Technische Universität München

Department of Mathematics

Analysing the relation of expected signatures to laws of stochastic processes

Master Thesis

by

Anton Hancharyk

Supervisor:	Prof. Dr. Blanka Horvath
Advisor:	Yannick Limmer
Submission Date:	Date Date

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Munich, DATE

Abstract

About 200 words and summarizing the thesis.

Contents

Chapter 1

Introduction

Here some introductory work.

Chapter 2

Basic theory

In this chapter, we cover some of the definitions, examples and theorems, in other words, key concepts to describe signatures and their properties. This chapter is mainly referring to Chevyrev and Kormilitzin [9].

2.1 Path

In this section, we define some of the basic elements for a theory and provide some examples. We will begin with the definition of a path, as this is the primary definition for our further discussion.

Definition 1. Let a and b are real numbers, such that the interval $[a, b] \in \mathbb{R}$. Define a *path* X to $\mathbb{R}^d$ as a continuous mapping from interval $[a, b]$ to $\mathbb{R}^d$, which can be written as $X : [a, b] \to \mathbb{R}^d$.

Note that we will use the short notation $X_t = X(t)$ to indicate parameter dependency on $t \in [a, b]$.

We will assume that all paths are piecewise differentiable, such a path that is continuous and differentiable on each subinterval of the existing partition of the interval $[a, b]$. Moreover, a path which has derivatives of all orders we will be considered as smooth. Let's now show some examples with graphs to give the reader an idea of these types of paths we are going to examine.

Example 1. So let's begin with two examples of smooth paths in $\mathbb{R}^2$ in Fig.2.1:

$$\begin{aligned} (a) &: X_t = \{X_t^1, X_t^2\} = \{t, \sin t\},\ t \in [-\pi, \pi] \\ (b) &: X_t = \{X_t^1, X_t^2\} = \{t, \exp t\},\ t \in [0, 3] \end{aligned} \tag{2.1}$$

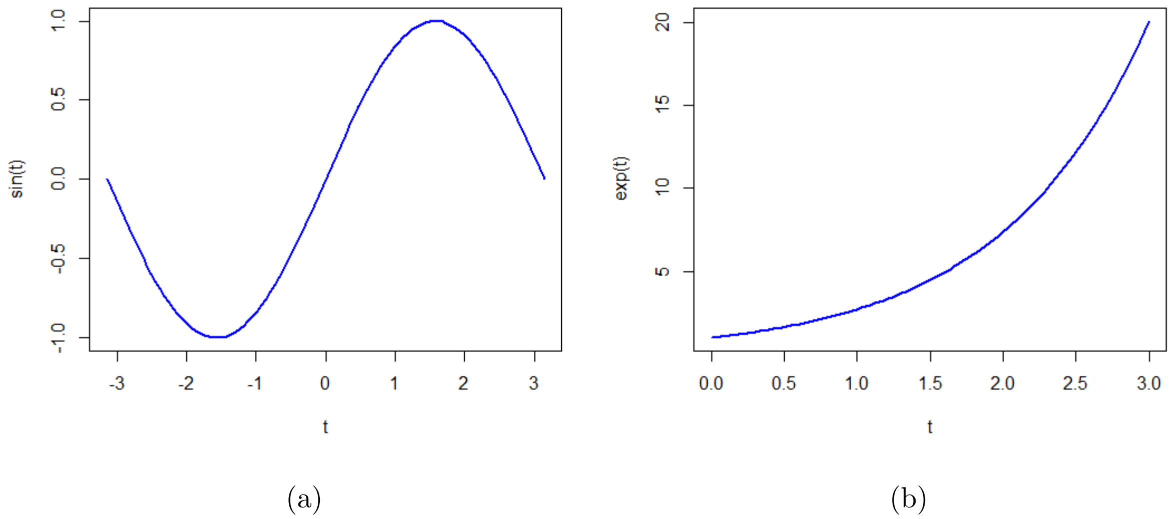


Figure 2.1: Examples of smooth paths in R^2

We will use the same parametrization for d-dimensions case $\mathbb{R}^d$, namely

$$X : [a, b] \to \mathbb{R}^d, X_t = \{X_t^1, X_t^2, ..., X_t^d\}. \tag{2.2}$$

Example 2. In figure 2.2 we show an example of a piecewise linear path. We have taken a stock prices at time t as an example of a function $f(t)$:

$$X_t = \{X_t^1, X_t^2\} = \{t, f(t)\},\ t \in [0, 1000] \tag{2.3}$$

Let us now define the *path integral*, which is also called the line integral, i.e. the integral where the integrable function is taken along the path. In particular, the line integral of one-dimensional path $X : [a, b] \to \mathbb{R}$ against function $f : \mathbb{R} \to \mathbb{R}$ is determined as follows

$$\int_a^b f(X_t)\, dX_t = \int_a^b f(X_t)\dot{X}_t\, dt, \tag{2.4}$$

where the right-hand side integral is the Riemann integral of continuous bounded functions. By $\dot{X}_t$ we denoted the derivative with respect to the time variable t, namely $\dot{X}_t = dX_t/dt$. Here we want to point out that the integrand $f(X_t)$ is a real-valued path which is defined on the interval $[a, b]$.

Generally, any path $Y : [a, b] \to \mathbb{R}$ can be integrated over a path $X : [a, b] \to \mathbb{R}$, by defining $f(X_t) = Y_t$ and writing

$$\int_a^b Y_t\, dX_t = \int_a^b Y_t\dot{X}_t\, dt. \tag{2.5}$$

Let us now consider two examples of path integrals in one- and two-dimensional cases.

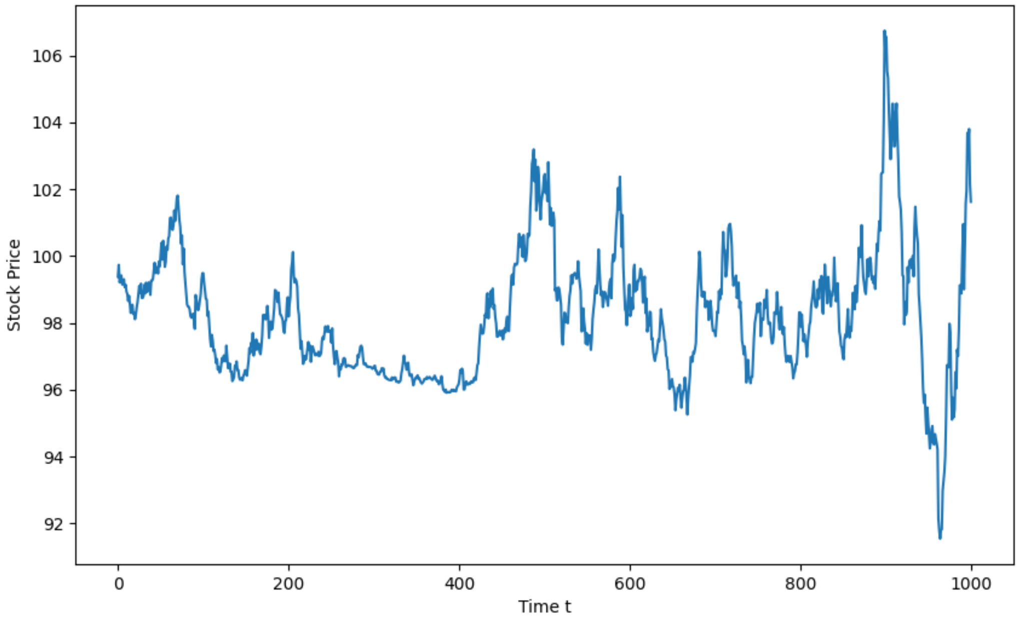


Figure 2.2: Example of piecewise linear path

Example 3. Assuming $Y_t = 1$ is a constant path for all $t \in [a, b]$. Then taking the path integral of Y against a path $X : [a, b] \to \mathbb{R}$ we obtain the increment of the path X:

$$\int_a^b dX_t = \int_a^b \dot{X}_t \, dt = X_b - X_a \tag{2.6}$$

Example 4. Consider the two-dimensional path as simple example of an iterated integral, a term that is crucial to define the path signature. The definition of which we will provide in the next section.

$$X_t = \{X_t^1, X_t^2\} = \{e^t, t^2\}, \; t \in [0, 1] \tag{2.7}$$

The numerical calculation of the path integral is obtained as follows

$$\int_0^1 X_t^1 \, dX_t^2 = \int_0^1 e^t 2t \, dt = 2 \tag{2.8}$$

2.2 The signature of a path

Before defining a signature, let us define what an *iterated integral* is. Here, for d-dimensional paths we will use the notations from the previous section, where each component $X^i : [a, b] \to \mathbb{R}$ is a real-valued path. Let us define the increment of i-th coordinate of the path at time $t \in \mathbb{R}$ by following expression:

$$S(X)^i_{a,t} = \int_{a<s<t} dX^i_s = X^i_t - X^i_a, \tag{2.9}$$

where $t \in [a,b]$, a is a starting point and $i \in \{1, ..., d\}$. Note that the mapping $S(X)^i_{a,b} : [a,b] \to \mathbb{R}$ is a real-valued path, since the components X^i are real-valued paths.

To define double-iterated integral for any index pair $i, j \in \{1, ..., d\}$, we use the following notation

$$S(X)^{i,j}_{a,t} = \int_{a<s<t} S(X)^i_{a,s}\, dX^j_s = \int_{a<r<s<t} dX^i_r\, dX^j_s \tag{2.10}$$

where $S(X)^i_{a,s}$ is defined in the same way as in (2.9) and a, r, s, t are such that $a < r < s$ and $a < s < t$. Note that $S(X)^i_{a,s}$ and X^j_s are real-valued paths, and therefore $S(X)^{i,j}_{a,t} : [a,b] \to \mathbb{R}$ are a real-valued path too.

Similarly we define triple-iterated integral for indices $i, j, k \in \{1, ..., d\}$:

$$S(X)^{i,j,k}_{a,t} = \int_{a<s<t} S(X)^{i,j}_{a,s}\, dX^k_s = \int_{a<q<r<s<t} dX^i_q\, dX^j_r\, dX^k_s \tag{2.11}$$

Note, that $S(X)^{i,j,k}_{a,t} : [a,b] \to \mathbb{R}$ is a real-valued path, since $S(X)^{i,j}_{a,s}$ and X^k_s are real-valued paths.

Continuing to do this $k \in 1, ..., d$ times, we determine the *k-fold iterated integral* of X, which is also a real-valued path, since $S(X)^{i_1,...,i_{k-1}}_{a,s}$ and $X^{i_k}_s$ are real-valued paths. For k-fold iterated integral we set following notation

$$S(X)^{i_1,...,i_k}_{a,t} = \int_{a<t_k<t} ... \int_{a<t_1<t_2} dX^{i_1}_{t_1} ... dX^{i_k}_{t_k}. \tag{2.12}$$

Now, with the definition and examples of iterated integrals, we can finally formulate a definition of the signature of a path X.

Definition 2. The *signature* of a path $X : [a,b] \to \mathbb{R}^d$, denoted by $S(X)_{a,b}$, is the sequence of all the iterated integrals of the path X. Formally, $S(X)_{a,b}$ is the sequence of real numbers

$$S(X)_{a,b} = (1, S(X)^1_{a,b}, ..., S(X)^d_{a,b}, S(X)^{1,1}_{a,b}, S(X)^{1,2}_{a,b}, ...) \tag{2.13}$$

where the first component, by convention, is equal to 1, and the superscripts run along the set of all *multi-indexes*

$$W = \{(i_1, ..., i_k) | k \geq 1, i_1, ..., i_k \in \{1, ..., d\}\} \tag{2.14}$$

The set W above is also frequently called the set of *words* on the *alphabet* $A = \{1, ..., d\}$ consisting of d letters.

Example 5. Let's say we have an alphabet that consists of only two letters $A = \{1, 2\}$. Then from these letters an infinite number of words can be formed.

$$W_3 = (1, 2, 11, 12, 21, 22, 111, 112, 121, 122, 211, 212, 222...) \tag{2.15}$$

Obviously, when using signatures in calculations, we cannot compute an infinite number of signature components. Therefore, in the implementation, we will use only truncated forms of signatures. However, even in this case, we may face the problem of calculating a large amount of data, as the number of terms in the signature grows exponentially. So we get the following formula for calculating the length of the signature vector of d-dimensional path with m level of truncation:

$$s_{len} = \frac{d(d^m - 1)}{d - 1}. \tag{2.16}$$

For a better illustration, we provide a small table with examples of signature dimensions.

	d=2	d=3	d=5	d=7
m=2	6	12	30	56
m=4	30	120	780	2800
m=5	62	364	3905	19607

Table 2.1: The sizes of signatures

Let now consider a simple example of a signature. For a better illustration, in the following example we will look at one-dimensional path.

Example 6. Assume we have $A = \{1\}$ as our index set, and the set of multi-indexes is $W = \{(1, ..., 1)|k \geq 1\}$, where "1" is written k times. It can be easily checked that the signature of path $X : [a, b] \to \mathbb{R}$, $X_t = t$ is calculated as

$$\begin{aligned} S(X)^1_{a,b} &= X_b - X_a, \\ S(X)^{1,1}_{a,b} &= \frac{(X_b - X_a)^2}{2!}, \\ S(X)^{1,1,1}_{a,b} &= \frac{(X_b - X_a)^3}{3!}, \\ &\vdots \end{aligned} \tag{2.17}$$

We can conclude that for one-dimensional paths, the signature of the path $X : [a, b] \to \mathbb{R}$ depends only on $X_b - X_a$, i.e. on its increments.

Example 7. Let now consider a two-dimensional path. In this case the set of indexes is $A = \{1, 2\}$ and multi-indexes set is given by

$$W = \{(i_1, ..., i_k) | k \geq 1, i_1, ..., i_k \in \{1, 2\}\}. \tag{2.18}$$

Let now assume that the path is defined in a following way

$$\begin{aligned} X_t &= \{X_t^1, X_t^2\} = \{t-1, t^3+1\},\ t \in [0,3], \\ dX_t &= \{\, dX_t^1,\, dX_t^2\} = \{\, dt, 3t^2\, dt\} \end{aligned} \tag{2.19}$$

In the next calculations, we will show that the signature terms are not computed simply by increments, as was the case before. Let's show the computation of some first components of the signature.

$$\begin{aligned} S(X)^1_{0,3} &= \int_{0<t<3} dX_t^1 = \int_0^3 dt = X_3^1 - X_0^1 = 3, \\ S(X)^2_{0,3} &= \int_{0<t<3} dX_t^2 = \int_0^3 3t^2 dt = X_3^2 - X_0^2 = 27, \\ S(X)^{1,1}_{0,3} &= \iint_{0<t_1<t_2<3} dX^1_{t_1}\, dX^1_{t_2} = \int_0^3 [\int_0^{t_2} dt_1]\, dt_2 = \frac{9}{2}, \\ S(X)^{1,2}_{0,3} &= \iint_{0<t_1<t_2<3} dX^1_{t_1}\, dX^2_{t_2} = \int_0^3 [\int_0^{t_2} dt_1]\, 3t_2^2 dt_2 = \frac{243}{4}, \\ S(X)^{2,1}_{0,3} &= \iint_{0<t_1<t_2<3} dX^2_{t_1}\, dX^1_{t_2} = \int_0^3 [\int_0^{t_2} 3t_1^2\, dt_1]\, dt_2 = \frac{81}{4}, \\ S(X)^{2,1}_{0,3} &= \iint_{0<t_1<t_2<3} dX^2_{t_1}\, dX^2_{t_2} = \int_0^3 [\int_0^{t_2} 3t_1^2\, dt_1] 3t_2^2\, dt_2 = \frac{729}{2}, \\ S(X)^{1,1,1}_{0,3} &= \iiint_{0<t_1<t_2<t_3<3} dX^1_{t_1}\, dX^1_{t_2}\, dX^1_{t_3} = \int_0^3 [\int_0^{t_3} [\int_0^{t_2} dt_1]\, dt_2]\, dt_3 = \frac{9}{2}, \\ &\ \ \vdots \end{aligned} \tag{2.20}$$

Proceeding in the same way, it is possible to calculate each term of the signature $S(X)_{0,3}$. But once we get the results of the first calculations, we can notice that we only use the information encoded in the differential dX to calculate iterated integrals (signature terms), rather than using the path information directly from X. Thus we can note that signatures depend on changes in path X, but not on a concrete position. Moreover, we observe that there were ∓ 1 shifts in the path coordinates, but this had no effect on the calculation of the iterated integrals.

2.3 Properties of signature

In this chapter we will describe some main properties of the signatures of path.

Let us state a lemma that describes one of the fundamental properties of the signature, which follows directly from the definition of iterated integrals and the fact that the starting point of the path does not affect iterated integrals. The proof for this lemma can be found in [9].

Lemma 1. Let define the path $\tilde{X}_t = X_t + x$, where $x \in \mathbb{R}^d$, then path $\tilde{X}_t$ will have the same signature as path X_t, i.e. $S(\tilde{X})^{i_1,...,i_k}_{a,b} = S(X)^{i_1,...,i_k}_{a,b}$.

2.3.1 Invariance under time reparameterizations

By *reparameterization* we will consider a continuous, increasing, surjective function $\phi : [a, b] \to [a, b]$.

Let assume that we have two real-valued paths $X, Y : [a, b] \to \mathbb{R}$ and reparameterization $\phi : [a, b] \to [a, b]$. Moreover, determine the reparameterized paths $\tilde{X}, \tilde{Y} : [a, b] \to \mathbb{R}$ such that $\tilde{X}_t = X_{\phi(t)}$ and $\tilde{Y}_t = Y_{\phi(t)}$. Note that using the substitutions with reparametrization we can write the following

$$\dot{\tilde{X}}_t = \dot{X}_{\phi(t)}\dot{\phi(t)}. \tag{2.21}$$

Using these notations we can write the path integral of $\tilde{X}$ and $\tilde{Y}$ as following

$$\int_a^b \tilde{Y}_t \, d\tilde{X}_t = \int_a^b Y_{\phi(t)}\dot{X}_{\phi(t)}\dot{\phi(t)} \, dt \overset{(2.21)}{=} \int_a^b Y_{\phi(t)} \, dX_{\phi(t)}, \tag{2.22}$$

what concludes the invariance of path integrals from time reparametrization.

The same finding can be applied to the d-dimensional path $X : [a, b] \to \mathbb{R}^d$. As a result, we can formulate following result for the signatures.

Lemma 2. The signature $S(X)_{a,b}$ is invariant under the time reparametrization of path X.

Proof. Consider reparameterization $\phi : [a, b] \to [a, b]$ and path $\tilde{X}_t = X_{\phi(t)}$. Thus, it follows from the above and from the fact, that every term of the signature $S(X)^{i_1,...,i_k}_{a,b}$ can be defined as an iterated linear integral of path X, that

$$S(\tilde{X})^{i_1,...,i_k}_{a,b} = S(X)^{i_1,...,i_k}_{a,b}, \forall k \geq 0, i_1, ..., i_k \in \{1, ..., d\} \tag{2.23}$$

This gives us the conclusion that the signatures are invariant to the time reparameterization. □

From this we can conclude that with the use of the signature we will not be able to understand and restore the exact speed of changes of the path.

2.3.2 Shuffle product

In this section we will define a shuffle product of multix-indexes and shuffle product identity for signatures, which was first shown by Ree [27]. Due to this fundamental property, the calculation of the product of two signature terms $S(X)^{i_1,...,i_k}_{a,b}$ and $S(X)^{j_1,...,j_k}_{a,b}$ can be replaced by the calculation of the sum of the signature terms of $S(X)_{a,b}$, whose indices depend only on the multi-indexes $(i_1, ..., i_k)$ and $(j_1, ..., j_k)$.

Firstly, we call by (k, m)-*shuffle* a permutation σ of the set $\{1, ..., k + m\}$, if $\sigma^{-1}(1) < ... < \sigma^{-1}(k)$ and $\sigma^{-1}(k + 1) < ... < \sigma^{-1}(k + m)$. For a better understanding of such permutations, one may recall the famous riffle shuffle method of two piles of cards. An important feature of shuffling is that when shuffling two packs of cards into one, the order of the cards of the original piles will remain the same in the final pile. In other words, with this method of card shuffling it would not matter if the cards of the first stack were mixed with one card or several cards of the other deck, the order of the cards will not be altered. Back to our notations, so by shuffle of indexes $(1, ..., k)$ and $(k + 1, ..., k + m)$ we will denote a list $(\sigma(1), ..., \sigma(k + m))$. Note, that the set of all (k, m)-shuffles we denote by $\mathrm{Suffles}(k, m)$.

Definition 3. Consider two multi-indexes $I = (i_1, ..., i_k)$ and $J = (j_1, ..., j_m)$ with $i_1, ..., i_k, j_1, ..., j_m \in \{1, ..., d\}$. Define the multi-index

$$(r_1, ..., r_k, r_{k+1}, ..., r_{k+m}) = (i_1, ..., i_k, j_1, ..., j_m) \tag{2.24}$$

The *shuffle product* of I and J, denoted $I \sqcup\!\sqcup J$, is a finite set of multi-indexes of length $k + m$ defined as follows

$$I \sqcup\!\sqcup J = \{(r_{\sigma(1)}, ..., r_{\sigma(k+m)}|\sigma \in \mathrm{Shuffles}(k, m)\}. \tag{2.25}$$

Let us now state the theorem, which shows us the shuffle product identity for signatures from Ree [27].

Theorem 1. For a path $X : [a, b] \to \mathbb{R}^d$ and two multi-indexes $I = (i_1, ..., i_k)$ and $J = (j_1, ..., j_m)$ with $i_1, ..., i_k, j_1, ..., j_m \in \{1, ..., d\}$, it holds that

$$S(X)^I_{a,b} S(X)^J_{a,b} = \sum_{K \in I \sqcup\!\sqcup J} S(X)^K_{a,b}. \tag{2.26}$$

The idea for the proof was taken from [14].

Proof. To show this identity we have to partition the domain of integration. As a result, we get the following calculations

$$\begin{aligned} S(X)^I_{a,b}S(X)^J_{a,b} &= \int \cdots \int_{a<u_1<\ldots<u_k<b} dX^{i_1}_{u_1} \ldots dX^{i_k}_{u_k} \int \cdots \int_{a<t_1<\ldots<t_m<b} dX^{j_1}_{t_1} \ldots dX^{j_m}_{t_m} \\ &= \sum_{\sigma \in \text{Shuffles}(k,m)} \int \cdots \int_{a<v_1<\ldots<v_{k+m}<b} dX^{r_{\sigma(1)}}_{v_1} \ldots dX^{r_{\sigma(k+m)}}_{v_{k+m}} \\ &= \sum_{K \in I \sqcup\!\sqcup J} S(X)^K_{a,b}, \end{aligned} \tag{2.27}$$

with $(r_1, ..., r_k, r_{k+1}, ..., r_{k+m}) = (i_1, ..., i_k, j_1, ..., j_m)$. $\square$

Example 8. Let we have a two-dimensional path $X : [a, b] \to \mathbb{R}^2$ and its signature $S(X)_{a,b}$. Thus from shuffle product we have

$$\begin{aligned} &S(X)^1_{a,b}S(X)^2_{a,b} = S(X)^{1,2}_{a,b} + S(X)^{2,1}_{a,b} \\ &S(X)^{1,2}_{a,b}S(X)^1_{a,b} = S(X)^{1,1,2}_{a,b} + S(X)^{1,1,2}_{a,b} + S(X)^{1,2,1}_{a,b} = 2S(X)^{1,1,2}_{a,b} + S(X)^{1,2,1}_{a,b}, \end{aligned} \tag{2.28}$$

where for the first equality we have the following multi-indices $I = (1)$ and $J = (2)$. As a result of shuffle product we obtain $(\mathbf{1}) \sqcup\!\sqcup (\mathit{2}) = \{(\mathbf{1}, \mathit{2}), (\mathit{2}, \mathbf{1})\}$. Accordingly, for the second equality we have $I = (1, 2)$ and $J = (1)$. Thus we get following $(\mathbf{1}, \mathbf{2}) \sqcup\!\sqcup (\mathit{1}) = \{(\mathit{1}, \mathbf{1}, \mathbf{2}), (\mathbf{1}, \mathit{1}, \mathbf{2}), (\mathbf{1}, \mathbf{2}, \mathit{1})\}$.

This example shows that product of two terms of the signature can be expressed as a sum of higher order terms. More generally, this property of signautres shows that non-linear operations on low-level terms can be expressed by a linear combination of high-level signature terms. Obtaining linear combinations is one of the reasons why signatures might be used in machine learning as an efficient feature transformation.

2.3.3 Chen's identity

In this part we will define Chen's equality, which was first presented in [6]. It will help us to describe the path and signature relationship from an algebraic perspective. First, however, let us define the algebra of formal power series, which will allow us to identify signatures in terms of formal power series.

Definition 4. Let $e_1, ..., e_d$ be d formal indeterminates. The algebra of noncommutative *formal power series* in d indeterminates is the vector space of all series of the form

$$\sum_{k=0}^{\infty} \sum_{i_1,\ldots,i_k \in \{1,\ldots,d\}} \lambda_{i_1,\ldots,i_k} e_{i_1}, \ldots, e_{i_k}, \tag{2.29}$$

where the second summation runs over all multi-indexes $(i_1, ..., i_k)$, $i_1, ..., i_k \in \{1, ..., d\}$, and $\lambda_{i_1,\ldots,i_k}$ are real numbers.

By formal polynomial of power $k \geq 1$ we mean a formal power series which has only a finite number k of coefficients that are not zeros. Moreover, the terms $e_{i_1}...e_{i_k}$ denote monomials. The space of such formal power series is called the *tensor algebra* of $\mathbb{R}^d$.

Let us define the tensor product $\otimes$ between monomials by

$$e_{i_1}...e_{i_k} \otimes e_{j_1}...e_{j_m} = e_{i_1}...e_{i_k} e_{j_1}...e_{j_m}. \tag{2.30}$$

Moreover, one can define addition and scalar product on the space of formal power series to obtain a vector space. Thus, let's define addition as

$$\begin{aligned}(\sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} \lambda_{i_1,...,i_k} e_{i_1}, ..., e_{i_k}) + (\sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} \mu_{i_1,...,i_k} e_{i_1}, ..., e_{i_k}) = \\ = \sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} (\lambda_{i_1,...,i_k} + \mu_{i_1,...,i_k}) e_{i_1}, ..., e_{i_k},\end{aligned} \tag{2.31}$$

and scalar product as

$$c(\sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} \lambda_{i_1,...,i_k} e_{i_1}, ..., e_{i_k}) = \sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} c\lambda_{i_1,...,i_k} e_{i_1}, ..., e_{i_k}, \tag{2.32}$$

In the following equation, let us show how $\otimes$ product extends to all formal power series

$$\begin{aligned}(\sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} \lambda_{i_1,...,i_k} e_{i_1}, ..., e_{i_k}) \otimes (\sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} \mu_{i_1,...,i_k} e_{i_1}, ..., e_{i_k}) \\ = \lambda_0 \mu_0 + \sum_{i=1}^{d} (\lambda_0 \mu_i + \lambda_i \mu_0) e_i + \sum_{i,j=1}^{d} (\lambda_0 \mu_{i,j} + \lambda_i \mu_j + \lambda_{i,j} \mu_0) e_i e_j + ...\end{aligned} \tag{2.33}$$

Thus, to obtain an algebra from the space of formal power series, we defined addition and scalar product to obtain a vector space structure and extended $\otimes$ product to all formal power series.

Now we can rewrite signatures $S(X)_{a,b}$ using formal power series, since the index set of monomials coincides with the multi-indexes of signatures $(i_1, ..., i_k)$, $i_1, ..., i_k \in \{1, ..., d\}$. Moreover, the coefficients of each monomials $e_{i_1}...e_{i_k}$ will be noted by $S(X)_{a,b}^{i_1,...,i_k}$.

$$S(X)_{a,b} = \sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} S(X)_{a,b}^{i_1,...,i_k} e_{i_1}...e_{i_k}, \tag{2.34}$$

where the first component of signature $S(X)_{a,b}^0 = 1$ (corresponding to $k = 0$).

Let state the last definition, which is needed for the Chen's identity, which will allow us to compute the signature of the concatenated paths using tensor multiplication of the signatures of the original paths.

Definition 5. For two paths $X : [a,b] \to \mathbb{R}^d$ and $Y : [b,c] \to \mathbb{R}^d$, we define their *concatenation* as the path $X * Y : [a,c] \to \mathbb{R}^d$ for which $(X*Y)_t = X_t$ for $t \in [a,b]$ and $(X*Y)_t = X_b + (Y_t - Y_b)$ for $t \in [b,c]$

Theorem 2 (Chen's identity)**.** Assume $X : [a,b] \to \mathbb{R}^d$ and $Y : [b,c] \to \mathbb{R}^d$ are two paths. Then

$$S(X*Y)_{a,c} = S(X)_{a,b} \otimes S(Y)_{b,c}. \tag{2.35}$$

In other words, using Chen's equality, from the concatenation product "$*$" we can get the tensor product $\otimes$.

Remark 1. Chen's identity allows us to compute the signatures of paths containing "streaming" information more efficiently. For example, if we have computed the signature of the path with information $X = (x_1, ...x_{100})$, but now we have additional information $X_{new} = (x_{101}, ..., x_{105})$. Using this identity, we can compute only the path signature of the new information X_{new}, and we do not need to recalculate the signature of the whole concatenated paths $X * X_{new}$.

The following proof of the Chen's identity was taken from [22].

Proof. Set $Z = X * Y$ and $S(Z) = (1, Z^1_{a,c}, Z^2_{a,c}, ...)$. Let choose $n \geq 1$ and compute $Z^n_{a,c}$. We obtain

$$Z^n_{a,c} = \int \cdots \int_{a<t_1<...<t_n<c} dZ_{t_1} \otimes ... \otimes dZ_{t_n} = \sum_{k=0}^{n} \int \cdots \int_{a<t_1<...<t_k<b<t_{k+1}<...<t_n<c} dZ_{t_1} \otimes ... \otimes dZ_{t_n} =$$
$$= \sum_{k=0}^{n} \int \cdots \int_{a<t_1<...<t_k<b} dZ_{t_1} \otimes ... \otimes dZ_{t_k} \otimes \int \cdots \int_{b<t_{k+1}<...<t_n<c} dZ_{t_{k+1}} \otimes ... \otimes dZ_{t_n}$$
$$= \sum_{k=0}^{n} X^k_{a,b} \otimes Y^{n-k}_{b,c}.$$

Hence, $S(Z) = S(X) \otimes S(Y)$.

□

2.3.4 Time-reversal

Next property of the signatures says, that under the tensor product $\otimes$ the signature $S(X)_{a,b}$ is inverse of the signature $S(\overleftarrow{X})_{a,b}$, where $X : [a,b] \to \mathbb{R}^d$ and $\overleftarrow{X}$ is given in a following definition.

Definition 6. Assume we have a path $X : [a,b] \to \mathbb{R}^d$, then by $\overleftarrow{X} : [a,b] \to \mathbb{R}^d$ define its time-reversal path, where $\overleftarrow{X}_t = X_{a+b-t}$ for all $t \in [a,b]$.

Proposition 1 (Chen, K.T. [7])**.** For a path $X : [a, b] \to \mathbb{R}^d$, it holds that

$$S(X)_{a,b} \otimes S(\overleftarrow{X})_{a,b} = 1 \tag{2.36}$$

Where 1 is an identity element under the tensor product $\otimes$, i.e. it is a formal power series which has $\lambda_0 = 1$ and all other coefficients $\lambda_{i_1,...,i_k} = 0$ for all $k \geq 1$ and $i_1, ..., i_k \in \{1, ..., d\}$. This proposition shows us that the signature of time-reversal path is inverse to the original path signature, $S(X)^{-1} = S(\overleftarrow{X})$.

Proof. The proof can be found in [14, Proposition 1.4]. □

2.3.5 Log signature

Here we will consider a definition of log signature, which can be described as a reduced form of the signature. In other words, log signatures describe the same information of the path as signatures, but requires fewer levels of truncation. Thus, to obtain the log signatures, one must take the logarithm of the signature in the algebra of formal power series.

Let us consider the following power series

$$x = \sum_{k=0}^{\infty} \sum_{i_1,...,i_k \in \{1,...,d\}} \lambda_{i_1,...,i_k} e_{i_1}, ..., e_{i_k} \tag{2.37}$$

where $\lambda_0 > 0$.

Define the logarithm of the power series x as following

$$\log x = \log(\lambda_0) + \sum_{n \geq 1} \frac{(-1)^n}{n} (1 - \frac{x}{\lambda_0})^{\otimes n}, \tag{2.38}$$

where $a^{\otimes n} = a \otimes .... \otimes a$, n times of a.

	d=2	d=3	d=5	d=7
m=2	6 **3**	12 **6**	30 **15**	56 **28**
m=4	30 **8**	120 **32**	780 **205**	2800 **728**
m=5	62 **14**	364 **80**	3905 **829**	19607 **4088**

Table 2.2: The sizes of signatures and log signatures(bold)

Definition 7. For a path $X : [a, b] \to \mathbb{R}^d$, the log signature of X is defined as the formal power series $\log S(X)_{a,b}$.

So to illustrate the difference in the signature and log signatures dimensionality, here is a table 2.2, where m is the truncation level and d is the path dimensions.

Although in some calculations log signatures have advantages than just signatures to save computation time. In this paper we will only consider signature calculations, as it also has disadvantages, of which those who are interested can find out about in [9].

Chapter 3

Expected signatures and stochastic processes

In this chapter we will consider basic definitions such as group-like elements group, polynomial identities, expected signatures, discuss the space of the signatures and also look at representations of space E. Moreover we will discuss the relation between expected signatures and laws of stochastic processes. This chapter is mainly referring to Chevyrev and Lyons [10].

In the last chapter we discussed signatures of certain paths, namely piecewise differentiable paths. This restriction was assumed in order to guarantee the existence of iterated path integrals as Riemann-Stieltjes integrals. However, in this chapter we will also consider integrals in the Young sense, since we briefly look at the signatures of rough paths, which extend the theory of classical smooth paths to the functions that may not be differentiable.

3.1 Universal locally m-convex algebra

Consider topological vector spaces V and W. We will assume that V is a normed, locally convex space. Define by $\mathbf{L}(V, W)$ the space of continuous linear maps between V and W. Moreover, denote by $V' = \mathbf{L}(V, \mathbb{R})$ and $\mathbf{L}(V) = \mathbf{L}(V, V)$. For a deeper understanding of topological vector spaces, algebras and their properties that we are using here, see Appendix 4.3.

In this section, to construct a universal topological algebra $E(V)$ of a specific category (namely, locally m-convex), we will use the method which was introduced by Lyons and Chevyrev [10]. It was based on the structure that was presented in [12] for cyclic cohomology and further studied in [29] for locally convex algebras with continuous multiplication and in [13] for commutative locally m-convex algebras.

Let us look at a topological algebra A, a tensor space $T(V) = \oplus_{k\geq 0} V^{\otimes k}$, where $V^{\otimes k} = V \otimes ... \otimes V$ (k times of topological vector space V), and the following statement:

$$\forall M \in \mathbf{L}(V, A),\ \textit{the extension } M : T(V) \to A \textit{ is continuous} \tag{3.1}$$

Then let us equip $T(V)$ with such a topology that for all topological algebras A the statement (3.1) holds. Furthermore, we will consider locally m-convex algebras in this work, the definition of such category of algebras is given in Appendix 4.3.

Assume that V is a locally convex space. Let us define by $E_a(V) = T(V)$ the space that equipped with the initial topology such that (3.1) is fulfilled for all locally m-convex algebras A (or in other words, all normed algebras A).Furthermore, let us denote the completion of space $E_a(V)$ by $E(V)$. Therefore, the set of continuous algebra homomorphisms $Hom(E_a, A)$ is in bijection with the space of continuous linear maps $\mathbf{L}(V, A)$.

From now on, we will omit the reference to V in the notations of the spaces E and E_a, unless clarification is required. We also have the following result for the spaces.

Lemma 3. The spaces E_a and E are m-convex algebra.

Proof. [see 23, p.14 and p.22] □

Remark 2. [see 10] If we start to construct a universal topological algebra with V as a general topological vector space, an easy verification shows that we arrive at the same space E_a as when we equip V with the finest locally convex topology coarser than its original.

Now we shall consider further definitions and corollaries that will help us to determine a representation of space E and the radius of convergence of elements from E.

Firstly, let define by Ψ a family of seminorms on V and ξ be a continuous seminorm on V. Then we call Ψ *fundamental* family if for any ξ on V, there $\exists \gamma \in \Psi$ and $\varepsilon > 0$ such that $\gamma \geq \varepsilon\xi$. Moreover, lets use the following notation $\Psi^* = \{n\gamma | n \geq 1, \gamma \in \Psi\}$.

Let V and W be locally convex spaces. Define *projective* seminorm (or tensor product of seminorms) by $\gamma \otimes \xi$, where γ and ξ are seminorms on V and W respectively:

$$\xi \otimes \gamma(x) := \inf\{\sum_{i=1}^{n} \xi(a_i)\gamma(b_i) : x = \sum_{i=1}^{n} a_i \otimes b_i,\ a_i \in V, b_i \in W\} \tag{3.2}$$

Let $V \otimes_\pi W$ denotes the *projective tensor product* and $V \hat{\otimes} W$ denotes the completion of this projective tensor product. Denote the operator norm of a continuous linear map M

from $\mathbf{L}(V,F)$ by $\gamma(M) = \sup_{\gamma(v)=1}\|Mv\|$, where F is a normed space and $\gamma(v)$ is the seminorm of the vector v, $\|\cdot\|$ is the norm of a linear operator. In other words, $\gamma(M)$ is the maximum value of $\|Mv\|$ over all vectors v in V such that $\gamma(v) = 1$.

By $exp(\gamma) = \sum_{k\geq 0}\gamma^{\otimes k}$ on E_a we denote the *projective extension* of a seminorm γ on V, i.e it is a way of "extending" γ to the projective tensor product $V^{\otimes_\pi k}$. Note, that it can be shown that the projective extension of γ is a seminorm on E_a, moreover it is a submultiplicative seminorm.

Let take a normed algebra A, such that $M \in \mathbf{L}(V,A)$, and γ be a seminorm on V such that $\gamma(M) \leq 1$. With these assumptions we can conclude that the projective extension of γ satisfies the inequality $exp(\gamma)(M_E) \leq 1$, where $M_E \in Hom(E_a, A)$ is a linear map that extends M. Hence we get the following proposition and its corollaries from [10]:

Proposition 2. Let Ψ be a family of seminorms on V. Then Ψ is a fundamental family of seminorms on V if and only if $exp(\Psi^*)$ is a fundamental family of seminorms on E.

Corollary 1. The space E is Hausdorff (resp. metrizable, separable) if and only if V is Hausdorff (resp. metrizable, separable).

In the following corollary we associate E with a subspace of $P(V) := \prod_{k\geq 0} V^{\hat{\otimes} k}$. Let denote x^k the projection of $x \in P$ on $V^{\hat{\otimes} k}$, such that x can be written as an infinite sequence $(x^0, x^1, x^2, ...)$. This notion of E space will allow us to associate the radius of convergence with E space.

Corollary 2. Let Ψ be a fundamental family of semi-norms on V. Then $E = \{x \in P | \forall\gamma \in \Psi^*, \sum_{k\geq 0}\gamma^{\otimes k}(x^k) < \infty\}$.

Intuitively, this means that E is the set of all elements of P that satisfy a certain convergence condition with respect to the projective seminorms in Ψ^*.

To get a better understanding of this, let's consider a case where $k = 2$. Let denote $P^{\hat{\otimes} 2} = \prod_{i,j\geq 0} V^{i,j}$, where $V^{i,j} \cong V^{\hat{\otimes}(i+j)}$. As a result from the corollary we have a following statement:

$$E^{\hat{\otimes} 2} = \{x \in P^{\hat{\otimes} 2} | \forall\gamma \in \Psi^*, \sum_{i,j\geq 0}\gamma^{\otimes(i+j)}(x^{i,j}) < \infty\}. \tag{3.3}$$

Assume $\rho^k : E \to V^{\hat{\otimes} k}$ to be the projection $\rho^k(x) = x^k$ and $\pi^k : E \to \oplus_{j=0}^k V^{\hat{\otimes} j}$ the projection $\pi^k(x) = (x^0, ..., x^k)$. Thus we get one more corollary.

Corollary 3. The operators $T^{(n)} := \sum_{k=0}^n \rho^k : E \to E$ converge uniformly on bounded sets to the identity operator on E.

Whenever the space V is normed, we assume that $V^{\otimes k}$ is equipped with a projective norm (i.e. tensor product).

Let take a look at the series

$$\sum_{k\geq 0}\|x^k\|\lambda^k. \tag{3.4}$$

Then we define a radius of convergence by $R(x)$ for $x \in P$ as the radius of convergence of the series 3.4. The radius of convergence is defined as the largest real number R such that the series converges for all complex numbers λ with magnitude less than R. Moreover, from Corollary 2, as was shown in [10], we have that the radius of convergence is infinite, $R(x) = \infty$, if and only if $x \in E$.

Intuitively, the radius of convergence measures how fast the terms in the series decay as k increases. A larger radius of convergence indicates that the terms in the series decay more slowly, which means that the series is more likely to converge for a wider range of values of λ. Further, by using the convergence radius, we will be able to link the distribution of the random variable X to the expected signature of that variable. This means that the radius of convergence can provide some information about the behavior of the series that defines the expected signature, and therefore about the distribution of x.

3.2 Group-like elements and expected signatures

In this section we assume that all measures are Borel. The space of probability measures on S is denoted by $\mathcal{P}(S)$, where S is a topological space endowed with the topology of weak convergence on $C_b(S, \mathbb{C})$. Note that we consider $T(V)$ as a Hopf algebra with coproduct $\Delta v = 1 \otimes v + v \otimes 1$, $\Delta : V \to T(V)^{\otimes 2}$, for all $v \in V$, which can be extended by universal property of $T(V)$ to a homomorphism $\Delta : T(V) \to T(V)^{\otimes 2}$. Moreover, from ([28], Proposition 1.10) in $T(V)$ we have *antipode* $\alpha(v_1...v_k) = (-1)^k v_k...v_1$, $\alpha : T(V) \to T(V)$, for all $v_1...v_k \in V^{\otimes k}$.

Further on we will assume V to be a locally convex space. Note, that from [23, p.378] we have that $E^{\hat{\otimes} 2}$ is also locally m-convex algebra. From this fact and that $\Delta \in \mathbf{L}(V, E^{\otimes_\pi 2})$, due to the universal property of E we get that extension $\Delta : E \to E^{\hat{\otimes} 2}$ is continuous. Moreover, let us mention a few more results from [10] in the remark below.

Remark 3. As a result of the findings above, we have that the antipode α can be extended to a continuous linear map $\alpha : E \to E$. This gives to the space E an "almost" Hopf algebra structure. Here we have "almost" since under the coproduct Δ space E mapped to the completion $E^{\hat{\otimes} 2}$ and not to $E^{\otimes 2}$ itself.

Let now us define two groups, which we will use further. Denote by $U(V) = \{g \in E | \alpha(g) = g^{-1}\}$ the *group of unitary elements* of E. Note, that U is closed in E, since the map $\varphi : x \to (\alpha(x)x, x\alpha(x))$, $\varphi : E \to E \times E$ is continuous and $U = \varphi^{-1}\{(\alpha(x)x, x\alpha(x))\} = \varphi^{-1}\{(1,1)\}$. Let us define the second group by $G(V) = \{g \in E | \Delta(g) = g \otimes g, g \neq 0\}$ the *group of group-like elements* of E. The group G is also closed in E since $g^0 = 1$ for all $g \in G$ and $G = \psi^{-1}\{0\} \setminus \{0\}$ for the continuous map $\psi : x \to x \otimes x - \Delta(x)$ from E into $E^{\hat{\otimes}2}$. Moreover, from the fact that multiplication and inversion in E are continuous [see 23, p.5, p.52] and if G and U are endowed with a subspace topology, then they are topological groups. Note, that we have an inclusion $G \subset U$.

As it was mentioned before, signature of a one-dimensional path X_t in $\mathbb{R}$ depends only on its increments $X_b - X_a$ and have a form $(1, X_t - X_0, (X_t - X_0)^2/2!, ...)$. If we assume X_t be a stochastic process, we can define expected signature in $\mathbb{R}$ case as expectations of its iterated integrals $(1, \mathbb{E}[X_t - X_0], \mathbb{E}[X_t - X_0]^2/2!, ...)$.

Assume F is a locally convex space, thus F-valued random variable X is weakly integrable (in sense of Gelfand-Pettis), i.e. $\mathbb{E}[X]$ exists, if $f(X)$ is integrable for all $f \in F'$ and if there exists $\mathbb{E}[X] := x \in F$ such that $\mathbb{E}[f(X)] = f(x)$. Moreover, denote by μ the probability measure associated with X and its barycenter we denote by $\mu^* = \mathbb{E}[X]$.

Definition 8. By *expected signature* of E-valued random variable X we denote the sequence

$$\mathbf{Sig}(X) := \mathbb{E}[S(X)] = (\mathbb{E}[X^0], \mathbb{E}[X^1], ...) \in P = \prod_{k \geq 0} V^{\hat{\otimes} k} \tag{3.5}$$

where X^k is integrable for all $k \geq 0$.

Lets now define radius of convergence for series of integrable and norm-integrable variables. Assuming V is normed space and taking a series

$$\sum_{k \geq 0} \mathbb{E}[\|X^k\|]\lambda^k, \tag{3.6}$$

we define $r_1(X)$ as a radius of convergence of (3.6), where $r_1(X) = 0$ if X^k is not norm-integrable for some $k \geq 0$.

Likewise define a radius of convergence $r_2(x)$ of the series

$$\sum_{k \geq 0} \|\mathbb{E}[X^k]\|\lambda^k, \tag{3.7}$$

such that $r_2(X) = 0$ if X_k is not integrable for some $k \geq 0$.

Note that for these radiuses of convergence we have following relations, namely $r_2(X) = R(\mathbf{Sig}(X))$, i.e. r_2 is equal to the radius of convergence of the expected signature, and $r_1(X) \leq r_2(X)$.

Due to the definition, expected signature $\mathbf{Sig}(X)$ exists when E-valued random variable X is integrable. Moreover, let now consider a G-valued random variable and show the opposite. We should keep in mind, though, that we defined E as a subspace of P previously.

Proposition 3. Let X be a G-valued random variable. Then X is weakly integrable if and only if $\mathbf{Sig}(X)$ exists and lies in E. In this case $\mathbb{E}[X] = \mathbf{Sig}(X)$.

In other words, the proposition shows that the random variable $X \in G$ is integrable (as an E-valued random variable) in the case when the projections X^k are (weakly) integrable and its normed expectation $\|\mathbb{E}[X^k]\|$ decays fast enough as $k \to \infty$. However, it is clear that this will not hold for an arbitrary E-valued random variable.

We should note that $\forall f \in E'$ we have $f^{\otimes 2} \circ \Delta \in E'$. Moreover, for all $g \in G$ it holds that $f(g)^2 = f^{\otimes 2}(\Delta g)$ and $\forall \mu \in \mathcal{P}(G)$ we have

$$\mu(|f|) \leq \sqrt{\mu(f^2)} = \sqrt{\mu(f^{\otimes 2} \circ \Delta)}. \tag{3.8}$$

Thus one can observe that if $\mathbb{E}[X]$ exists and $\mu \in \mathcal{P}(G)$ then we have $\forall f \in E'$ the real random variable $f(X)$ has moments of all orders which are finite.

To show Proposition 3 we use proof from [10, Prop. 3.2]. The main idea of which is that assuming existence of $\mathbb{E}[X]$ for all $k \geq 0$, we approximate $\mathbb{E}[f(X)]$ by $\sum_{k=0}^{n} \mathbb{E}[f(X^k)]$. Moreover, using inequality (3.8) and grading of the coproduct Δ, we apply dominated convergence to obtain $\mathbb{E}[f(X)] = \sum_{k=0}^{n} \mathbb{E}[f(X^k)]$.

Proof. [see 10, Prop. 3.2] The "only if" direction is clear. Assume that $\mathbf{Sig}(X)$ exists and $\mathbf{Sig}(X) \in E$. As usual, let μ be the measure on G associated to X. We are required to show that f is μ-integrable and that $\mu(f) = \langle f, \mathbf{Sig}(X)\rangle$ for all $f \in E'$.

We recall the projection $\rho^k : E \to V^{\hat{\otimes} k}$ and canonically embed $(V^{\hat{\otimes} k})'$ into E' by $f \to f\rho^k =: f^k$ for all $f \in (V^{\hat{\otimes} k})'$. By Corollary 3, $\sum_{k=0}^{n} f^k$ converges uniformly on bounded sets (and a fortiori pointwise) to f.

Remark that for any $f \in E'$, $f \in (V^{\hat{\otimes} k})'$ if and only if $f = f^k$. Recall that Δ is a graded linear map from $T(V)$ to $T(V)^{\otimes 2}$. In particular, for all $f_1 \in (V^{\hat{\otimes} k})'$, $f_2 \in (V^{\hat{\otimes} m})'$ and $x \in T(V)$, it holds that

$$(f_1 \otimes f_2)\Delta(x) = (f_1 \otimes f_2)\Delta(x^{k+m}). \tag{3.9}$$

As $T(V)$ is dense in E, (3.9) holds for all $x \in E$, from which it follows that $(f_1 \otimes f_2) \circ \Delta \in (V^{\hat{\otimes} k+m})'$.

Let $f \in E'$ and note that $\mu(f^k) = \langle f^k, \mathbb{E}[X^k] \rangle$ for all $k \geq 0$. Since μ has support on G, it follows from (3.8) and (3.9) that

$$\mu\left(\sum_{k\geq 0}|f^k|\right) \leq \sum_{k\geq 0}\sqrt{\mu((f^k)^{\otimes 2}\circ\Delta)} = \sum_{k\geq 0}\sqrt{(f^k)^{\otimes 2}\Delta\mathbb{E}[X^{2k}]}. \tag{3.10}$$

Without loss of generality, we can assume that $|f(1)| \leq 1$. Let γ be a semi-norm on Vsuch that $exp(\gamma) \geq |f|$ and ξ a semi-norm on E such that $\xi \geq exp(\gamma)^{\otimes 2} \circ \Delta$. It follows that $exp(\gamma) \geq |f^k|$ for all $k \geq 0$, and thus $\xi \geq |(f^k)^{\otimes 2} \circ \Delta|$ for all $k \geq 0$.

Since $\mathbf{Sig}(X) \in E$, it follows from Corollary 2 that $\sum_{k\geq 0}\sqrt{\xi(\mathbb{E}[X^k])}$ is finite, and hence (3.10) is finite. By dominated convergence, we obtain

$$\mu(f) = \lim_{n\to\infty}\mu(\sum_{k=0}^{n} f^k) \tag{3.11}$$

It then follows that $\mu(f) = \langle f, \mathbf{Sig}(X)\rangle$ as desired since

$$\mu(\sum_{k=0}^{n} f^k) = \sum_{k=0}^{n}\langle f^k, \mathbb{E}[X^k]\rangle \to \langle f, \mathbf{Sig}(X)\rangle. \tag{3.12}$$

$\square$

Corollary 4. [see 10] Assume we have a G-valued random variable X and a normed space V. Then we have that $\mathbb{E}[X] \in E$ exists if and only if radius of convegence $r_2(X) = \infty$, i.e., $\mathbf{Sig}(X)$ exists and has an infinite radius of convergence. In this case $\mathbb{E}[X] = \mathbf{Sig}(X)$.

Let us state another proposition in which we show a partially opposite relation of the radius of convergence r_1 and r_2 when $V = \mathbb{R}^d$ and X is an G-valued random variable. But assume first that V is a normed space. Furthermore, we note that for all $v \in V$ we have that $\|\Delta v\| = 2\|v\|$. As a result we have $\|\Delta|_{V^{\hat{\otimes} k}}\| = 2^k$ and following equation

$$\|\Delta x^k\| \leq 2^k\|x^k\| \text{ for all } x \in E. \tag{3.13}$$

Let we have $V = \mathbb{R}^d$ and define $e_I = e_{i_1}...e_{i_k} \in V^{\otimes k}$ for an alphabet $\{1, ..., d\}$ and a word $I = i_1...i_k$, where $e_{i_1}, ..., e_{i_k}$ is a standard basis of V. Moreover, let us equip the space V with the l^1 norm from this basis. Hence, from the grading of Δ we have

$$\begin{aligned}\mathbb{E}[\|X^k\|^2] = \mathbb{E}[(\sum_{|I|=k} |\langle e_I, X^k\rangle|)^2] &\leq d^k\mathbb{E}[(\sum_{|I|=k}\langle e_I, X^k\rangle^2)] \\ &= d^k \sum_{|I|=k} e_I^{\otimes 2}\Delta\mathbb{E}[X^{2k}] \\ &\leq d^k\|\Delta\mathbb{E}[X^{2k}]\|,\end{aligned} \tag{3.14}$$

note that for the last inequality we used the argument that $(e_I \otimes e_J)_{|I|=|J|=k}$ is an l^1 basis for $V^{\otimes 2k}$. As a consequence of equation (3.13) we have the following proposition.

Proposition 4. Let X be a $G(\mathbb{R}^d)$-valued random variable. It follows that

$$\mathbb{E}[\|X^k\|^2] \leq d^k 2^{2k}\|\mathbb{E}[X^{2k}]\|. \tag{3.15}$$

In particular, $r_1(X) \leq r_2(X) \leq 2\sqrt{d}r_1(X)$.

For the proof of the proposition look at [10, Chapter 3]. This proposition shows how the convergence radiuses r_1 and r_2 are related to each other. Moreover, we further state a theorem in which we show the conditions for the radius r_1 to be infinite for an E-valued variable.

3.3 Representations of space E

In this section we will discuss the representations and separability of space E. It is based on the theory of Chevyrev and Lyons from [10]. Lets us remind that for a Hopf algebra, we can define dual of representations via the antipode, namely $M^*(x) := M(\alpha(x))^*$. Moreover, one may determine the tensor product using the coproduct, $M_1 \otimes M_2(x) := (M_1 \otimes M_2)\Delta(x)$. Let consider the representations of the space E which are continuous and take them over Hilbert spaces with finite dimension. Then we should note that such represenations are closed under duals and tensor products since coproduct Δ and antipode α are continuous.

Definition 9. [see 10] Denote by $\mathcal{A}(V)$ the family of finite dimensional representations of E which arise from extensions of all linear maps $M \in \mathbf{L}(V, \mathfrak{u}(H_M))$, where H_M ranges over all finite dimensional Hilbert spaces and $\mathfrak{u}(H_M)$ denotes the Lie algebra of the anti-Hermitian operators on H_M . Denote by $\mathcal{C}(V)$ the set of corresponding matrix coefficients, i.e., the set of linear functionals $M_{u,v} \in \mathbf{L}(E, \mathbb{C})$, where $M_{u,v}(x) = \langle M(x)u, v\rangle$ for all $M \in \mathcal{A}$ and $u, v \in H_M$

One of the key features of the family $\mathcal{A}(V)$ is that it is closed when taking the duals of representations and tensor products. One can also note that the family $\mathcal{A}(V)$ contains representations of E such that they are of finite dimension and for all $x \in E$ we have $M(\alpha x) = M(x)^*$, in other words, it is an involutory family. Hence, it implies that every representation $M \in \mathcal{A}$ is an unitary representation of the group U, therefore of the group G.

We note that for any representations M_1, M_2 of the space E the tensor product that we mentioned before, $M_1 \otimes M_2$, and tensor product of representations from group theory agree on the group of group-like elements G. Furthermore, for $M \in \mathcal{A}$ one may define on a group U the dual represenation M^* as the conjugate representation of M on U. Hence, we have that $\mathcal{C}|_G$ forms an involutive subalgebra ($*$-subalgebra) of $C_b(G, \mathbb{C})$.

Define by S a topological space and by F a $*$-subalgebra of $C_b(G, \mathbb{C})$, which is separating. Let μ and ν be tight Borel measures on S. Thus, from the Stone-Weierstrass theorem we have that $\mu = \nu$ if and only if $\mu(f) = \nu(f)$ for all $f \in F$ [see 4, Exercise 7.14 .79]. We now may state the following lemma.

Lemma 4. [see the proof in 10] Assume that the family of represenations $\mathcal{A}$ separates the points of group G. Then for tight Borel measures μ, ν on G, $\mu = \nu$ if and only if $\mu(f) = \nu(f)$ for all $f \in \mathcal{C}$, or equivalently, $\mu(M) = \nu(M)$ for all representations $M \in \mathcal{A}$.

Further we will consider the requirements under which algebra homomorphisms of E separate points in general case. However, in Theorem 4 we will have a look at the case when $V = \mathbb{R}^d$ and show that $\mathcal{A}(\mathbb{R}^d)$ separates the points in $E(\mathbb{R}^d)$.

3.3.1 Separation of points

Assume A be a Banach algebra and M be a continous liner map from $\mathbf{L}(V, A)$. Denote the algebra homomorphism on E by (λM) which is induced by $\lambda M \in \mathbf{L}(V, A)$, where λ might be complex if we consider algebra A over $\mathbb{C}$. In case $\lambda \in \mathbb{R}$, lets us denote by $\delta_\lambda : E \to E$ a dilation operator $\delta_\lambda(x^0, x^1, ...) = (\lambda^0 x^0, \lambda^1 x^1, ...)$. Moreover, we can note that $(\lambda M) = M\delta_\lambda$ for $\lambda \in \mathbb{R}$.

Lemma 5. Let V be a locally convex space, A be a Banach algebra and $M \in \mathbf{L}(V, A)$. Let $x \in E$ such that $M(x^k) \neq 0$ for some $k \geq 0$. Then there exists $\epsilon > 0$ sufficiently small such that $(\epsilon M)(x) \neq 0$.

Proof. [see 10]. Since $\|M(x)\|$ is a seminorm on E, $\sum_{k\geq 0} \|M(x^k)\|$ converges by Corollary 2, from which the conclusion follows. □

Let's now have a look at the polynomial identity, more details of which can be found in [18]. Assume A be an algebra over a field $\mathbb{F}$. By polynomial identity over the field $\mathbb{F}$

on a subset $Q \subseteq A$ we mean the polynomial in non-commuting indeterminates $x_1, ...x_k$, such that all coefficients are in $\mathbb{F}$. Moreover, the coefficients are not all equal to zero and the polynomial evaluates to zero when the indeterminates are substituted with variables $x_1, ..., x_k \in Q$.

Assume now that V be a vector space and Θ be its Hamel basis. Define the Hamel basis for $V^{\otimes k}$ by the set of pure tensors $\Theta^{\otimes k} = \{v_1...v_k | v_j \in \Theta, 1 \leq j \leq k\}$. Thereby for every element x from $V^{\otimes k}$ define Θ_x as a finite set of vectors in Θ, such that it is included in the representations of vector x in the Hamel basis $\Theta^{\otimes k}$. Denote by f_x^Θ the canonical formal non-commuting polynomial in indeterminates Θ_x, which is related to vector x. Using the fact that the set Θ_x is finite, we can state the following lemma from [10], which follows from the Hahn-Banach theorem.

Lemma 6. Let V be a locally convex space with Hamel basis Θ, A an algebra which is a topological vector space, and $Q \subseteq A$ a subset. Let $k \geq 0$ and $x \in V^{\otimes k}$.The following two assertions are equivalent.

(i) f_x^Θ is not a polynomial identity over $\mathbb{R}$ on Q

(ii) There exists a continuous linear map $M : V \mapsto span(Q)$ such that $M(x)$ is non-zero and $M(v)$ is in Q for all $v \in \Theta_x$.

Note, that if one does not wonder about topological considerations, a similar assertion would hold if we replace $\mathbb{R}$ by a field $\mathbb{F}$, V by a vector space over $\mathbb{V}$, A by an $\mathbb{F}$-algebra and in the previous lemma drop the continuity assumption in (ii).

3.3.2 Polynomial identities

In this section we consider polynomial identities in unitary Lie algebras in order to understand how representations of the family $\mathcal{A}(\mathbb{R}^d)$ separate points of space $E(\mathbb{R}^d)$. For this we use the results of Lemmas 5 and 6.

Let us denote a symplectic involution on $M_{2m}(\mathbb{C})$ by $\cdot^s$, where $m \geq 1$ is an integer, which also represents an involution of the first kind, namely $(\alpha a)^* = \alpha a^*$ for all $\alpha \in \mathbb{C}$ and $a \in M$, one can see in more details in [17]. Recall the Lie algebra associated with compact symplectic group $Sp(m)$ and denote it by $\mathfrak{sp}(m) = \{u \in \mathfrak{u}(\mathbb{C}^{2m}) | u^s + u = 0\}$. Consider a general linear Lie algebra $\mathfrak{gl}(\mathbb{C}^{2m})$ and closely related complex subalgebra $\mathfrak{sp}(m, \mathbb{C}) = \{u \in M_{2m}(\mathbb{C}) | u^s + u = 0\}$, which is also a complexification of $\mathfrak{sp}(m)$. Further on, we will be concerned about Lie algebras $\mathfrak{sp}(m)$ and $\mathfrak{sp}(m, \mathbb{C})$.

Based on Giambruno and Valenti [17] results and remark that $\mathfrak{sp}(m, \mathbb{C}) = \{u - u^s | u \in M_{2m}(\mathbb{C})\}$, we can state the following theorem.

Theorem 3 ([17, Theorem 6])**.** Let $m \geq 2$ and $f(x_1, ..., x_k)$ a polynomial identity over $\mathbb{C}$ on $\mathfrak{sp}(m, \mathbb{C}) \subset M_{2m}(\mathbb{C})$. Then $deg(f) > 3m$.

Proof. [see 17] □

The following lemma we can formulate by generalising Theorem 1.3.2 of [18], the proof of which one finds in [10].

Lemma 7. Let $\mathbb{F}$ be an infinite field, A be an algebra over the field $\mathbb{F}$ and Q a linear subspace of A. If f is a polynomial identity over $\mathbb{F}$ on Q, then every multi-homogeneous component of f is a polynomial identity over $\mathbb{F}$ on Q.

Note that if we have a multi-homogeneous polynomial identity over $\mathbb{C}$ and especially over $\mathbb{R}$ on $\mathfrak{sp}(m) \in M_{2m}(\mathbb{C})$, then it is also a polynomial identity over $\mathbb{C}$ on $sp(\mathfrak{m}, \mathbb{C})$. Hence, using the results of the previous Theorem 3 and Lemma 7 we obtain that any polynomial identity f over $\mathbb{R}$ on $\mathfrak{sp}(m)$ for $m \geq 2$ has degree of each of its multi-homogeneous components greater than or equal to $3m$. Then, using the conclusions from Lemmas 5 and 6, we can formulate the following theorem.

Theorem 4. [see proof in 10] Let $x \in E(\mathbb{R}^d)$ such that $x^k \neq 0$ for some $k \geq 0$. Then for any integer $m \geq \max\{2, k/3\}$ there exists $M \in \mathbf{L}(\mathbb{R}^d, \mathfrak{sp}(m))$ such that $M(x) \neq 0$. In particular, $\mathcal{A}(\mathbb{R}^d)$ separates the points of $E(\mathbb{R}^d)$.

In the previous theorem we were required to use a finite dimensional space $V = \mathbb{R}^d$ in order to have the following equality $V^{\otimes k} = V^{\hat{\otimes} k}$. The result of this theorem could be extended to an infinite dimensional space V if we could find an alternative to Lemma 6 for $x \in V^{\hat{\otimes} k}$, or an equivalent result of Theorem 3 corresponding to a series of polynomials with finite degree, but with an unbounded number of indeterminants.

Corollary 5. The group $U(\mathbb{R}^d)$ is maximally almost periodic.

In a following remark we show that topological groups $G(\mathbb{R}^d)$ and $U(\mathbb{R}^d))$ are not locally compact for $d \geq 2$.

Remark 4. Assume we have a field $V = \mathbb{R}^d$ and $L(V)$ be the smallest Lie algebra in $T(V)$ which contains V. With Theorem 1.4 from [28] we obtain that for any continuous linear map $l \in L(V)$ we have a coproduct in the following form $\Delta(l) = 1 \otimes l + l \otimes 1$, which leads by computation to $exp(l) \in G$.

Let us take two linear independent vectors u and v from V and denote their span by $W = span(u, v)$. Note that for $k \geq 1$ the space $L(W)$ has an element in $W^{\otimes k}$ which is nonzero. Recall Proposition 2 about the fundamentality of the family of seminorms Ψ, then we can construct a sequence $(l_n)_{n\geq 1} \in B$ for any neighborhood of zero B of $L(W)$ such that $\gamma(exp(l_i) - exp(l_j)) \geq 1$ for all $i \neq j$ and some seminorms on E. Moreover, from the theorem [2, Theorem 3] we have that $exp : L(W) \to G$ is continuous, as a result of which we obtain that there is no such a neighborhood of the identity in G that would lie in a sequentially compact set.

Together with Corollary 1 we have that, in case the space V is metrizable and separable, E is a Polish space, moreover G is also Polish, as $G \subset E$ and it is closed. From Lemma 4 and Theorem 4 we get the following.

Corollary 6. For Borel probability measures μ and ν on $G(\mathbb{R}^d)$, it holds that $\mu = \nu$ if and only if $\mu(f) = \nu(f)$ for all $f \in \mathcal{C}(\mathbb{R}^d)$, or equivalently, $\mu(M) = \nu(M)$ for all $M \in \mathcal{A}(\mathbb{R}^d)$.

As a result in this section we discussed representations of space $E(V)$. One of the essential results was Theorem 4. There we have explicitly characterised the set of representations $E(\mathbb{R}^d)$ that preserves unitary elements and separates points. Moreover, from the last corollary we can conclude that a characteristic function for $G(\mathbb{R}^d)$-valued random variable can be defined, but that is not the focus of this work.

3.4 Expected signatures and laws

In this section we discuss the expected signature of G-valued random variables, the conditions under which the expected signatures of $G(\mathbb{R}^d)$-valued random variables determine its laws and also in Theorem 6 we show how this criterion can be verified, even without explicit knowledge of the expected signature. Moreover, we note that the signature of a rough path can be represented as an element of the space $E(V)$.

Firstly, note that to denote *simplex* we will use the notation $\Delta_{[0,T]} = \{(s,t) : 0 \leq s \leq t \leq T\}$. Let denote *truncated tensor algebra* by $T^n(V) = \oplus_{0\leq k\leq n} V^{\hat{\otimes} k}$, where $V^{\otimes 0} = \mathbb{R}$ and $V^{\hat{\otimes} k} = V \otimes ... \otimes V$ (k times of V).

Definition 10. [see 10] Let us now by $\Omega_p(V)$ denote the *set of the rough paths* in $T^{([p])}(V)$ with p-variation. That is the multiplicative functionals $X : \Delta_{[0,T]} \to T^{[p]}$ with control ω, such that

- for all $0 \leq s \leq t \leq u \leq T$ we have

$$X^0_{(s,t)} = 1 \text{ and } X_{(s,t)} X_{(t,u)} = X_{(s,u)} \tag{3.16}$$

  and

- for some control ω one has

$$\sup_{0\leq k\leq [p]}((k/p)!\beta_p\|X^k_{(s,t)}\|)^{p/k}\leq \omega(s,t),\ \forall (s,t)\in\Delta_{[0,T]}, \tag{3.17}$$

where β_p is constant that only depends on p.

The definitions of multiplicative functionals and a control ω can be found in the Appendix 4.3.

Assume for some control function ω the map $X\in\Omega_p$ satisfaies condition (3.17). Thus using following theorem we obtain one of the essential results, that follows from the extension theorem [see 22, Theorem 3.7].

Theorem 5. [see 8, Theorem 3.1.3] Let $X\in\Omega_p$. Then for all $n\geq [p]$ there exists a unique map $S_n(x):\Delta_{[0,T]}\to T^n$ such that (3.16) and (3.17) remain true for the same control ω and with $sup_{1\leq k\leq [p]}$ replaced by $sup_{1\leq k\leq n}$ in (3.17).

Thus, the theorem shows, that for every multiplicative functional of degree n and finite p-variation we have an unique extension to a multiplicative functional of arbitraty high degree.

From the factorial decay in (3.17) we can conclude that a map $S(X):\Delta_{[0,T]}\to P=\prod_{k\geq 0}V^{\hat{\otimes}k}$ takes values in the space E for any $p\geq 1$ (see Corollary 2).

Let now define the extension of the space of rough paths Ω_p.

Definition 11. The space ΩE_p is defined as a set of maps $X:\Delta_{[0,T]}\to E$ which satisfy (3.16) and (3.17) with $\sup_{0\leq k\leq [p]}$ replaced by $\sup_{0\leq k}$ in (3.17).

Thus we have that the map S from Ω_p to ΩE_p is bijective, such that the $[p]$-th level truncation $(X^0_{s,t},X^1_{s,t},...)\to(X^0_{s,t},X^1_{s,t},...,X^{[p]}_{s,t})$ provide the inverse. Moreover, the $X^k_{s,t}$ components could be named as iterated integrals. Further by $S(X)_{0,T}\in E$ we can denote the signature of a rough path $X\in\Omega_p$. Note that for $1\leq p<2$, $S(x)_{0,T}$ is a sequence of iterated integrals (in terms of Young theory) of the path $X_{0,.}:[0,T]\to V$.

As was mentioned in [10, Corollary 5.5], the signature map $\mathbb{I}^p_{[0,T]}:\Omega_p\to E, X\to S(X)_{0,T}$, i.e. the map that sends the rough path to its signature, is continuous. Hence, we have that for the Ω_p-valued rough path X its signature $S(X)_{0,T}$ is a well-defined E-valued random variable.

Let's recall the Corollary 4, that is, assume we have normed space V and G-valued random variable X, such that its expected signature exists and has $r_2(X)=\infty$, then $\mathbb{E}[X]$ exists and $\mathbf{Sig}(X)=\mathbb{E}[X]$. Hence, for all $f\in E'$ and, in addition, for all $M\in\mathcal{A}$, the expected signature $\mathbf{Sig}(X)$ determine the expected value of $f(X)$, $\mathbb{E}[f(X)]$. Then, applying the results of Corollary 6 on the uniqueness of the probability measure, we can write the following proposition.

Proposition 5. [see 10] Let X and Y be $G(\mathbb{R}^d)$-valued random variables such that $\mathbf{Sig}(X) = \mathbf{Sig}(Y)$ and $\mathbf{Sig}(X) \in E$, i.e., $\mathbf{Sig}(X)$ has an infinite radius of convergence. Then $X \stackrel{D}{=} Y$.

Example 9. [see 10, Example 6.2] We apply Proposition 5 to the Lévy–Khintchine formula established in [15]. Recall that every Lévy process in $\mathbb{R}^d$ admits a lift to a $G\Omega_p(\mathbb{R}^d)$-valued random variable X for any $p > 2$ by adding appropriate adjustments for jumps [see 30, Section 2]. Let (a, b, K) denote the triplet of the Lévy process. It follows from [15, Section 9.1] that $\mathbf{Sig}(X)$ exists (as an element of $P(\mathbb{R}^d) = \prod_{k\geq 0}(\mathbb{R}^d)^{\otimes k}$) whenever the Lévy measure K has finite moments of all orders. Furthermore, $\mathbf{Sig}(X) \in E$ exactly when

$$\int_{\mathbb{R}^d} (e^{\lambda\|y\|} - 1 - \lambda \mathbf{1}_{\|y\|\leq 1}\|y\|)K(dy) < \infty \text{ for all } \lambda > 0. \tag{3.18}$$

It follows by Proposition 5 that whenever (3.18) is satisfied, $S(\mathbf{X})_{0,T}$ is uniquely determined as a $G(\mathbb{R}^d)$-valued random variable by its expected signature.

Now we determine a theorem that shows the conditions which are necessary to make the radius of convergence $r_1(X) > 0$ or $r_1(X) = \infty$ without explicit knowledge of $\mathbf{Sig}(X)$. Recall that radiuses of convergence have the following relations $r_1(X) \leq r_2(X) = R(\mathbf{Sig}(X))$. In this way, we set the conditions under which the radius of convergence r_2 will be equal to infinity, that is, in the case $r_1 = \infty$.

Let A be an topological algebra, B is its subset and $n \geq 1$. Define $B^n = \{x_1...x_n|\ x_1, ..., x_n \in B\}$ and $B(x) = \inf\{n \geq 1 | x \in B^n\}$ for any $x \in A$ (assuming $B(x) = \infty$ if $x \notin B^n$ for all $n \geq 1$). Moreover, for measurable set $B \subset A$ and random variable $X \in A$ we have that $B(X)$ is well-defined random variable in $\{1, 2, ...\} \cup \{\infty\}$.

Theorem 6. [see 10] Let V be a normed space and X an E-valued random variable. Suppose there exists a bounded, measurable set $B \subset E$ such that $B(X)$ has an exponential tail, i.e., $\mathbb{E}[e^{\lambda B(X)}] < \infty$ for some $\lambda > 0$. Then $r_1(X) > 0$. If moreover $\mathbb{E}[e^{\lambda B(X)}] < \infty$ for all $\lambda > 0$, then $r_1(X) = \infty$.

Proof. Equip E with the projective extension of the norm on V. For any $r > 0$ and $\lambda > 0$ such that $\sup_{x\in B}\|\delta_r(x)\| < e^{\lambda}$, it holds that

$$\sum_{k\geq 0} r^k \mathbb{E}[\|X^k\|] = \mathbb{E}[\|\delta_r(X)\|] \leq \mathbb{E}[e^{\lambda B(X)}], \tag{3.19}$$

where the inequality follows from the fact that $\delta_r(X) = \delta_r(X_1)...\delta_r(X_{B(X)})$ for some $X_1, ..., X_{B(X)} \in B$.

Suppose first that $\mathbb{E}[e^{\lambda B(X)}] < \infty$ for all $\lambda > 0$. For any $r > 0$ let $\lambda > 0$ be sufficiently large such that $\sup_{B(X)} \|\delta_r(x)\| < e^{\lambda}$. Then (3.19) implies that $r_1(X) \geq r$, and thus $r_1(X) = \infty$.

Suppose now that $\mathbb{E}[e^{\lambda B(X)}] < \infty$ for some $\lambda > 0$. By Proposition 2.10 from [10], the functions δ_r converge strongly to δ_0 as $r \to 0$ and, in particular, uniformly on B. Thus there exists $r > 0$ such that $\sup_{x \in B} \|\delta_r(x)\| < e^{\lambda}$. Then (3.19) implies that $r_1(X) \geq r > 0$ as desired.

$\square$

Chapter 4

Implementation

To understand how the implementation of signatures works in practice, we decided to simulate the calculation of signatures and expected signatures based on securities prices. We have used Heston's pricing model to calculate securities prices. Since it is one of the most popular stochastic volatility models for derivatives pricing. So in this chapter we will look at the Heston model, examples of security price simulations using the Monte Carlo method, calculation of signatures and expected signatures using the Python "Signatory" library, discuss the results and draw conclusions. This chapter is based on [16], [19], [25], [20]...

4.1 Heston Model

Let us first consider two stochastic differential equations, which corresponds to the stock price S and its variance v respectively:

$$dS_t = \mu_t S_t dt + \sqrt{v_t} S_t dW_t^S,\ S_0 \geq 0, \tag{4.1}$$

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t} dW_t^v,\ \sigma_0^2 = v_0 = 0, \tag{4.2}$$

with

$$dW_t^S dW_t^v = \rho dt, \tag{4.3}$$

where W_t^S and W_t^v are two standard Brownian motions, rho is its correlation. Moreover, κ, θ, σ, ρ, v_0 are parameters of the model, namely:

- Correlation parameter: ρ
- Initial variance: v_0

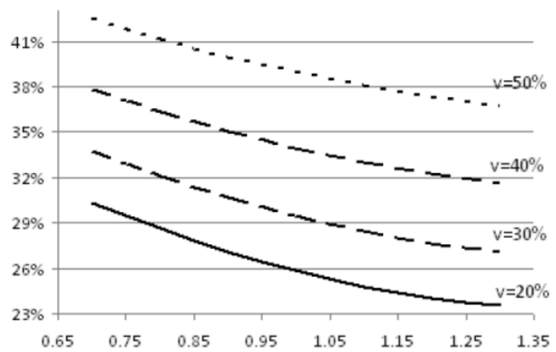
- Mean reversion rate: κ
- Volatility of variance: σ
- Long run variance: θ.

Such a model we will call the Heston model, which was first presented by Heston in [19]. The Heston model considers asset volatility as a stochastic process, which means that it can vary randomly over time. This differs from traditional models, which assume that volatility is constant (such as the Black-Scholes model). Heston's model is often used to evaluate options and other derivatives, as well as for portfolio risk analysis.

Since such a model depends on the initial parameters that we mentioned earlier, it is important to understand the meaning of these parameters and how they affect the model in general.

Let us start with **initial variance** v_0. Adjusting the initial level of variance can impact the height of the volatility smile curve rather than its shape. For example, increasing the initial volatility level, denoted as $\sqrt{v_0}$, causes the implied volatility smile to move upwards (as shown in Figure 4.1 on the left). This effect can be intuitively understood and does not require further explanation.

Actually, both the **long-term variance** θ and the initial variance v_0 have a similar effect on the smile of the implied volatility. This can be seen in Figure 4.1 on the right, which shows the effect of changing the long-run variance.



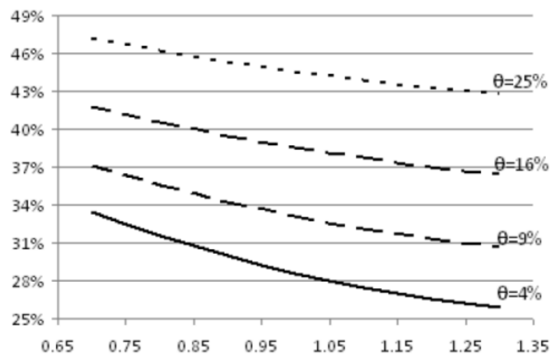


Figure 4.1: Changes in the volatility smile from different values of the initial variance $v = \sqrt{v_0}$ (left) and long-run variance θ (right). Source of pictures: [16].

The **mean reversion rate** κ is a measure of the degree of volatility clustering. This means that large movements are often followed by large movements, while small movements are more likely to be followed by small movements. The mean reversion parameter affects the shape of the curve reflecting this phenomenon. A higher mean reversion parameter will result in a flatter curve, while a lower mean reversion parameter will result in a curve with more curvature, see figure 4.2, similar to the effect of increased variance volatility. This can be seen lower in figure 4.5.

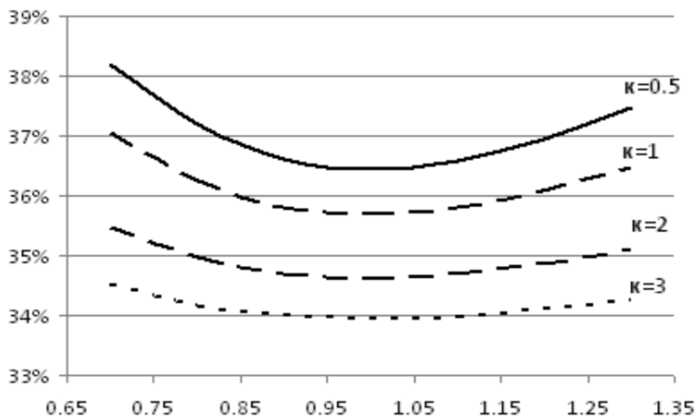


Figure 4.2: Changes in the implied volatility smile from different values of the mean reversion κ. Source of pictures: [16].

The **correlation** coefficient, denoted by ρ, measures the relationship between the log-returns and volatility of an asset. When ρ is positive, an increase in the asset's price or return is accompanied by an increase in volatility, which leads to a distribution with a heavy right tail. On the other hand, when ρ is negative, an increase in the asset's price or return is associated with a decrease in volatility, resulting in a distribution with a heavy left tail. The correlation, therefore, determines the skewness of the distribution. Figure 4.3 illustrates how different values of ρ affect the skewness of the density function.

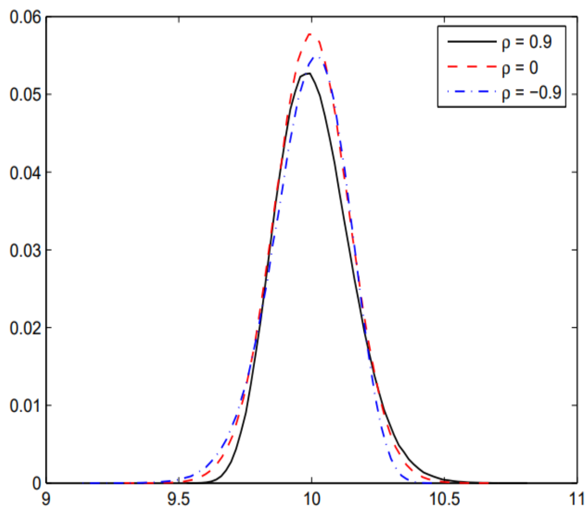


Figure 4.3: Changes in the distribution from different values of the correlation coefficient ρ. Source of pictures: [25].

The skewness of a distribution also affects the shape of the implied volatility surface, and thus the correlation coefficient ρ also plays a role in this. By changing the value of ρ, the model is able to generate a range of different volatility surfaces, which allows it to better capture the varying levels of volatility at different strikes that are not accounted for in the Black-Scholes model. Figure 4.7 illustrate the impact of different values of ρ on the implied volatility surface.

The **volatility of variance**, represented by the symbol σ, has an impact on the kurtosis, or peak, of a distribution. When σ is equal to zero, the volatility is deterministic and the resulting log-returns will be normally distributed. However, if σ is increased, it will only increase the kurtosis and create heavy tails on either side of the distribution. This relationship is illustrated in Figure 4.4.

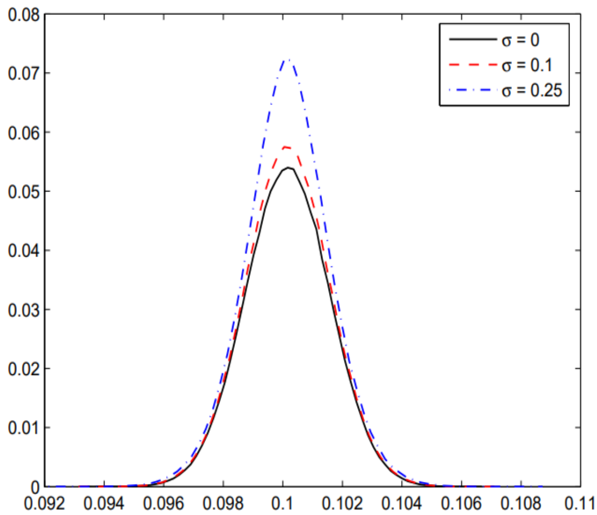


Figure 4.4: Changes in the distribution from different values of the volatility of variance σ. Source of pictures: [25].

Moreover, the kurtosis of the distribution also affects the implied volatility, as shown in Figures 4.5 and 4.6. Higher values of σ lead to a more prominent skew or smile, which makes sense in relation to the leverage effect. Higher values of σ indicate more volatile volatility, meaning that the market is more prone to extreme movements.

4.2 Libraries

4.3 Results and conclusions

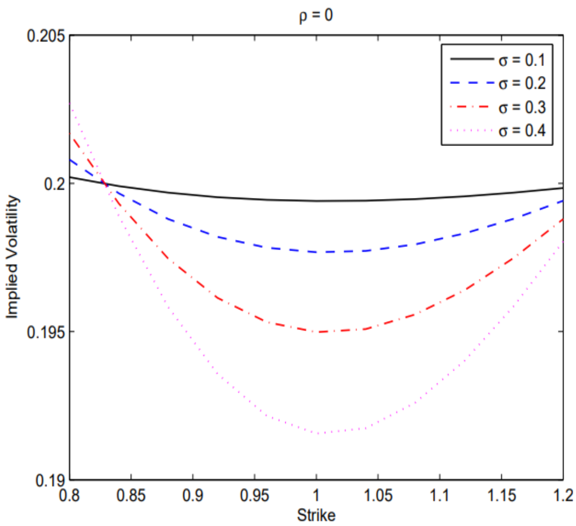


Figure 4.5: Implied volatility curve, $\rho = 0$, $\kappa = 2$, $\theta = 0.04$, $\sigma = 0.1$, $V_0 = 0.04$, $r = 0.01$, $S_0 = 1$, strikes: $0.8 - 1.2$, maturities : $0.5 - 3$ years. Source of pictures: [25].

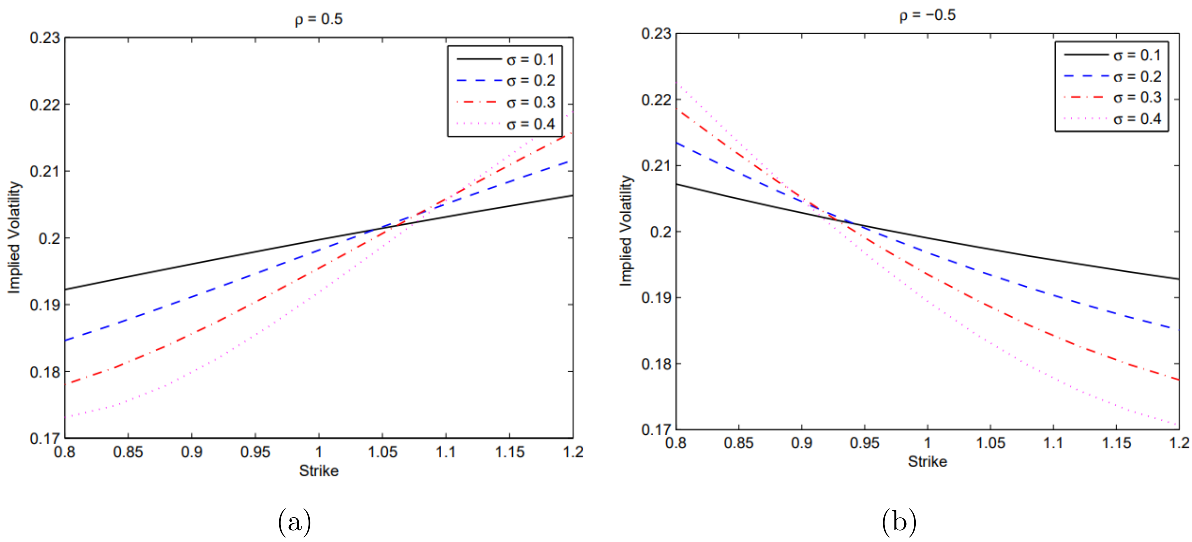
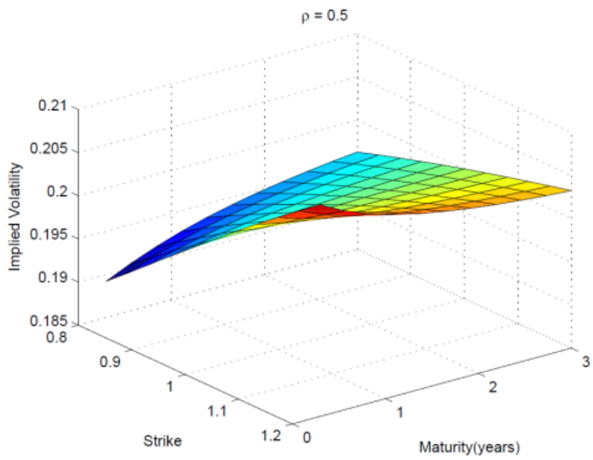
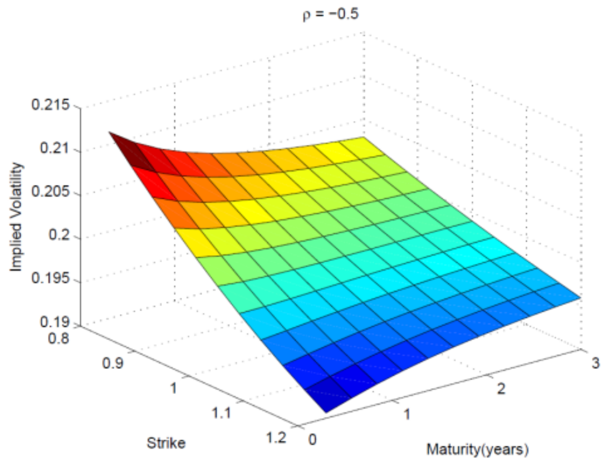


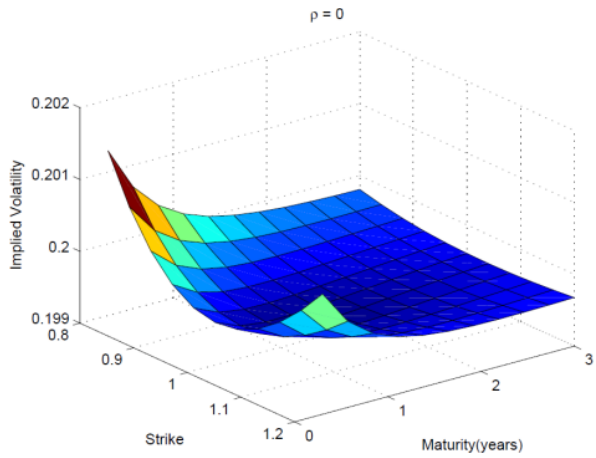
Figure 4.6: Implied volatility curves, $\rho = 0.5$(left), $\rho = -0.5$(right), $\kappa = 2$, $\theta = 0.04$, $\sigma = 0.1$, $V_0 = 0.04$, $r = 0.01$, $S_0 = 1$, strikes: $0.8 - 1.2$, maturities : $0.5 - 3$ years. Source of pictures: [25].



(a)



(b)



(c)

Figure 4.7: Implied volatility surface, $\rho = 0.5$(left), $\rho = -0.5$(right), $\rho = 0.5$(bottom) , $\kappa = 2$, $\theta = 0.04$, $\sigma = 0.1$, $V_0 = 0.04$, $r = 0.01$, $S_0 = 1$, strikes: $0.8 - 1.2$, maturities : $0.5 - 3$ years. Source of pictures: [25].

Appendix A

Definitions, theorems, lemmas

In this appendix we will state definitions, theorems and propositions from functional analysis, which are necessary for a better understanding of this paper, but which are not the main focus of this thesis.

A.1 Basic definitions

By *seminorms* on the vector space V we define continuous real-valued functions $\gamma : V \to \mathbb{R}$ which satisfy following conditions:

- Triangle inequality: for all $x, y \in \mathbb{R}$, $\gamma(x+y) \leq \gamma(x) + \gamma(y)$
- Non-negativity: $\gamma(x) \geq 0$ for all $x \in \mathbb{R}$
- Positive homogeneity: $\gamma(hx) = |h|\gamma(x)$ for all $x \in \mathbb{R}$ and scalar h

Moreover, if a seminorm γ satisfies the condition of positive definiteness, i.e. $\gamma(x) = 0$ implies $x = 0$, then γ is called a *norm*. Therefore, we will define a *normed space* as a pair (V, γ).

Definition 12. We define a *topological vector space* as a vector space endowed with a topology with respect to which addition and scalar multiplication operations are continuous.

Definition 13. [see 24]. By a *topological algebra* we mean an algebra E which is a topological vector space in such a way that the ring multiplication in E is separately continuous (i.e., continuous in each one of the two variables, the latter operation being a map $E \times E$ into E).

Let by N we denote a subset of topological space E. Then N is *dense* in E if its closure $\bar{N} = E$.

Definition 14. The topological space E we call *separable* if it has countable dense subset $N \subset E$.

Definition 15. We call the topological space E *metrizable* if its topology is induced by some metric on E.

Definition 16. [see 8] A *locally m-convex algebra* A is an algebra equipped with a locally convex topology for which there exists a fundamental family of defining seminorms Ψ such that for all $\gamma \in \Psi$

$$\gamma(xy) \leq \gamma(x)\gamma(y) \ \forall x, y \in A. \tag{A.1}$$

When a seminorm γ satisfies (A.1), we say it is *submultiplicative*. An locally m-convex algebra A is called *normed* if there exists a single submultiplicative norm $\| \cdot \|$ which defi'nes the topology of A.

Definition 17. [1] Hermitian operator is a a linear operator A on a Hilbert space H with a dense domain of definition $D(A)$ and such that $<Ax, y> = <x, Ay>$ for any $x, y \in D(A)$. An anti-hermitian operator is respectively an operator A that satisfies $<Ax, y> = -<x, Ay>$.

Definition 18. [3] A finite Borel measure μ on S we call tight if for every $\epsilon > 0$ there exists a compact set $M \subset X$ such that $\mu(S\ M) < \epsilon$, or, equivalently, $\mu(M) \geq \mu(S) - \epsilon$.

Definition 19. A vector space V over a field F, with an operation $V \times V \to V$, denoted $(x, y) \to [xy]$ and called the *bracket* or *commutator* of x and y, is called a *Lie algebra* over F if the following axioms are satisfied:

1. The bracket operation is bilinear.
2. $[xx] = 0$ for all x in V.
3. $[x[yz]] + [y[zx]] + [z[xy]] = 0$, where $x, y, z \in V$.

Definition 20. The norm l_1 of a vector $v \in V$ is denoted by $\|v\|$ and is defined as the sum of the absolute values of its components:

$$\|v\| = \sum_{i=1}^{n} |v_i| \tag{A.2}$$

Theorem 7. [see 5, Stone-Weierstrass theorem] If X is a compact Hausdorff space, then a subalgebra $\mathcal{A}$ of $C(X)$, which contains $f = 1$ and separates points, is dense in $C(X)$.

Definition 21. [11] We say that f is *multi-homogenous* of multidegree $(d_1, ..., d_m)$ if $f = f(x_1, ..., x_m) \in \mathbb{F}(Q)^{(d_1,...,d_m)}$. Moreover, we can write

$$f = \sum_{d_1 \geq 0, \ldots, d_m \geq 0} f^{(d_1,\ldots,d_m)}, \tag{A.3}$$

where $f^{(d_1,\ldots,d_m)} \in \mathbb{F}(Q)^{(d_1,\ldots,d_m)}$. The polynomials $f^{(d_1,\ldots,d_m)}$ are called the *multihomogenous components* of f.

Definition 22. [26] We call a group X maximally almost periodic if the group X^+ is Hausdorff, where $X^+ := (X, \tau^+)$ is the group X endowed with the Bohr topology τ^+ induced from the Bohr compactification bX.

Definition 23. [31] The subset F of the Banach space V is called sequentially compact if every sequence in F has a subsequence converging to a point of F.

Definition 24. [21] Define a *control function* as a continuous non-negative function ω on simplex $\Delta_{[0,T]} = \{(s,t) : 0 \leq s \leq t \leq T\}$ which is super-additive, such that $\omega(s,t) + \omega(t,u) \leq \omega(s,u)$, $\forall s \leq t \leq u \in I$ and $\omega(t,t) = 0$ for all $t \in I$.

Definition 25. [21] A continuous map X from the simplex $\Delta_{[0,T]}$ into a truncated tensor algebra $T^{(n)}(V)$, and written as

$$X_{s,t} = (X^0_{s,t}, X^1_{s,t}, ..., X^n_{s,t}),\ with\ X^k_{s,t} \in V^{\otimes k},\ for\ any\ (s,t) \in \Delta_{[0,T]}, \tag{A.4}$$

is called a *multiplicative functional* of degree n ($n \in \mathbb{N}, n \geq 1$) if $X^0_{s,t} \equiv 1$ (for all $(s,t) \in \Delta_{[0,T]}$) and

$$X_{s,t} \otimes X_{t,u} = X_{s,u},\ \forall (s,t), (t,u) \in \Delta_{[0,T]}, \tag{A.5}$$

where the tensor product $\otimes$ is taken in $T^{(n)}(V)$.

A.2 Second Part

List of Figures

Bibliography

[1] Naum Ilich Akhiezer, Naum Ilich Ahiezer, and Izrail Markovich Glazman. *Theory of linear operators in Hilbert space*. 9. Pitman publishing, 1981.

[2] Richard Arens. "The space L^ω and convex topological rings". In: *Bulletin of the American Mathematical Society* 52.10 (1946), pp. 931–935.

[3] P Billingsley. "Convergence of probability measures Wiley and Sons". In: *New York* (1968).

[4] VI Bogachev. "Measure theory vol. II Springer-Verlag". In: *Berlin Go to reference in article Crossref MathSciNet ZentralBlatt Math* (2007).

[5] Peter Borwein and Tamas Erdelyi. *Polynomials and polynomial inequalities*. Vol. 161. Springer Science & Business Media, 1995.

[6] Kuo-Tsai Chen. "Integration of paths–A faithful representation of paths by noncommutative formal power series". In: *Transactions of the American Mathematical Society* 89.2 (1958).

[7] Kuo-Tsai Chen. "ITERATED INTEGRALS AND EXPONENTIAL HOMOMORPHISMSŤ". In: *Collected Papers of KT Chen* (2001), p. 54.

[8] Ilya Chevyrev. "Characteristic functions of path signatures and applications". PhD thesis. University of Oxford, 2015.

[9] Ilya Chevyrev and Andrey Kormilitzin. "A primer on the signature method in machine learning". In: *arXiv preprint arXiv:1603.03788* (2016).

[10] Ilya Chevyrev and Terry Lyons. "Characteristic functions of measures on geometric rough paths". In: *The Annals of Probability* 44.6 (2016), pp. 4049–4082.

[11] L Cioletti, JA Freitas, and DJ Goncalves. "Polynomial identities of Banach algebras". In: *Beitrage zur Algebra und Geometrie/Contributions to Algebra and Geometry* 58.1 (2017), pp. 189–194.

[12] Joachim Cuntz. "Excision in periodic cyclic theory for topological algebras". In: *Fields Inst. Comm* 17 (1997), pp. 43–53.

[13] Sean Dineen. *Complex analysis in locally convex spaces*. English. North-Holland Pub. Co. ; Sole distributors for the U.S.A. and Canada, Elsevier North-Holland Amsterdam ; New York : New York, N.Y, 1981, xiii, 492 p.

[14] Adeline Fermanian. "Learning time-dependent data with the signature transform". PhD thesis. Sorbonne Université, 2021.

[15] Peter K Friz and Atul Shekhar. "General rough integration, Lévy rough paths and a Lévy–Kintchine-type formula". In: *The Annals of Probability* 45.4 (2017), pp. 2707–2765.

[16] Pierre Gauthier and Pierre-Yves Henri Rivaille. "Fitting the smile, smart parameters for SABR and Heston". In: *Smart Parameters for SABR and Heston (October 30, 2009)* (2009).

[17] Antonio Giambruno and Angela Valenti. "On minimal $*$-identities of matrices". In: *Linear and Multilinear Algebra* 39.4 (1995), pp. 309–323.

[18] Antonio Giambruno, Mikhail Zaicev, and Michail V Zaicev. *Polynomial identities and asymptotic methods*. 122. American Mathematical Soc., 2005.

[19] Steven L Heston. "A closed-form solution for options with stochastic volatility with applications to bond and currency options". In: *The review of financial studies* 6.2 (1993), pp. 327–343.

[20] Patrick Kidger and Terry Lyons. "Signatory: differentiable computations of the signature and logsignature transforms, on both CPU and GPU". In: *arXiv preprint arXiv:2001.00706* (2020).

[21] Terry Lyons, Zhongmin Qian, et al. *System control and rough paths*. Oxford University Press, 2002.

[22] Terry J Lyons, Michael Caruana, and Thierry Lévy. *Differential equations driven by rough paths*. Springer, 2007.

[23] A Mallios. *Topological Algebras. Selected Topics, North-Holland Mathematics Studies*. 1986.

[24] Anastasios Mallios. *Topological algebras: Selected topics*. Elsevier, 2011.

[25] Nimalin Moodley. "The heston model: A practical approach with matlab code". In: *University of Witwatersrand* (2005).

[26] JV Neumann. "Almost periodic functions in a group. I". In: *Transactions of the American Mathematical Society* 36.3 (1934), pp. 445–492.

[27] Rimhak Ree. "Lie elements and an algebra associated with shuffles". In: *Annals of Mathematics* (1958), pp. 210–220.

[28] C. Reutenauer and P.D. de Mathematiques-Informatique Christophe Reutenauer. *Free Lie Algebras*. LMS monographs. Clarendon Press, 1993. ISBN: 9780198536796. URL: `https://books.google.de/books?id=cBvvAAAAMAAJ`.

[29] Christian Valqui. "Universal extension and excision for topological algebras". In: *K-theory* 22.1-2 (2001), pp. 145–160.

[30] David RE Williams. "Path-wise solutions of stochastic differential equations driven by Lévy processes". In: *Revista Matemática Iberoamericana* 17.2 (2001), pp. 295–329.

[31] Adriaan C Zaanen. *Introduction to operator theory in Riesz spaces*. Springer Science and Business Media, 2012.