lip}(1+\alpha)} \|Y - Y'\|_{\mathcal{V}^p} \|X\|_{p, [0, t]}$$

Now, choose t such that the following inequality holds

$$\|X\|_{p, [0, t]} < C_{p, \alpha, M, f} = (2C_{p,p} \|f\|_{\text{Lip}(1)})^{-1} \wedge \left(2C_{p, \frac{p}{\alpha}} C_{\alpha,p,M} \|f\|_{\text{Lip}(1+\alpha)} \right)^{-1}$$

Using the notation $a \wedge b = \min(a, b)$. If t is chosen properly, then we have on the ball $B_M = \{\|\cdot\|_{\mathcal{V}^p([0,t],W)} \leq M\}$

$$\begin{aligned}\|F(Y)\|_{\mathcal{V}^p([0,t],W)} &\leq M \\ \|F(Y) - F(Y')\|_{\mathcal{V}^p} &< \|Y - Y'\|_{\mathcal{V}^p}\end{aligned}$$

This implies that $F(B_M) \subset B_M$ and that it is a contraction mapping on this ball. From Banach's fixed point theorem, it then follows that our functional has a unique fixed point, which is our unique solution to (1) on the interval $[0, t]$.

For global existence and uniqueness on $J = [0, T]$ we can apply the usual argument, by subdividing it on suitable small intervals such that the previous argument applies. Finally stitching together these local solutions yields the final solution.

Now we check that the Itô mapping is continuous. Let $Y = I_f(X, \xi)$ denote the unique solution starting at ξ . Choose $X \in \mathcal{V}^p$, $M > 0$ and t such that (5.2) holds. Then it follows from (5.2) that if $|\xi| \leq \frac{M}{2}$ and $\|Y\|_{\mathcal{V}^p([0,t],W)} \leq M$, then for all $n \geq 1$,

$$\|F^n(Y) - I_f(X, \xi)\|_{\mathcal{V}^p([0,t],W)} \leq \frac{M}{2^{n-2}}$$

Where F^n means that we compose F with itself n times. From this inequality it follows that the sequence of continuous mappings $(\xi, X, Y) \mapsto (\xi, X, F^n(Y))$ converges uniformly to the mapping $(\xi, X, Y) \mapsto (\xi, X, I_f(X, \xi))$ on the domain $\{(\xi, X, Y) : 2|\xi| \leq M, \|X\|_{p,[0,t]} < C_{p,\alpha,M,f}, \|Y\|_{\mathcal{V}^p([0,t],W)} \leq M\}$

Since we wrote I_f as a uniformly converging sequence of continuous mappings, we conclude that the Itô functional

$$I_f : \{(\xi, X) : 2|\xi| \leq M, \|X\|_{p,[0,t]} < C_{p,\alpha,M,f}\} \rightarrow \mathcal{V}^p([0, t], W)$$

is continuous on this domain. □

5.3 Some remarks and further questions

After a long journey we finally know and have proven when we can expect a solution to a differential equation driven by an irregular signal, and when it is unique. Even though the proofs and the statements mirror the classical case quite closely, the proofs are still rather different and significantly more involved.

Although we have answered the main question we posed, there are still multiple other questions one could have. For example, we always assume that X is of bounded p -variation with $p < 2$. The reason for this restriction is simple, because we need $1/p + 1/q > 1$ for the Young integral to make sense, since its proof relies on the convergence of the Riemann-zeta function. Is there some way that we can relax this requirement? This is important, as we know that a Brownian motion only has bounded p -variation for $p > 2$.

Secondly, at this point we have no clue what a solution looks like. We know when it exists and when it is unique, but that's it. Can we find an explicit solution to (1) for some function f ? If not, can we numerically approximate a solution?

The answers to all these questions is yes. It is possible to define a differential equation driven by a path with bounded p -variation for $p \geq 2$ and infinite p -variation for $p < 2$. But it needs a completely different theory, named rough path theory. The theory is quite young and as stated in the introduction, its use has already resulted in a Fields medal for Martin Hairer.

As for explicit solutions, these are quite hard to find. There are some known closed form solutions when X_t is a fractional Brownian motion with $H > 1/2$, as we will see in the next section. In the common case that we don't have an explicit solution, we will need to numerically approximate them. We shall detail two algorithms for which convergence can be proved and use them to simulate solutions.

Finally, it should be mentioned that the path we took here, using p -variation spaces and methods, is not the only way to do this. After Lyons developed this theory, Nualart and Rascanu came to similar results using fractional calculus.

6 Computing solutions

6.1 Explicit solutions

In this section we will find an explicit solution to a class of differential equations driven by irregular signals. Let B_t^H be a fBm with Hurst parameter $H \in (\frac{1}{2}, 1)$. We follow the paper [42] for this part.

Proposition 41. *Consider the following stochastic differential equation*

$$dY_t = aY_t dB_t^H, \quad Y_0 = \xi \neq 0 \quad (17)$$

With $a \in \mathbb{R}$. Then this equation has the following unique solution

$$Y_t = \xi \exp(aB_t^H) \quad (18)$$

Before we can prove this, we need the following theorem

Theorem 42. *Let $f \in \mathcal{V}^p([0, T], \mathbb{R})$, let $F \in C^1(\mathbb{R})$ be a real valued continuously differentiable function such that $F \circ f \in \mathcal{V}^q([0, T], \mathbb{R})$ for p, q such that $1/p + 1/q > 1$, then we have for all $y \in [0, T]$,*

$$F(f(y)) - F(f(a)) = \int_a^y F'(f(t)) df(t)$$

Proof. Let \mathcal{D} be an arbitrary partition, then it follows from the mean value theorem for F and the continuity of f that

$$\begin{aligned} F(f(y)) - F(f(a)) &= \sum_{i=0}^{r-1} F(f(t_{i+1})) - F(f(t_i)) \\ &= \sum_{i=0}^{r-1} F'(f(\tilde{t}_i))(f(t_{i+1}) - f(t_i)) \end{aligned}$$

For some $\tilde{t}_i \in [t_i, t_{i+1}]$. Now if we let $|\mathcal{D}| \rightarrow 0$, the last expression converges to $\int_a^y F'(f(t)) df(t)$. \square

A generalization of this is given by, with $F \in C^1(\mathbb{R} \times (a, b))$

$$F(f(y), y) - F(f(a), a) = \int_a^y F'_1(f(x), x) df(x) + \int_a^y F'_2(f(x), x) dx \quad (19)$$

We will now prove Proposition 41.

Proof. We simply apply the Theorem 42 with $F(z) = \xi \exp(z)$ and $f(t) = aB_t^H$, from which the conclusion follows. \square

It is actually possible to solve a much bigger set of such equations, using the Doss-Sussman representation, for which we refer to [33]

6.2 Numerically approximating solutions

In this subsection we will consider the following one dimensional differential equation driven by a fractional Brownian motion B^H for $H > \frac{1}{2}$,

$$dY_t = \sigma(Y_t)dB_t^H, \quad Y_0 = \xi \quad (20)$$

Where we assume that σ is regular enough so that there exists a unique solution to this equation, which we understand as

$$Y_t = \xi + \int_0^t \sigma(Y_s)dB_s^H \quad (21)$$

We will consider two algorithms to numerically to compute this solution. Both are well known in the deterministic case, namely the explicit Euler and the Crank-Nicolson schemes. In the setting we are considering, things are a bit different as we will see. As one might expect, the convergence of these algorithms are slower in our setting than in the ordinary, less rough, deterministic case. We will provide the convergence estimates and will test them on examples.

This section is not completely mathematically rigorous. The theory necessary for proving the convergence estimates is rather involved and could warrant another bachelor's thesis to do it justice. Furthermore, the theory is rather new and expanding quickly. The paper detailing the Crank-Nicolson scheme has only been published in September 2017. So there is active research in this field and there are numerous applications in finance and physics.

In all that follows, we will only consider the one dimensional case. Extension to higher dimensions are possible and do exist, but are significantly more involved and hence omitted.

6.2.1 Euler-Maruyama scheme

We will try to numerically approximate (20) on the interval $[0, 1]$. For simplicity, we will only consider uniform partitions of this interval, $t_k = k/n$. We will consider the following iteration scheme

$$\begin{aligned} Y_0^{(n)} &= \xi \\ Y_{(k+1)/n}^{(n)} &= Y_{k/n}^n + \sigma(Y_{k/n}^n) \left(B_{(k+1)/n}^H - B_{k/n}^H \right), \quad k \in \{0, \dots, n-1\} \end{aligned}$$

It has been proven in [31] that

$$n^{2H-1} \|Y^{(n)} - Y\|_\infty \xrightarrow{\text{a.s.}} \frac{1}{2} \left(\sup_{t \in [0,1]} \left| \int_0^1 D_s Y_t ds \right| \right) \quad (22)$$

Where $D_s Y_t$ is the Malliavin derivative at time s of Y_t with respect to the fBm B^H . We will take this result for granted, but its proof can be found in the reference above. Furthermore, we refer the reader to [44] for a good introduction on the Malliavin derivative. We note that this result is just a random variable, which only depends on H .

Our convergence estimate has the implication that if $H \rightarrow \frac{1}{2}$, i.e. B^H is a standard Brownian motion, then the convergence deteriorates. Furthermore, if $H < \frac{1}{2}$, then there is no convergence. It also shows that if convergence is worse the lower the Hurst parameter gets. This makes sense, as we have seen that a fBm becomes more regular as H increases.

6.2.2 Crank-Nicholson

Once more, we will consider the interval $[0, 1]$ and a uniform partition $t_k = k/n$. Then the Crank-Nicolson scheme is given by [21]

$$\begin{aligned} Y_0^n &= \xi \\ Y_{t_{k+1}}^n &= Y_{t_k}^n + \frac{1}{2} \left(\sigma(Y_{t_{k+1}}^n) + \sigma(Y_{t_k}^n) \right) (B_{t_{k+1}}^H - B_{t_k}^H) \quad k = 0, \dots, n-1 \end{aligned}$$

Let $Y_t^n = Y_{t_k}^n + \frac{1}{2} \left(\sigma(Y_{t_{k+1}}^n) + \sigma(Y_{t_k}^n) \right) (B_t^H - B_{t_k}^H)$ be the continuous time linear interpolation of this scheme for $t \in [t_k, t_{k+1})$. We then have the following convergence result: If Y is the solution to the equation, σ is a bounded three times continuously differentiable function with $\sigma(0) = 0$, then there exists a constant $K = K_p$ independent of n such that we have

$$\sup_{t \in [0,1]} (\mathbb{E} |Y_t - Y_t^n|^p)^{1/p} \leq K n^{-2H}$$

We remark that if we formally set $H = 1$, then we recover the convergence rate of the normal (i.e. deterministic) Crank-Nicolson scheme. Also, in the case that $\sigma(0) \neq 0$, we still have convergence, but it is slower. Then it is of order $N^{-H-\frac{1}{2}}$. It can also be generalized to higher dimensional equations. This yields another deterioration of convergence, as the Levy area does not vanish in this case. For the details we refer to [21].

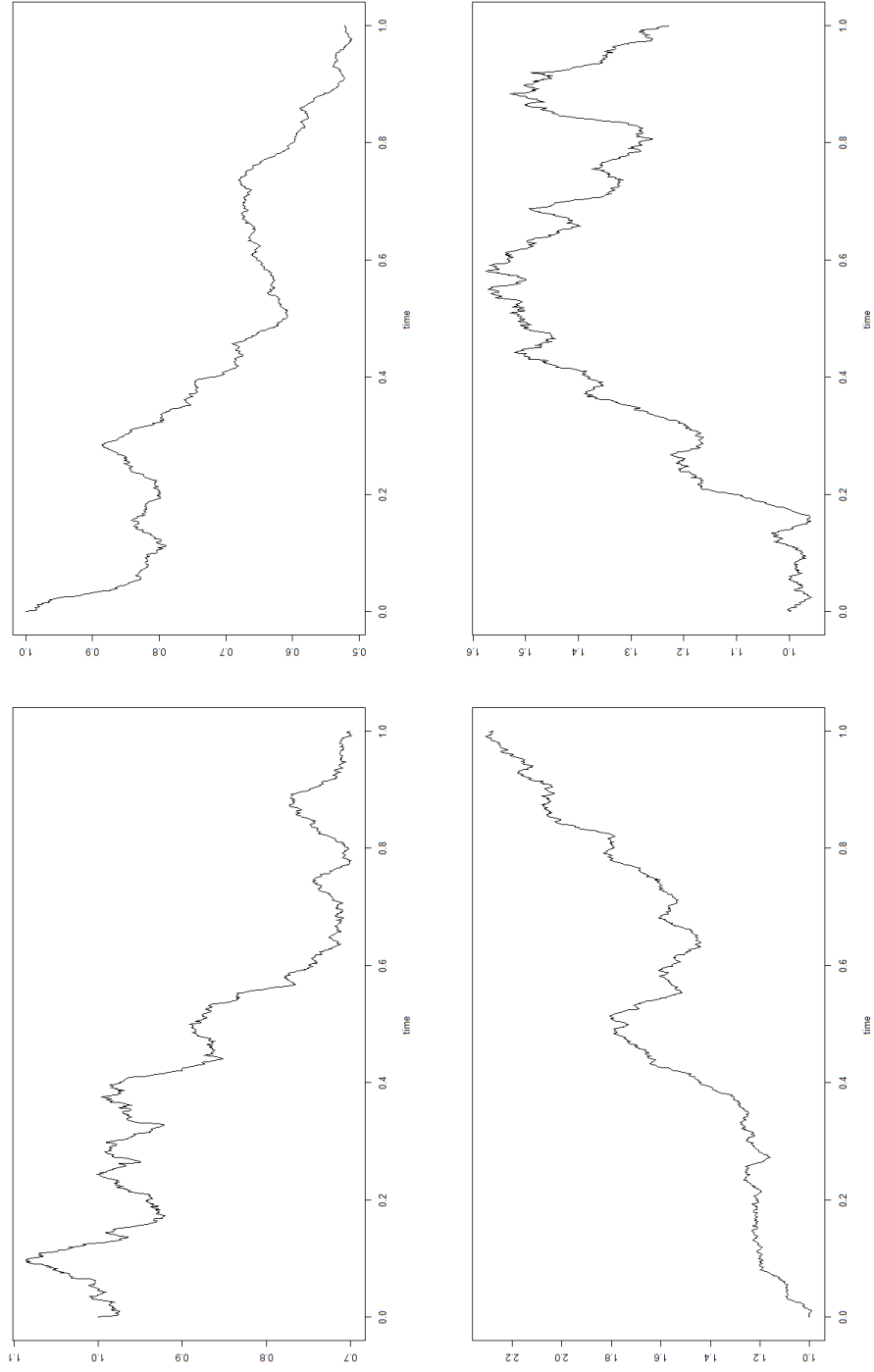


Figure 5: Four possible solutions to the differential equation $dY_t = Y_t \sin(Y_t) dB_t^H$ with $H = 0.7$

Notice that we have a different kind of convergence compared to the one for the Euler-Maruyama method, there we had so called strong convergence, here we have weak convergence. Strong convergence implies weak convergence, but as we can see, the orders of convergence may be different.

6.3 Numerical results for the Euler-Maruyama scheme

In this section we will try to empirically confirm the convergence estimate (22). In the previous section we solved a linear equation explicitly. We will use this as the equation to test the above convergence result. To make matters precise: we consider the equation $dY_t = 2Y_t dB_t^H$ with $Y_0 = 1$ on the interval $[0, 1]$. We use the R package *somebm* to generate a sample of a fractional Brownian motion. We consider uniform grids.

To determine the validity of the convergence estimate (22), we will use the following procedure for every $H \in \{0.55, 0.65, 0.75, 0.85, 0.95\}$:

1. First we generate a sample of size $n = 2^{15}$ of a fBm with a certain H on the $[0, 1]$;
2. For $dt = 2^{-15} = 1/n$, we calculated the applied the Euler-Maruyama scheme with $dt, 2dt, 4dt, \dots, 2^{11}dt$ and calculated the error;
3. We do this 1000 times and calculate the mean error;
4. Lastly, we estimate the speed of convergence using log-log regression.

When we apply this procedure, we produce the following graphs, in which the red line is the theoretical convergence result. The code can be found in appendix A. Notice that we look at a slightly weaker statement, as we just consider the error at $t = 1$.

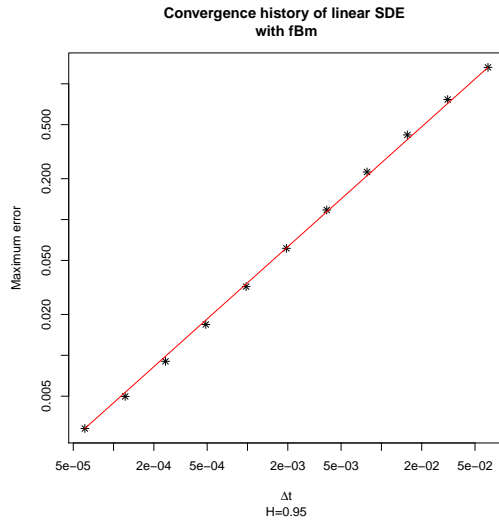
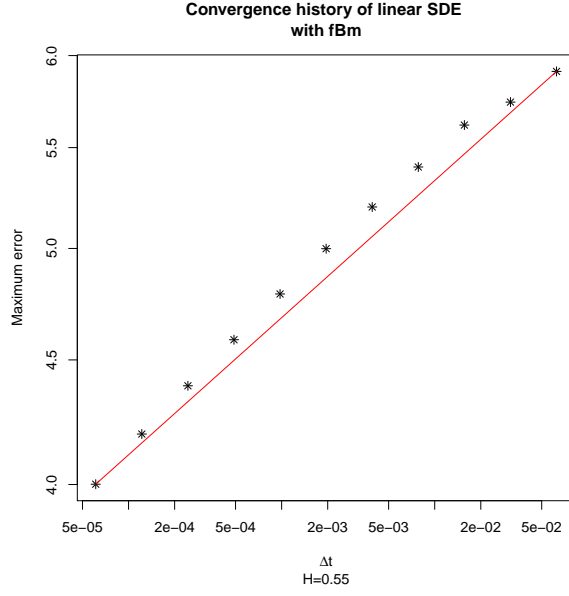
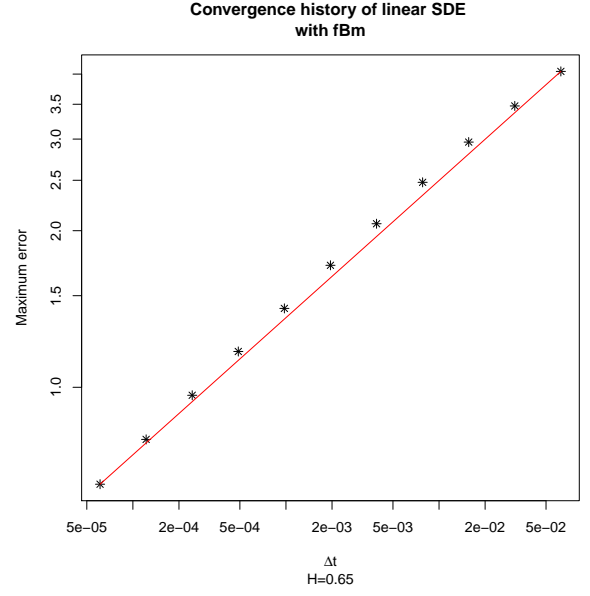


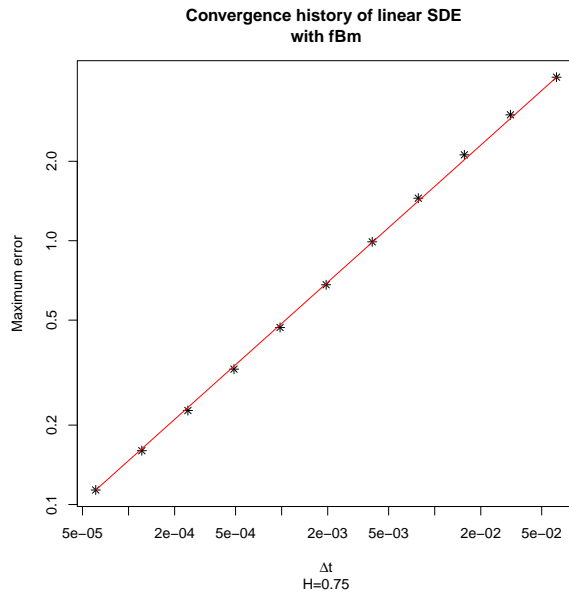
Figure 6: We estimated a convergence rate of $n^{-0.9018}$



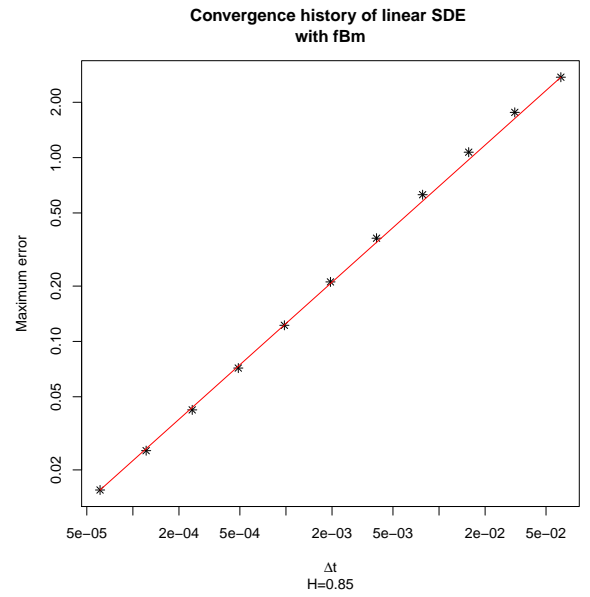
(a) We estimated a convergence rate of $n^{-0.05711}$



(b) We estimated a convergence rate of $n^{-0.2659}$



(a) We estimated a convergence rate of $n^{-0.5268}$



(b) We estimated a convergence rate of $n^{-0.7597}$

Some remarks on the results are in order: As expected, the closer we get to $H = 0.5$, the more the convergence rate deteriorates. If we look at the maximum error when $H = 0.55$, even with an uniform grid of size 2^{15} , we still have an error of about 4.0. The convergence we get is also lower than the theoretical result. I can think of two reasons for this:

1. Since our H is pretty close to 0.5, the sample paths of the fBm are quite wild. This may imply that the variance of the random variable in (22) which results from the Malliavin derivative has a very high variance and is possibly very skewed to the right. A very high number of samples may be necessary to accurately determine the empirical convergence rate. A Monte-Carlo study to this random variable may be a good option to determine its properties, but unfortunately out of the scope of this project;
2. The pseudo random number generator used in the somebm package does not have enough entropy.

The second one seems unlikely, as it would imply that we'd see the same behaviour in the other estimates. To test the first possible reason, the following strategy can be applied. Instead of calculating only one thousand samples, we try a significant greater amount of samples. Preliminary tests with 30k and 500k samples shows that the converge rate does indeed improve slightly, so we can't rule out this option yet. Unfortunately, due to the lack of sufficient computing resources, we can't further investigate the situation.

Since we see the same behaviour with $H = 0.65$, it seems likely that the previous reasoning is sound. Interestingly, we see the same behaviour but reversed with $H \in \{0.75, 0.85\}$. Now the estimated convergence rate is faster than the theoretical rate. So in this case the random variable in the convergence estimate may help speed up the convergence rate. But this is just speculation, more research has to be done. In any case, it is known that the converge estimate (22) is strict, in the sense that $n^{2H-1}|X_t^n - X_t|$ converges almost surely to a finite and nonzero limit, as was proved in [31].

Interestingly, for $H = 0.95$, the estimated convergence is as expected. This is once more evidence that more samples are necessary to make conclusions.

7 The fractional Black-Scholes model

One of the most famous models in financial mathematics is the (classical) Black-Scholes option pricing model. Developed by the economists Black, Scholes and Merton and awarded a Nobel (Memorial) price in economics in 1997, it was the starting point of a very extensive mathematical theory of option pricing.

The original Black-Scholes model is described as follows: Consider a market where there are at least two assets, one risk-less, for instance a bond of a trusted country, and a risky one, which may be a stock of a company. We want to price an European option, which is simply a contract between parties to have the right, but not an obligation, to buy or sell an asset at a predetermined price (the 'strike' price) and time. There are many kinds of options, but for this paper we will focus only on European ones.

We have some more assumptions on the market, such that no dividends are paid out, the market is efficient ², there are no transaction costs. Lastly, in the classical Black-Scholes model it is assumed that the log-returns follow a Brownian motion with drift.

Even though the model is celebrated, it is not without its flaws. Statistical analysis of market behaviour has shown [37] that there is some kind of long range dependence. It is well known that a standard Brownian motion is unable to capture such dynamics. Previously, we have calculated the Hurst parameter of some stock indices and a currency exchange rate. We saw that for the Dutch AEX we had $H \approx 0.55$, so a standard Brownian motion may be inappropriate. So recently, a more general model has been proposed, the fractional Black-Scholes model. In this model the dynamics of the market are driven by a fractional Brownian motion. For our discussion, we will restrict the Hurst parameter to $(0.5, 1)$ as usual, so we can use the theory developed previously in this thesis. Furthermore, the fBm only shows long term dependence in this parameter range.

7.1 A brief introduction to mathematical finance

Before we dive into the main subsections, we will first give a short introduction in mathematical finance for the uninitiated. We assume knowledge of the standard Black-Scholes model, information on it can be found readily on the internet.

Consider the fractional Black-Scholes market on the time interval $[0, 1]$ with a risky asset S , say a stock, and a non-risky asset B , for instance a

²See also the efficient market hypothesis

bond. We assume that the interest rate r_t is bounded by a deterministic constant. We assume that S and B solve the following equations

$$\begin{aligned} dB_t &= r_t B_t dt \\ dS_t &= \mu S_t dt + \sigma S_t dB_t^H \end{aligned}$$

or, as we have previously proved³, equivalently with $B_0 = 1$,

$$\begin{aligned} B_t &= \exp \left\{ \int_0^t r_s ds \right\} \\ S_t &= S_0 \exp \{ \mu t + \sigma B_t^H \} \end{aligned}$$

Let \mathcal{F} be the (left-continuous) filtration generated by B and S , $\mathcal{F}_t = \sigma\{B_u, S_u, u \leq t\} = \sigma\{B_u, B_u^H, u \leq t\}$.

Definition 43. *A portfolio or trading strategy is a \mathcal{F} -predictable process $\Pi = (\Pi_t)_{t \in [0,1]} = (\pi_t^B, \pi_t^S)_{t \in [0,1]}$, where π_t^B denotes the number of bonds and π_t^S denotes the number of shares that the investor has at time t . The value of this portfolio a time t is given by*

$$V_t^\Pi = \pi_t^B B_t + \pi_t^S S_t$$

We call a portfolio self-financing if

$$dV_t^\Pi = \pi_t^B dB_t + \pi_t^S dS_t$$

which states that the changes in the portfolio are only due to changes in asset prices. In other words, there is no money in or outflow, our theoretical market is rather limited.

Define the discounted value of the portfolio by

$$C_t = V_t^\Pi B_t^{-1}$$

This is essentially a way to compare the portfolio to the risk-free alternative. Then it follows that

$$dC_t = \pi_t^S S_t B_t^{-1} =: \pi_t^S X_t$$

where X_t can be interpreted as the discounted risky asset price process.

³This statement is slightly more general than the one we have proved, because of the deterministic drift term. The proofs for this case are nearly the same as for the restricted case and hence were left out.

Next we are going to define what arbitrage in a market means. Consider the following scenario: you are a stock trader that is active on two stock exchanges that use different currencies and you don't have to pay any transaction costs. Furthermore, both markets are perfectly fluid. So any selling or buying of stocks goes instantaneous. On both the stock exchanges you can buy the stock of company A. During your market research, you notice that you can buy a stock of company A for a certain price in currency x and then sell the same stock on the other stock exchange for a price in currency y. When you trade your currency y for currency x, you conclude that you have made a nice profit on these transactions.

The concept described here is known as arbitrage, which means that you have made a profit without taking any risks. As one might think, such opportunities are the holy grail for stock traders; risk-less profit. In reality such opportunities are very rare, as the markets corrects itself very quickly. So when making a mathematical model of the market, one does not want to allow the possibility of arbitrage. Before we look at the possibility of arbitrage in the fractional Black-Scholes model, we shall first define arbitrage in a mathematical sense.

Definition 44. *A self-financing portfolio Π is arbitrage if $V_0^\Pi = 0$, $V_1^\Pi \geq 0$ a.s. and $\mathbb{P}(V_1^\Pi > 0) > 0$. Moreover, if there exists a $c > 0$ such that $V_1^\Pi \geq c$ a.s. then it is called strong arbitrage.*

In other words, in markets where there exists strong arbitrage, the portfolio holder can choose the amount of profit he/she wants to make. Hence it is not a good model for a realistic market. In the next section, we will show that the fractional Black-Scholes model under the standard assumption, has strong arbitrage.

7.2 Arbitrage in the fractional Black-Scholes model

In this section we will proof that the fractional Black-Scholes model unfortunately has strong arbitrage. Instead of proving it directly, we will proof a stronger result from which the following theorem will follow:

Theorem 45. *The fractional Black-Scholes model admits strong arbitrage*

Before we proceed we need the following lemma from [30].

Lemma 46. *There exists a \mathcal{F} -adapted process $\varphi = \{\varphi_t : t \in [0, 1]\}$ such that*

1. For any $t < 1$ and $\alpha \in (1 - H, 0.5)$ $\|\varphi\|_{1,\alpha,t} < \infty$ a.s., so the integral $v_t = \int_0^t \varphi_s dB_s^H$ exists;
2. $\lim_{t \rightarrow 1^-} v_t = \infty$ a.s.

The norm $\|\cdot\|_{1,\alpha,t}$ is some kind of Hölder norm, for the definition we refer to the reference, just as for the proof of the lemma, which is rather lengthy and technical. We now state and proof our result:

Theorem 47. *For any distribution function F , there is a self-financing portfolio Π with $V_0^\Pi = 0$ such that its discounted terminal capital C_1^Π has distribution F .*

Proof. Take a nondecreasing function g such that $g(B_{\frac{1}{2}}^H)$ has distribution F , let φ be as in Lemma 46. Set $v_t = \int_{\frac{1}{2}}^t \varphi_s dX_s$. Furthermore, let

$$\tau = \min\{t \geq \frac{1}{2} : v_t = g(B_{\frac{1}{2}}^H)\}, \quad \pi_t^S = \varphi_s \mathbf{1}_{[\frac{1}{2}, \tau)}(t) \quad (23)$$

Since we have that $\lim_{t \rightarrow 1^-} v_t = \infty$ by Lemma 46, we have that $\tau < 1$ a.s. Then it is possible to construct a self-financing portfolio $\Pi = (\pi^B, \pi^S)$ with initial value zero ($V_0^\Pi = 0$). By definition, $\pi_t^S = 0$ for $t \in [0, \frac{1}{2})$, hence $C_{\frac{1}{2}}^\Pi = 0$. Moreover,

$$C_1^\Pi = C_{\frac{1}{2}}^\Pi + \int_{\frac{1}{2}}^1 \pi_s^S dX_s = \int_{\frac{1}{2}}^\tau \phi_s dX_s = g(B_{\frac{1}{2}}^H) \quad \square$$

If we let F be the distribution function of a constant $A > 0$ and observe that B_1^{-1} is greater than some deterministic constant by assumption, we can let the terminal value of the portfolio be any number that we want, in other words, we have shown the existence of strong arbitrage.

So in this setting, the fractional Black-Scholes model has no value as it is unrealistic. The literature details a couple of solutions to repair this problems. We list three of them:

1. Incorporate proportional transaction costs;
2. Allow trading only at discrete times (i.e. limit the amount of times trades can be made);
3. Add a standard Brownian motion to the model, which results in the Mixed Fractional Black-Scholes model.

These options are still actively being researched and show promising results so far.

8 Summary

In this thesis we developed the theory necessary to solve differential equations driven by irregular signals in a pathwise sense. Our focus was mostly on proving sufficient conditions for the existence and/or uniqueness of solutions. To do this, we first tried to use the classical ODE theory, by rewriting it as an integral equation and trying to solve it using fixed point techniques. In the case of a mildly irregular signal (when it is Lipschitz continuous), we could use the Riemann-Stieltjes integral. But in the class of signals we would like to consider, such relatively regular signals are seldom seen.

Since the process we considered most important in this thesis, a fractional Brownian motion, does not satisfy this property. Even in the case that $H > \frac{1}{2}$, the Riemann-Stieltjes integral is of no use to us. To be able to solve (1), we need to extend the Riemann-Stieltjes integral.

The mathematician Young provided us with an extension that proved to be fruitful for our applications. In 1936 he showed that if f is of bounded p -variation and if g is of bounded q -variation such that $1/p + 1/q > 1$, then the integral $\int f dg$ is well-defined. Unfortunately, this theory restricts us to the case $q < 2$. For our first purpose, this is good enough. But we need to ask if we can use another extension.

Using the Young integration theory, we were able to show when there are solutions and when they are unique. The conditions resemble the conditions from the classical ODE theory: for uniqueness we need exactly one degree of regularity more than for existence. But since we are integrating against irregular signals, we need more regularity on our function f . We have shown how much exactly.

Next we used the tools and knowledge developed previously to try and determine solutions of differential equations driven by irregular signals. We have shown that for linear equations, the solution very much resembles the solution of a standard linear ODE. For more complex cases, we often need to numerically approximate the solution. We have described two algorithms, both well known in the deterministic and smooth case. For the Euler-Maruyama method, we also looked at the convergence rate in practice.

Lastly, we detailed the fractional Black-Scholes model. Statistical analysis has shown that the standard Black-Scholes model driven by a Brownian motion may not be the optimal form of the model. So an extension to incorporate a fractional Brownian motion was proposed. Unfortunately, the resulting model allows for arbitrage, which renders the model useless in practice. We finished with a short discussion on how to prevent arbitrage in this setting.

A R code for estimating convergence rate of Euler-Maruyama scheme

In this section we present the R code used to compute the convergence rates of the Euler-Maruyama scheme. To speed up the computing, we have parallelised the program. All results were calculated on the DLR-SC cluster, for which I am grateful. An improvement would be to use GPGPU, since the outer loop is trivially parallelizable. There convergence history is plotted and saved to a file. In the terminal the estimated convergence speed is outputted as "convSpeed".

```
library(foreach)
library(doParallel)

N <- 2^15
Yzero <- 1
dt <- 1/N
numSamples <- 500000
numTrials <- 11
Yerr <- matrix(0, nrow=numSamples, ncol=numTrials)
Hurst <- 0.85
lambda <- 2

cores <- 55
cl <- makeCluster(cores, outfile="")
registerDoParallel(cl)

Yerr <- foreach (idx=1:numSamples, .combine='rbind') %dopar%
{
  library(somebm)
  tmpYerr <- c(numTrials)
  fBmSample <- fbm(Hurst, N)
  dB <- diff(fBmSample)
  Yexact <- Yzero * exp(lambda * fBmSample)

  for (dist in 1:numTrials)
  {
    R <- 2^(dist - 1)
    Dt <- R * dt
    L <- N / R
```

```

Ytmp <- Yzero

for (iter in 1:L)
{
  tmp <- R * (iter - 1) + 1
  tmp2 <- R * iter
  Binc <- sum(dB[tmp : tmp2])
  Ytmp <- Ytmp + lambda * Ytmp * Binc
}
tmpYerr[dist] <- abs(Ytmp- Yexact[length(Yexact)])
}
tmpYerr
}
rhs <- log(colMeans(Yerr))
DtVals <- dt * 2^(1:numTrials)
convSpeed <- log(DtVals)
lm(rhs ~ convSpeed)

stopCluster(cl)

```

References

- [1] Anonymous, "R-S Integrability of Cts. Functions with Integrators of Bounded Variation", <http://mathonline.wikidot.com/riemann-stieltjes-integrability-of-continuous-functions-with>
- [2] E. Azmoodeh, "RIEMANN-STIELTJES INTEGRALS WITH RESPECT TO FRACTIONAL BROWNIAN MOTION AND APPLICATIONS", PhD thesis, Helsinki University of Technology Institute of Mathematics (2010)
- [3] F. Baudion, "Rough Paths Theory", Unknown
- [4] F. Biagini, Y. Hu, B. Øksendal, T. Zhang, "Stochastic Calculus for Fractional Brownian Motion and Applications", Probability and Its Applications (2008)
- [5] F.E. Burke, "A Garden of Integrals", Mathematical Association of America (2007)
- [6] V.V. Chistyakov, "On Mappings of Bounded Variation", Journal of Dynamical and Control Systems (1997)
- [7] V.V. Chistyakov, O.E. Galkin, "On Maps Of Bounded p -Variation With $p > 1$ ", Positivity (1998)
- [8] L. Coutin, "An Introduction to (Stochastic) Calculus with Respect to Fractional Brownian Motion", Séminaire de Probabilités XL (2007)
- [9] P.F. Craigmile, "Simulating a Class of Stationary Gaussian Processes Using the Davies-Harte Algorithm, with Application to Long Memory Processes", NRCSE (200)
- [10] M.S. Crouse, R.G. Baraniuk, "Fast, Exact Synthesis of Gaussian and nonGaussian Long-Range-Dependent Processes", IEEE Transactions on Information Theory (1999)
- [11] J. Diehl, J. Zhang, "Backward stochastic differential equations with Young drift", Probability, Uncertainty and Quantitative Risk (2017)
- [12] A.B. Dieker, "Simulation of fractional Brownian motion", CWI (2004)
- [13] A.B. Dieker, M. Mandjes, "On spectral simulation of fractional Brownian motion", CWI (2002)

- [14] T. Duncan, D. Nualart, "Existence of strong solutions and uniqueness in law for stochastic differential equations driven by fractional Brownian motion", (2009)
- [15] B.K. Driver, "Rough Path Analysis", (2013)
- [16] D. Freedman, "Brownian Motion and Diffusion", Springer-Verlag (1983)
- [17] P.K. Friz, M. Hairer, "A Course on Rough Paths - With an introduction to regularity structures", Springer (2014)
- [18] P.K. Friz, N.B. Victoir, "Multidimensional Stochastic Processes as Rough Paths: Theory and Applications", (2009)
- [19] M. Gubinelli, "Ramification of rough paths", arXiv:math/0610300v1 [math.CA] (2006)
- [20] M. Hairer, J. Maas, H. weber, "APPROXIMATING ROUGH STOCHASTIC PDES", arXiv:1202.3094v2 [math.PR] (2013)
- [21] Y. Hu, Y. Liu, D. Nualart, "Crank-Nicolson scheme for stochastic differential equations driven by fractional Brownian motions", arXiv:1709.01614v1 [math.PR] (2017)
- [22] Y. Hu, Y. Liu, D. Nualart, "RATE OF CONVERGENCE AND ASYMPTOTIC ERROR DISTRIBUTION OF EULER APPROXIMATION SCHEMES FOR FRACTIONAL DIFFUSIONS", arXiv:1408.6471v3 [math.PR] (2016)
- [23] H. Koch, D. Tataru, M. Visan, "Dispersive Equations and Nonlinear Waves", Birkhäuser (2014)
- [24] A. Lejay. "A short introduction to rough paths: outline and selected bibliography", Workshop on Long Range Dependence: from Fractional Calculus to Financial Applications (2009)
- [25] A. Lejay, "Controlled differential equations as Young integrals: a simple approach", Journal of Differential Equations (2010)
- [26] J.A. Léon, D. Nualart, S. Tindel, "YOUNG DIFFERENTIAL EQUATIONS WITH POWER TYPE NONLINEARITIES", arXiv:1606.02258v1 [math.PR] (2016)

- [27] T.J. Lyons, "DIFFERENTIAL EQUATIONS DRIVEN BY ROUGH SIGNALS (I): AN EXTENSION OF AN INEQUALITY OF L. C. YOUNG", Mathematical Research Letters (1994)
- [28] T.J. Lyons, M. Caruana, T. Lévy, "Differential Equations Driven by Rough Paths", Springer (2004)
- [29] Y.S. Mishura, "Stochastic Calculus for Fractional Brownian Motion and Related Processes", Springer-Verlag (2008)
- [30] Y. Mishura, G. Shevchenko and E. Valkeila, "Random variables as pathwise integrals with respect to fractional Brownian motion", arXiv:1111.1851 [math.PR] (2012)
- [31] A. Neuenkirch, I. Nourdin, "Exact rate of convergence of some approximation schemes associated to SDEs driven by a fractional Brownian motion", arXiv:math/0601038v3 [math.PR] (2006)
- [32] I. Nourdin, "Schémas d'approximation associés à une équation différentielle dirigée par une fonction höldérienne; cas du mouvement brownien fractionnaire", Probabilités
- [33] I. Nourdin, "A simple theory for the study of SDEs driven by a fractional Brownian motion, in dimension one." Séminaire de Probabilités XLI. Lecture Notes in Mathematics (Séminaire de Probabilités), Springer, Berlin, Heidelberg
- [34] D. Revuz, M. Yor, "Continuous martingales and Brownian motion.", Springer (1999)
- [35] F. Russo, "Stochastic Differential Equations", Encyclopedia of Mathematical Physics (2005)
- [36] G. Shevchenko, "FRACTIONAL BROWNIAN MOTION IN A NUT-SHELL", arXiv:1406.1956v1 [math.PR] (2014)
- [37] F. Shokrollahi, "Fractional Delta Hedging Strategy For Pricing Currency Options With Transaction Costs", arXiv:1702.00037 [q.fin.PR] (2017)
- [38] E. Stein, "Singular Integrals and Differentiability Properties of Functions", Princeton University Press (1971)

- [39] H. Sussmann, "On the gap between deterministic and stochastic ordinary differential equations", The Annals of Probability, Vol. 6, No. 1 (1978)
- [40] T. Tao, "245C, Notes 1: Interpolation of L^p spaces", <https://terrytao.wordpress.com/2009/03/30/245c-notes-1-interpolation-of-lp-spaces/> (2009)
- [41] B. Williams, "Probability with Martingales", Cambridge University Press (1991)
- [42] M. Zähle, "Integration with respect to fractal functions and stochastic calculus. I", Probability Theory and Related Fields (1998)
- [43] M. Zähle, "Integration with respect to fractal functions and stochastic calculus. II", Unknown
- [44] H. Zhang, "THE MALLIAVIN CALCULUS", BSc Thesis (2004)