



**HAL**  
open science

## Path-dependent processes from signatures

Eduardo Abi Jaber, Louis-Amand Gérard, Yuxing Huang

► **To cite this version:**

Eduardo Abi Jaber, Louis-Amand Gérard, Yuxing Huang. Path-dependent processes from signatures. 2025. <hal-04637402v2>

**HAL Id: hal-04637402**

**<https://hal.science/hal-04637402v2>**

Preprint submitted on 24 Nov 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Path-dependent processes from signatures

Eduardo Abi Jaber<sup>\*1</sup>, Louis-Amand Gérard<sup>†2,3</sup>, and Yuxing Huang<sup>‡1</sup>

<sup>1</sup>Ecole Polytechnique, CMAP

<sup>2</sup>Université Paris 1 Panthéon-Sorbonne, CES

<sup>3</sup>Gefip

November 4, 2025

## Abstract

We provide explicit series expansions to certain stochastic path-dependent integral equations in terms of the path signature of the time augmented driving Brownian motion. Our framework encompasses a large class of stochastic linear Volterra and delay equations and in particular the fractional Brownian motion with a Hurst index  $H \in (0, 1)$ . Our expressions allow to disentangle an infinite dimensional Markovian structure and open the door to straightforward and simple approximation schemes, that we illustrate numerically. A key application is the derivation of explicit series representations for both conditional and unconditional moments.

**Mathematics Subject Classification (2010):** 60L10, 60L70, 60H20, 60G22

**Keywords:** Path-signatures, Volterra processes, Stochastic Delay equations, fractional Brownian motion.

## 1 Introduction

We provide explicit solutions to certain stochastic path-dependent integral equations in the form

$$X_t = X_0 + \int_0^t b(t, s, (X_u)_{0 \leq u \leq s}) ds + \int_0^t \sigma(t, s, (X_u)_{0 \leq u \leq s}) dW_s,$$

for some linear coefficients  $b, \sigma$ , in terms of the path signature of the time-augmented driving Brownian motion  $(t, W_t)_{t \geq 0}$ .

We consider two non-Markovian specifications:

- **The Volterra case**, where the coefficients  $b, \sigma$  are of the form

$$g(t, s, (X_u)_{0 \leq u \leq s}) = K(t - s)(a + bX_s),$$

for some locally integrable deterministic kernel  $K$  and  $a, b \in \mathbb{R}$ . The Riemann-Liouville Fractional Brownian motion and Gaussian Volterra processes constitute a particular case.

- **The Delay case**, where the coefficients  $b, \sigma$  are of the form

$$g(t, s, (X_u)_{0 \leq u \leq s}) = a + bX_s + \int_0^s K(s - u)X_u du.$$

---

<sup>\*</sup>eduardo.abi-jaber@polytechnique.edu. The first author is grateful for the financial support from the Chaires FiME-FDD, Financial Risks, Deep Finance & Statistics and Machine Learning and systematic methods in finance at Ecole Polytechnique.

<sup>†</sup>louis-amand.gerard@etu.univ-paris1.fr

<sup>‡</sup>yuxing.math@gmail.com.

The authors would like to thank Xuyang Lin for fruitful discussions and the two anonymous referees for their insightful suggestions.

In both cases, the unique strong solution  $X$  is a non-anticipative measurable function of the whole path of the driving Brownian motion, that is

$$X_t = f(t, (W_u)_{0 \leq u \leq t}), \quad t \geq 0,$$

for some measurable functional  $f$ .

Our main contribution is to explicit this functional  $f$ . More precisely, we show that the solution  $X$  can be written in terms of an infinite linear combination of the signature process  $\widehat{\mathbb{W}}_t$  of the time extended Brownian motion  $\widehat{W}_t := (t, W_t)$  defined by the infinite sequence of iterated integrals in the Stratonovich sense:

$$\widehat{\mathbb{W}}_t = \left( 1, \binom{t}{W_t}, \left( \int_0^t W_s ds, \int_0^t s dW_s \right), \left( \int_0^t \int_0^s W_u du ds, \int_0^t \int_0^s u dW_u ds, \int_0^t \int_0^s W_s^2 ds, \int_0^t \int_0^s s^2 dW_s \right), \left( \int_0^t \int_0^s W_u du \circ dW_s, \int_0^t \int_0^s u dW_u \circ dW_s, \int_0^t \int_0^s \frac{W_t^3}{3!} \right), \dots \right),$$

that is

$$X_t = \langle \boldsymbol{\ell}_t, \widehat{\mathbb{W}}_t \rangle, \quad t \geq 0, \quad (1.1)$$

where the (possibly) time-dependent deterministic coefficients  $\boldsymbol{\ell}_t$ , given by a sequence of numbers indexed by words, are explicit.

To highlight our strategy, consider a generalized geometric Brownian motion process  $X$  solution to

$$X_t = x + \int_0^t (a + bX_s) ds + \int_0^t (\alpha + \beta X_s) \circ dW_s, \quad \text{with } x, a, b, \alpha, \beta \in \mathbb{R}, \quad (1.2)$$

where  $\circ$  denotes Stratonovich integration rule.

### Our recipe:

1. Translate the probabilistic equation into an algebraic linear equation on  $\boldsymbol{\ell}$ :

$$\boldsymbol{\ell} = x\boldsymbol{\emptyset} + (a + b\boldsymbol{\ell})\mathbf{1} + (\alpha + \beta\boldsymbol{\ell})\mathbf{2}, \quad (1.3)$$

where concatenation with the letters  $\mathbf{1}$  and  $\mathbf{2}$  play respectively the role of integration with respect to  $dt$  and  $dW_t$  (in the sense of Stratonovich).

2. Solve the linear algebraic equation (1.3) by writing

$$\boldsymbol{\ell} = \boldsymbol{p} + \boldsymbol{\ell}\boldsymbol{q}, \quad (1.4)$$

with

$$\boldsymbol{p} = (x\boldsymbol{\emptyset} + a\mathbf{1} + \alpha\mathbf{2}) \quad \text{and} \quad \boldsymbol{q} = b\mathbf{1} + \beta\mathbf{2},$$

to get

$$\boldsymbol{\ell} = \boldsymbol{p}(\boldsymbol{\emptyset} - \boldsymbol{q})^{-1} = \boldsymbol{p} \sum_{n \geq 0} \boldsymbol{q}^{\otimes n}.$$

3. Argue that the infinite series  $\langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_t \rangle$  is well-defined and solves the stochastic differential equation (1.2).

The recipe is simple yet powerful, as it remains robust across non-Markovian settings and for multi-dimensional Brownian motions, with the same algebraic equation (1.4) but with different  $\boldsymbol{p}$  and  $\boldsymbol{q}$ . To make the analysis rigorous, we proceed as follows. First, we study the existence of solutions of the algebraic equation (1.3) taking values in the extended tensor algebra together with some estimates for the solutions, see Proposition 2.2 and Theorem 3.9. Second, since  $\boldsymbol{\ell}$  can have infinitely many non-zero elements, a delicate analysis is needed to make

sense of the series expansions in (1.1), this is the object of Section 3.2. In particular, we derive a tractable criterion for the global convergence in the supremum norm which is new compared to the related literature on infinite series of signature elements, see for instance Ben Arous [9], Cuchiero, Svaluto-Ferro, and Teichmann [17]. The third step, a central ingredient for the derivation of our representation formulas in Section 4 is an Itô's formula for possibly infinite linear combinations of signature elements, see Theorem 3.3. Finally, in Section 3.4 we derive a tractable criterion based on the domination by a shuffle exponential of  $\mathbf{p}$  and  $\mathbf{q}$  that validates the aforementioned steps, i.e. that the linear functional is well-defined, converges against the signature and is an Itô process.

We provide two types of representations: (i) time-independent representations (1.1) for non-Markovian stochastic Volterra and delay equations with smooth enough kernels  $K$  where the coefficients  $\ell$  do not depend on time, see Theorems 4.2 and 4.4. Such  $\ell$  would in principal correspond to the coefficients of a Stochastic-Taylor expansion applied in a path-dependent setting, except that here we compute  $\ell$  algebraically and establish the convergence of the Taylor series. For more general kernels, we provide approximation results in Corollary 4.3; but also (ii) time-dependent representations (1.1) with  $\ell_t$  depending on time which allowed us to recover more general Gaussian processes in Theorem 4.5 with singular kernels. For instance, for an Ornstein-Uhlenbeck process, a time-dependent representation involves

$$\ell_t^{\text{OU}} = e^{\sqcup - \kappa(t\sigma - 1)} \mathbf{2},$$

which is the algebraic translation of the probabilistic expression

$$X_t = \int_0^t e^{-\kappa(t-s)} dW_s,$$

with  $e^{\sqcup}$  the shuffle exponential (2.7).

The representation (1.1) has at least two main features. First, it allows to disentangle an infinite dimensional Markovian structure in terms of the process  $\widehat{\mathbb{W}}_t$ . Second it opens the door to straightforward and simple approximation schemes by truncating  $\ell_t$ , numerical illustrations are provided in Section 4.4. A key application is the derivation of explicit series representations for both conditional and unconditional moments of Volterra and delay processes, which can be efficiently computed numerically without relying on Monte Carlo simulations, as shown in Section 5. In particular, the linear Volterra process is a polynomial Volterra process in the terminology of Abi Jaber et al. [5], where unconditional moments were derived in terms of deterministic integral equations, without any treatment of conditional moments or numerical methods. Our results provide an alternative explicit representation for both conditional and unconditional moments and, most importantly, offer the first numerical implementation for such process.

**Motivation.** The elegant framework of path signatures, as introduced by Chen [12] and extensively used in rough path theory [25, 36], not only captures the intricate dynamics of path-dependent processes but also facilitates their integration into various fields of applications. A notable attribute driving the recent prominence of signatures is their inherent linearization property, or equivalently the universal approximation property of linear functionals of the signature, and the fact that signatures characterize paths, see [15, 27]. These properties have found many practical applications in Machine Learning (Chevyrev and Kormilitzin [14], Fermanian [22], Perez Arribas, Goodwin, Geddes, Lyons, and Saunders [42]), Mathematical Finance (Abi Jaber and Gérard [3], Bayer, Hager, Riedel, and Schoenmakers [7], Bayer, Pelizzari, and Schoenmakers [8], Cartea, Perez Arribas, and Sanchez-Betancourt [11], Cuchiero, Gazzani, Möller, and Svaluto-Ferro [16], Dupire and Tissot-Daguette [20], Lyons, Nejad, and Arribas [38])...

Our motivation for studying exact representation formulae in the form (1.1) for path-dependent processes arises from the importance of non-Markovian stochastic modeling in various fields. Specifically, stochastic Volterra and delay equations capture intricate temporal dependencies in time series data. For such non-Markovian processes, the representation formulae (1.1) not only reveal the underlying Markovian structure in the whole signature process  $\widehat{\mathbb{W}}_t$  but also enable us to re-frame them within a rather universal framework of linear functionals of the signature, for which one can develop generic methods. As illustration, recent works by Abi Jaber and Gérard [3], Cuchiero, Svaluto-Ferro, and Teichmann [17] have demonstrated interest in infinite linear combinations of signature elements, unveiling a surprising infinite-dimensional affine structure underlying a

broad spectrum of stochastic models. Certain Fourier-Laplace transforms of linear functionals of the signature can be computed in terms of non-standard tensor valued infinite-dimensional Riccati equations and applied in practice for risk management purposes in mathematical finance for instance.

**Related literature on series and functional expansions.** Stochastic Taylor expansions of solutions to Markovian stochastic differential equations (SDEs) in terms of the signature of the time augmented driving Brownian motion  $(t, W_t)_{t \geq 0}$  have appeared in Kloeden et al. [33, Chapter 5] and have been used to develop a cubature formula on the Wiener space by Lyons [39]. Ben Arous [9] establishes convergence criterion for such stochastic Taylor series for Markovian SDEs with analytic coefficients; the so-called Chen–Fliess [23, 24] approximation that generalizes the classic stochastic Taylor expansion for path dependent functionals of Markovian SDEs is studied by Litterer and Oberhauser [35]. A functional Taylor expansion in terms of a finite series of higher order functional derivatives multiplied by iterated integrals is presented by Dupire and Tissot-Daguette [20], a nice overview on series and functional expansions is given there. Variation of constants formulae for linear forward and backward stochastic Volterra integral equations using certain sequence of iterated Volterra integrals, adapting Wiener–Itô chaos expansions, are derived by Hamaguchi [29]. Related finite order expansions also appeared in Bayer, Friz, Gassiat, Martin, and Stemper [6], Hu [32] and Neuenkirch, Nourdin, Rößler, and Tindel [40].

Our expansions can be seen as an infinite Stochastic-Taylor series expansion of the solution of path-dependent stochastic equations whose coefficients depend linearly on the whole trajectory. In particular, we believe that the coefficients of our time-independent expansions, obtained in ‘one shot’ by solving an algebraic equation, can be recovered from successive functional derivatives of the coefficients of the stochastic equation. None of the works cited above rely on such algebraic method. When it comes to studying the convergence of the stochastic Taylor series, our work is closest in spirit to Ben Arous [9] which operates in a Markovian setting. Just like Taylor expansions [33, 20], and unlike Volterra and Wiener expansions, the coefficients  $\ell$  that depend on the model parameters and the signature  $\widehat{W}_t$  of the driving noise are disentangled. Combined with the explicit knowledge of the coefficients, this allows for a straightforward and simple numerical implementation of non-Markovian processes.

Moreover, finite expansions of Volterra and delay processes have also appeared in the rough path literature, using tailor-made constructions of signature objects. We mention, for instance, Volterra-type signatures, where the iterated integrals involve weights coming from certain kernels, as developed in Deya and Tindel [18, 19], Harang and Tindel [31], Harang et al. [30], Bruned and Katsetiadis [10], as well as the discrete delay signature introduced in Neuenkirch, Nourdin, and Tindel [41]. Compared to these works, in the present paper we do not rely on such specialized signature constructions. Instead, we leverage the standard (time-augmented) signature, which offers several advantages: it provides a unified and versatile framework for expanding both Volterra and delay processes, it preserves the classical Chen identity and shuffle product, and it yields a Markovian lift of our inherently non-Markovian dynamics, making it particularly convenient for computations (e.g. conditional moments in Section 5). Beyond these considerations, we stress that our contribution goes significantly further than previous works that focus primarily on formal expansions. While existing literature, including Volterra-based approaches, typically produces finite stochastic Taylor expansions, we rigorously establish the convergence of the infinite signature expansion for the class of processes we consider.

**Outline.** Section 2 deals with some preliminaries on path-signatures. Section 3 is concerned with infinite linear combinations of signature elements and their properties. Section 4 collects our representations formulae together with a numerical illustration. Section 5 presents series expansions for the moments. The proofs are postponed to later sections and to the Appendices.

## 2 Preliminaries

In this section, we setup the framework for dealing with signatures of semimartingales. We not only collect some (well-known) properties of finite linear combinations of signature elements, refer also to the first sections in [7, 17, 38], but we also introduce and study the resolvent object in Section 2.2 which plays a crucial role in our main results.

## 2.1 Tensor algebra

Let  $d \in \mathbb{N}$  and denote by  $\otimes$  the tensor product over  $\mathbb{R}^d$ , e.g.  $(x \otimes y \otimes z)_{ijk} = x_i y_j z_k$ , for  $i, j, k = 1, \dots, d$ , for  $x, y, z \in \mathbb{R}^d$ . For  $n \geq 1$ , we denote by  $(\mathbb{R}^d)^{\otimes n}$  the space of tensors of order  $n$  and by  $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}$ . In the sequel, we will consider mathematical objects, path signatures, that live on the extended tensor algebra space  $T((\mathbb{R}^d))$  over  $\mathbb{R}^d$ , that is the space of (infinite) sequences of tensors defined by

$$T((\mathbb{R}^d)) := \{ \ell = (\ell^n)_{n=0}^\infty : \ell^n \in (\mathbb{R}^d)^{\otimes n} \}.$$

Similarly, for  $M \geq 0$ , we define the truncated tensor algebra  $T^M(\mathbb{R}^d)$  as the space of sequences of tensors of order at most  $M$  defined by

$$T^M(\mathbb{R}^d) := \{ \ell \in T((\mathbb{R}^d)) : \ell^n = 0, \quad \forall n > M \},$$

and the tensor algebra  $T(\mathbb{R}^d)$  as the space of all finite sequences of tensors defined by

$$T(\mathbb{R}^d) := \bigcup_{M \in \mathbb{N}} T^M(\mathbb{R}^d).$$

We clearly have  $T(\mathbb{R}^d) \subset T((\mathbb{R}^d))$ . For  $\ell = (\ell^n)_{n \in \mathbb{N}}, \varphi = (\varphi^n)_{n \in \mathbb{N}} \in T((\mathbb{R}^d))$  and  $\lambda \in \mathbb{R}$ , we define the following operations:

$$\ell + \varphi := (\ell^n + \varphi^n)_{n \in \mathbb{N}}, \quad \ell \otimes \varphi := \left( \sum_{k=0}^n \ell^k \otimes \varphi^{n-k} \right)_{n \in \mathbb{N}}, \quad \lambda \ell := (\lambda \ell^n)_{n \in \mathbb{N}}.$$

For the rest of the paper, we will use  $\ell \varphi$  and  $\ell \otimes \varphi$  interchangeably. These operations are closed on  $T^M(\mathbb{R}^d)$  and  $T(\mathbb{R}^d)$ .

**Important notations.** Let  $\{e_1, \dots, e_d\} \subset \mathbb{R}^d$  be the canonical basis of  $\mathbb{R}^d$  and  $A_d = \{\mathbf{1}, \mathbf{2}, \dots, \mathbf{d}\}$  be the corresponding alphabet. To ease reading, for  $i \in \{1, \dots, d\}$ , we write  $e_i$  as the blue letter  $\mathbf{i}$  and for  $n \geq 1$ ,  $i_1, \dots, i_n \in \{1, \dots, d\}$ , we write  $e_{i_1} \otimes \dots \otimes e_{i_n}$  as the concatenation of letters  $\mathbf{i}_1 \dots \mathbf{i}_n$ , that we call a word of length  $n$ . We note that the family  $(e_{i_1} \otimes \dots \otimes e_{i_n})_{(i_1, \dots, i_n) \in \{1, \dots, d\}^n}$  is a basis of  $(\mathbb{R}^d)^{\otimes n}$  that can be identified with the set of words of length  $n$  defined by

$$V_n := \{ \mathbf{i}_1 \dots \mathbf{i}_n : \mathbf{i}_k \in A_d \text{ for } k = 1, 2, \dots, n \}. \quad (2.1)$$

Moreover, we denote by  $\emptyset$  the empty word and we set  $V_0 = \{\emptyset\}$  which serves as a basis for  $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}$ . Equipped with the product topology,  $T((\mathbb{R}^d))$  admits  $V := \bigcup_{n \geq 0} V_n$  as a Schauder topological basis. In particular, every  $\ell \in T((\mathbb{R}^d))$  can be decomposed as

$$\ell = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbf{v}, \quad (2.2)$$

where  $\ell^{\mathbf{v}} \in \mathbb{R}$  is the real coefficient of  $\ell$  at coordinate  $\mathbf{v}$ . Representation (2.2) induces a natural linear homeomorphism from  $T((\mathbb{R}^d))$  to  $\mathbb{R}^{\mathbb{N}}$ , showing that it is a locally convex space. This representation will be frequently used in the paper. We stress again that in the sequel, every blue ‘word’  $\mathbf{v} \in V$  represents an element of the canonical basis of  $T((\mathbb{R}^d))$ , i.e. there exists  $n \geq 0$  such that  $\mathbf{v}$  is of the form  $\mathbf{v} = \mathbf{i}_1 \dots \mathbf{i}_n$ , which represents the element  $e_{i_1} \otimes \dots \otimes e_{i_n}$ . The concatenation  $\ell \mathbf{v}$  of an element  $\ell \in T((\mathbb{R}^d))$  and the word  $\mathbf{v} = \mathbf{i}_1 \dots \mathbf{i}_n$  means  $\ell \otimes e_{i_1} \otimes \dots \otimes e_{i_n}$ .

In addition to the decomposition (2.2) of elements  $\ell \in T((\mathbb{R}^d))$ , we introduce the projection  $\ell|_{\mathbf{u}} \in T((\mathbb{R}^d))$  as

$$\ell|_{\mathbf{u}} := \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{u}\mathbf{v}} \mathbf{v} \quad (2.3)$$

for all  $\mathbf{u} \in V$ . The projection plays an important role in the space of iterated integrals as it is closely linked to partial differentiation, in contrast with the concatenation that relates to integration. It will be used throughout the paper.

*Remark 2.1.* The projection allows us to decompose elements of the extended tensor algebra  $\ell \in T((\mathbb{R}^d))$  as

$$\ell = \ell^\emptyset \emptyset + \sum_{\mathbf{i} \in A_d} \ell|_{\mathbf{i}}.$$

For instance, for  $\ell = 4 \cdot \emptyset + 3 \cdot \mathbf{1} - 1 \cdot \mathbf{12} + 2 \cdot \mathbf{2212}$ , we have that  $\ell^\emptyset = 4$ ,  $\ell|_{\mathbf{1}} = 3 \cdot \emptyset$ ,  $\ell|_{\mathbf{2}} = -1 \cdot \mathbf{1} + 2 \cdot \mathbf{221}$  and  $\ell|_{\mathbf{3}} = 0$ .  $\blacksquare$

We define the bracket between  $\ell \in T(\mathbb{R}^d)$  and  $\mathbb{X} \in T((\mathbb{R}^d))$  by

$$\langle \ell, \mathbb{X} \rangle = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbb{X}^{\mathbf{v}}. \quad (2.4)$$

Notice that it is well-defined as  $\ell$  has finitely many non-zero terms. For  $\ell \in T((\mathbb{R}^d))$ , the series in (2.4) involves infinitely many terms and requires special care, this will be discussed in Section 3 below.

We will also consider another operation on the space of words  $V$ , the shuffle product. The shuffle product plays a crucial role for an integration by parts formula on the space of iterated integrals, see Lemma 2.5 below.

**Definition 2.1** (Shuffle product). *The shuffle product  $\sqcup\sqcup: V \times V \rightarrow T(\mathbb{R}^d)$  is defined inductively for all words  $\mathbf{v}$  and  $\mathbf{w}$  and all letters  $\mathbf{i}$  and  $\mathbf{j}$  in  $A_d$  by*

$$\begin{aligned} (\mathbf{vi}) \sqcup\sqcup (\mathbf{wj}) &= (\mathbf{v} \sqcup\sqcup (\mathbf{wj}))\mathbf{i} + ((\mathbf{vi}) \sqcup\sqcup \mathbf{w})\mathbf{j} \\ \mathbf{w} \sqcup\sqcup \emptyset &= \emptyset \sqcup\sqcup \mathbf{w} = \mathbf{w}. \end{aligned}$$

*With some abuse of notation, the shuffle product on  $T((\mathbb{R}^d))$  induced by the shuffle product on  $V$  will also be denoted by  $\sqcup\sqcup$ . The shuffle product is clearly commutative. See [26, 44] for more information on the shuffle product.*

The shuffle product corresponds to the shuffling of two decks of cards together, while keeping the order of each single deck as illustrated on the following example:  $\mathbf{1} \sqcup\sqcup \mathbf{23} = \mathbf{123} + \mathbf{213} + \mathbf{231}$ .

## 2.2 Resolvent and linear equation

For  $n \in \mathbb{N}$  and  $\ell \in T((\mathbb{R}^d))$ , we define the concatenation power of  $\ell$  by

$$\ell^{\otimes n} := \overbrace{\ell \otimes \ell \otimes \dots \otimes \ell}^{n \text{ times}},$$

with the convention  $\ell^{\otimes 0} = \emptyset$ .

For every  $\ell \in T((\mathbb{R}^d))$  such that  $\ell^\emptyset = 0$ , we define its *resolvent* by

$$(\emptyset - \ell)^{-1} = \sum_{n=0}^{\infty} \ell^{\otimes n}. \quad (2.5)$$

*Remark 2.2.* The condition  $\ell^\emptyset = 0$  can easily be loosened to  $\ell^\emptyset \neq 1$  by scaling and shifting  $\ell$ , i.e.

$$(\emptyset - \ell)^{-1} = \frac{1}{1 - \ell^\emptyset} \left( \emptyset - \frac{1}{1 - \ell^\emptyset} (\ell - \ell^\emptyset \emptyset) \right)^{-1} = \sum_{n=0}^{\infty} \frac{1}{(1 - \ell^\emptyset)^{n+1}} (\ell - \ell^\emptyset \emptyset)^{\otimes n}.$$

We thus write  $(\emptyset - \ell)^{-1}$  for all  $\ell \in T((\mathbb{R}^d))$  such that  $\ell^\emptyset \neq 1$ .  $\blacksquare$

*Remark 2.3.* It is clear that  $(\emptyset - \ell)(\emptyset - \ell)^{-1} = (\emptyset - \ell)^{-1}(\emptyset - \ell) = \emptyset$ .  $\blacksquare$

The purpose of the resolvent is to solve linear algebraic equations as shown in the next proposition.

**Proposition 2.2.** *Let  $\mathbf{p}, \mathbf{q} \in T((\mathbb{R}^d))$  such that  $\mathbf{q}^\emptyset \neq 1$ , then the unique solution  $\ell \in T((\mathbb{R}^d))$  to the linear algebraic equation*

$$\ell = \mathbf{p} + \ell \mathbf{q} \quad (2.6)$$

*is given by*

$$\ell = \mathbf{p}(\emptyset - \mathbf{q})^{-1}.$$

*Proof.* It is easy to verify that  $\ell = \mathbf{p}(\boldsymbol{\phi} - \mathbf{q})^{-1}$  is a solution of (2.6). On the other hand, for every  $\boldsymbol{\varphi} \in T((\mathbb{R}^d))$  s.t.  $\boldsymbol{\varphi} = \mathbf{p} + \boldsymbol{\varphi}\mathbf{q}$ , we directly derive  $\boldsymbol{\varphi}(\boldsymbol{\phi} - \mathbf{q}) = \mathbf{p} = \ell(\boldsymbol{\phi} - \mathbf{q})$  and then  $(\boldsymbol{\varphi} - \ell)(\boldsymbol{\phi} - \mathbf{q}) = 0$ . Hence  $\boldsymbol{\varphi} = \ell$ , verifying the uniqueness.  $\square$

Interestingly, whenever  $\ell$  is a linear combination of single letters, the resolvent of  $\ell$  is equal to the shuffle exponential  $e^{\sqcup\sqcup\ell}$  defined by

$$e^{\sqcup\sqcup\ell} := \sum_{n=0}^{\infty} \frac{\ell^{\sqcup\sqcup n}}{n!} \quad (2.7)$$

with

$$\ell^{\sqcup\sqcup n} := \overbrace{\ell \sqcup\sqcup \ell \sqcup\sqcup \ell \cdots \sqcup\sqcup \ell}^{n \text{ times}}, \quad n \geq 1, \quad \ell^{\sqcup\sqcup 0} = \boldsymbol{\phi}.$$

**Proposition 2.3.** *Let  $\mathbf{b} = \sum_{\mathbf{i} \in A_d} \mathbf{b}^{\mathbf{i}}$ , with  $\mathbf{b}^{\mathbf{i}} \in \mathbb{R}$ , we have that*

$$(\boldsymbol{\phi} - \mathbf{b})^{-1} = e^{\sqcup\sqcup\mathbf{b}}. \quad (2.8)$$

*In particular, this implies*

$$e^{\sqcup\sqcup\mathbf{b}} = \boldsymbol{\phi} + e^{\sqcup\sqcup\mathbf{b}}\mathbf{b} = \boldsymbol{\phi} + \mathbf{b}e^{\sqcup\sqcup\mathbf{b}}. \quad (2.9)$$

*Proof.* We first observe that  $\mathbf{b}^{\boldsymbol{\phi}} = 0$ , so  $(\boldsymbol{\phi} - \mathbf{b})^{-1}$  is well-defined. We then prove that  $\frac{1}{k!}\mathbf{b}^{\sqcup\sqcup k} = \mathbf{b}^{\otimes k}$ . By induction, it suffices to prove for every  $k \in \mathbb{N}^*$ ,

$$\mathbf{b}^{\otimes k-1} \sqcup\sqcup \mathbf{b} = k\mathbf{b}^{\otimes k}. \quad (2.10)$$

It is easy to check that (2.10) is satisfied for  $k = 1$ . Now, assuming it holds for some  $k \geq 1$ , we can deduce

$$\begin{aligned} \mathbf{b}^{\otimes k} \sqcup\sqcup \mathbf{b} &= \sum_{\mathbf{i} \in A_d} \mathbf{b}^{\otimes k} \mathbf{b}^{\mathbf{i}} + \sum_{\mathbf{i} \in A_d} (\mathbf{b}^{\otimes k} |_{\mathbf{i}} \sqcup\sqcup \mathbf{b}) \mathbf{i} \\ &= \mathbf{b}^{\otimes k+1} + \sum_{\mathbf{i} \in A_d} (\mathbf{b}^{\otimes k-1} \sqcup\sqcup \mathbf{b}) \mathbf{b}^{\mathbf{i}} \\ &= \mathbf{b}^{\otimes k+1} + (\mathbf{b}^{\otimes k-1} \sqcup\sqcup \mathbf{b}) \mathbf{b} \\ &= \mathbf{b}^{\otimes k+1} + (k\mathbf{b}^{\otimes k}) \mathbf{b} \\ &= (k+1)\mathbf{b}^{\otimes k+1}, \end{aligned}$$

which proves (2.10) by induction, hence proving the equality (2.8). Noticing that  $(\boldsymbol{\phi} - \mathbf{b})^{-1} = \boldsymbol{\phi} + \mathbf{b}(\boldsymbol{\phi} - \mathbf{b})^{-1} = \boldsymbol{\phi} + (\boldsymbol{\phi} - \mathbf{b})^{-1}\mathbf{b}$ , we derive (2.9).  $\square$

### 2.3 Signature of a continuous semimartingale

We define the (path) signature of a semimartingale as the sequence of iterated stochastic integrals in the sense of Stratonovich. Throughout the paper, the Itô integral is denoted by  $\int_0^\cdot Y_t dX_t$  and the Stratonovich integral by  $\int_0^\cdot Y_t \circ dX_t$ . If both  $X$  and  $Y$  are semimartingales then,  $\int_0^\cdot Y_t \circ dX_t := \int_0^\cdot Y_t dX_t + \frac{1}{2}[X, Y]$ .

**Definition 2.4** (Signature). *Fix  $T > 0$ . Let  $(X_t)_{t \geq 0}$  be a continuous semimartingale on  $\mathbb{R}^d$  on a complete filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ . The signature of  $X$  is defined by*

$$\begin{aligned} \mathbb{X} : \Omega \times [0, T] &\rightarrow T((\mathbb{R}^d)) \\ (\omega, t) &\mapsto \mathbb{X}_t(\omega) := (1, \mathbb{X}_t^1(\omega), \dots, \mathbb{X}_t^n(\omega), \dots), \end{aligned}$$

where

$$\mathbb{X}_t^n := \int_{0 < u_1 < \dots < u_n < t} \circ dX_{u_1} \otimes \dots \otimes \circ dX_{u_n}$$

takes value in  $(\mathbb{R}^d)^{\otimes n}$ ,  $n \geq 0$ . Similarly, the truncated signature of order  $M \in \mathbb{N}$  is defined by

$$\begin{aligned} \mathbb{X}^{\leq M} : [0, T] &\rightarrow T^M(\mathbb{R}^d) \\ (\omega, t) &\mapsto \mathbb{X}_t^{\leq M}(\omega) := (1, \mathbb{X}_t^1(\omega), \dots, \mathbb{X}_t^M(\omega), 0, \dots, 0, \dots). \end{aligned} \quad (2.11)$$

The signature plays a similar role as polynomials on path-space. Indeed, in dimension  $d = 1$ , the signature of  $X$  is the sequence of monomials  $\left(\frac{(X_t - X_0)^n}{n!}\right)_{n \in \mathbb{N}}$ . In particular, for  $\ell \in T^M(\mathbb{R})$ ,  $\langle \ell, \mathbb{X}_t \rangle$  is a polynomial in  $(X_t - X_0)$  of degree no greater than  $M$ .

*Example 2.4.* For  $X_t = (t, W_t^1, W_t^2, \dots, W_t^d)$ , where  $(W_t^1, W_t^2, \dots, W_t^d)$  is a  $d$ -dimensional Brownian motion, we can compute the first level of the signature

$$\mathbb{X}_t^1 = (t, W_t^1, W_t^2, \dots, W_t^d),$$

and the second level

$$\mathbb{X}_t^2 = \begin{pmatrix} \frac{t^2}{2} & \int_0^t s dW_s^1 & \cdots & \int_0^t s dW_s^d \\ \int_0^t W_s^1 ds & \frac{(W_t^1)^2}{2} & \cdots & \int_0^t W_s^1 \circ dW_s^d \\ \vdots & \vdots & \ddots & \vdots \\ \int_0^t W_s^d ds & \int_0^t W_s^d \circ dW_s^1 & \cdots & \frac{(W_t^d)^2}{2} \end{pmatrix}.$$

*Remark 2.5.* It follows from Definition 2.4 that the level  $n$  of the signature  $\mathbb{X}_t^n := (\mathbb{X}_t^{\mathbf{i}_1 \cdots \mathbf{i}_n})_{(\mathbf{i}_1 \cdots \mathbf{i}_n) \in V_n} \in (\mathbb{R}^d)^{\otimes n}$  can be written in the following iterated form

$$\mathbb{X}_t^{\mathbf{i}_1 \cdots \mathbf{i}_n} = \int_0^t \mathbb{X}_s^{\mathbf{i}_1 \cdots \mathbf{i}_{n-1}} \circ dX_s^{\mathbf{i}_n}, \quad \mathbf{i}_1 \cdots \mathbf{i}_n \in V_n. \quad (2.12)$$

## 2.4 Finite linear combinations of signature elements

In this section, we fix  $X$  a continuous semimartingale on  $\mathbb{R}^d$  and we let  $\mathbb{X}$  be its signature. We recall some important properties of finite linear combinations of signature elements  $\langle \ell, \mathbb{X}_t \rangle$  for  $\ell \in T(\mathbb{R}^d)$ , recall (2.4).

The first one highlights a crucial linearization property of the signature, the product of two linear combinations of the signature is again a linear combination of the signature where the coefficients are given by the shuffle product, recall Definition 2.1. This property will be later extended to infinite linear combinations of signatures in Proposition 3.1 below.

**Lemma 2.5** (Shuffle product property). *For  $\ell, \varphi \in T(\mathbb{R}^d)$ ,*

$$\langle \ell, \mathbb{X}_t \rangle \langle \varphi, \mathbb{X}_t \rangle = \langle \ell \sqcup \varphi, \mathbb{X}_t \rangle.$$

*Proof.* This follows from the standard chain rule of Stratonovich integrals, see Gaines [26, Proposition 2.2].  $\square$

The shuffle product can be seen as an extension of the Cauchy product to the space of signatures. For example, set  $d = 1$ ,  $X_0 = 0$  and  $\ell \in T^M(\mathbb{R})$ ,  $\varphi \in T^N(\mathbb{R})$  for some  $N, M \in \mathbb{N}$ , i.e. there exists  $(a_m)_{m \leq M}, (b_n)_{n \leq N} \in \mathbb{R}$  such that  $\ell = \sum_{m=0}^M a_m \mathbf{1}^{\otimes m}$ ,  $\varphi = \sum_{n=0}^N b_n \mathbf{1}^{\otimes n}$ . Then,

$$\langle \ell, \mathbb{X}_t \rangle \langle \varphi, \mathbb{X}_t \rangle = \sum_{m=0}^M a_m \frac{(X_t)^m}{m!} \sum_{n=0}^N b_n \frac{(X_t)^n}{n!} = \sum_{k=0}^{M+N} c_k \frac{(X_t)^k}{k!}$$

with  $c_k = \sum_{i=0}^k \binom{k}{i} a_i b_{k-i}$ . In this case  $\ell \sqcup \varphi = \sum_{k=0}^{M+N} c_k \mathbf{1}^{\otimes k}$ .

The second property shows that any finite linear combination of signature elements is a semi-martingale with an explicit decomposition given in terms of the projection elements, recall (2.3). This semi-martingality property is no longer always true when infinite linear combinations of signature elements are considered, see Theorem 3.3 below.

**Lemma 2.6.** *Let  $\ell \in T(\mathbb{R}^d)$ , then  $(\langle \ell, \mathbb{X}_t \rangle)_{t \geq 0}$  is a semimartingale with decomposition*

$$d \langle \ell, \mathbb{X}_t \rangle = \sum_{\mathbf{i} \in A_d} \langle \ell, \mathbf{i} \rangle \circ dX_t^{\mathbf{i}}.$$

*Proof.* It follows from the equivalent definition of the signature (2.12) that

$$\langle \ell |_{\mathbf{i}}, \mathbb{X}_t \rangle = \int_0^t \langle \ell |_{\mathbf{i}}, \mathbb{X}_s \rangle \circ dX_s^{\mathbf{i}}, \quad \mathbf{i} \in A_d.$$

Then, using the projection relation (2.3) and the linearity of the bracket, we get that

$$\langle \ell, \mathbb{X}_t \rangle = \left\langle \ell^\emptyset + \sum_{\mathbf{i} \in A_d} \ell |_{\mathbf{i}}, \mathbb{X}_t \right\rangle = \ell^\emptyset + \sum_{\mathbf{i} \in A_d} \int_0^t \langle \ell |_{\mathbf{i}}, \mathbb{X}_s \rangle \circ dX_s^{\mathbf{i}}.$$

□

Finally, we show how to transform the Stratonovich integral into the Itô integral in the context of finite linear combination of the signature.

**Lemma 2.7.** *Let  $\ell \in T(\mathbb{R}^d)$ , then for every  $\mathbf{i} \in A_d$ :*

$$\langle \ell, \mathbb{X}_t \rangle \circ dX_t^{\mathbf{i}} = \langle \ell, \mathbb{X}_t \rangle dX_t^{\mathbf{i}} + \frac{1}{2} \sum_{\mathbf{j} \in A_d} \langle \ell |_{\mathbf{j}}, \mathbb{X}_t \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_t.$$

*Proof.* Denote by  $Y_t = \langle \ell, \mathbb{X}_t \rangle$  and fix  $\mathbf{i} \in A_d$ . It follows from Lemma 2.6 that  $Y$  is a semimartingale with dynamics

$$dY_t = \sum_{\mathbf{j} \in A_d} \langle \ell |_{\mathbf{j}}, \mathbb{X}_t \rangle \circ dX_t^{\mathbf{j}}.$$

Therefore, by the definition of the Stratonovich integral:

$$Y_t \circ dX_t^{\mathbf{i}} = Y_t dX_t^{\mathbf{i}} + \frac{1}{2} d[Y, X^{\mathbf{i}}]_t = Y_t dX_t^{\mathbf{i}} + \frac{1}{2} \sum_{\mathbf{j} \in A_d} \langle \ell |_{\mathbf{j}}, \mathbb{X}_t \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_t.$$

□

### 3 Infinite linear combinations of signature elements

From now on, we fix a complete filtered probability space  $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$  satisfying the usual conditions, and  $(X_t)_{0 \leq t \leq T}$  a continuous  $\mathbb{F}$ -semimartingale on  $\mathbb{R}^d$  and we let  $\mathbb{X}$  be its signature. In Section 2.3, we recalled well-known properties of finite linear combinations of signature elements  $\langle \ell, \mathbb{X}_t \rangle$  for  $\ell \in T(\mathbb{R}^d)$ . In this section, we study infinite linear combinations  $\langle \ell, \mathbb{X}_t \rangle$  for certain admissible  $\ell \in T(\mathbb{R}^d)$  for which the infinite series will make sense. Our central results in this section are an Itô formula for such elements in Theorem 3.3 below, a control for their supremum norm in Propositions 3.5 and 3.6, and an estimate of solutions of algebraic linear equations taking values in the extended tensor algebra in Theorem 3.9.

For this, we consider the space  $\mathcal{A}(\mathbb{X})$  of admissible elements  $\ell$  using the associated stochastic semi-norm:

$$\|\ell\|_t^{\mathcal{A}(\mathbb{X})} := \sum_{n=0}^{\infty} \left| \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbb{X}_t^{\mathbf{v}} \right|, \quad t \geq 0,$$

recall the definition of  $V_n$  in (2.1) and the decomposition (2.2).  $\|\ell\|_t^{\mathcal{A}(\mathbb{X})}$  is actually random variables. Whenever  $\|\ell\|_t^{\mathcal{A}(\mathbb{X})} < \infty$  a.s., the infinite linear combination

$$\langle \ell, \mathbb{X}_t \rangle := \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbb{X}_t^{\mathbf{v}}$$

is well-defined. This leads to the following definition for the admissible set  $\mathcal{A}(\mathbb{X})$ :

$$\mathcal{A}(\mathbb{X}) := \left\{ \ell \in T(\mathbb{R}^d) : \|\ell\|_t^{\mathcal{A}(\mathbb{X})} < \infty \text{ for all } t \in [0, T] \text{ a.s.} \right\}.$$

For simplicity, we write  $\mathcal{A}$  (resp.  $\|\cdot\|_t^{\mathcal{A}}$ ), in place of  $\mathcal{A}(\mathbb{X})$  (resp.  $\|\cdot\|_t^{\mathcal{A}(\mathbb{X})}$ ), when it does not cause ambiguity. Note that  $T(\mathbb{R}^d) \subset \mathcal{A}$  and that  $\langle \ell, \mathbb{X}_t \rangle$  extends (2.4), as the two bracket operations  $\langle \cdot, \cdot \rangle$  coincide whenever  $\ell \in T(\mathbb{R}^d)$ .

*Remark 3.1.* The space  $\mathcal{A}$  plays a crucial role in this paper. Whenever we talk about  $\langle \ell, \mathbb{X}_t \rangle$  for  $\ell \in T((\mathbb{R}^d))$ , we have to first verify that  $\ell$  belongs to  $\mathcal{A}$  to ensure that  $\langle \ell, \mathbb{X}_t \rangle$  is well-defined. The set  $\mathcal{A}$  is a natural set to consider. It has also been studied in Cuchiero, Svaluto-Ferro, and Teichmann [17, Example 3.2 (iii)]. In practice, however, we will use a stronger but more tractable criterion in Section 3.2 below to obtain that a given  $\ell$  belongs to  $\mathcal{A}$ . ■

*Remark 3.2.* If  $\ell \in \mathcal{A}$ , then  $\langle \ell, \mathbb{X}_t \rangle$  is a progressively measurable process, as it is the a.s.-limit of a series of progressively measurable processes. ■

*Remark 3.3.* We could restrict to a slightly smaller space  $\tilde{\mathcal{A}}$ :

$$\tilde{\mathcal{A}}(\mathbb{X}) := \left\{ \ell \in T((\mathbb{R}^d)) : \sup_{t \in [0, T]} \|\ell\|_t^{\mathcal{A}(\mathbb{X})} < \infty \text{ a.s.} \right\},$$

and show that whenever  $X$  and  $Y$  follow the same law,  $\tilde{\mathcal{A}}(\mathbb{X}) = \tilde{\mathcal{A}}(\mathbb{Y})$ . This would however not be the case in general for  $\mathcal{A}$  due to measurability issue. ■

It is straightforward to see that the sets  $\mathcal{A}$  is closed under linear operations, i.e. for  $\alpha, \beta \in \mathbb{R}$  and  $\ell, \varphi \in \mathcal{A}$ , then  $\alpha\ell + \beta\varphi \in \mathcal{A}$ . Moreover,  $\mathcal{A}$  is closed under the shuffle product.

The next proposition extends the shuffle property, derived in Proposition 2.5, to the case of infinite linear combination of signatures. We note that it can be obtained from a particular instance of the more general result in [17, Lemma 4.4] dealing with so-called shuffle compatible partitions. For completeness we provide a proof for our specific setting in Appendix C.

**Proposition 3.1** (Extended Shuffle property). *If  $\ell, \varphi \in \mathcal{A}$ , then  $\ell \sqcup \varphi \in \mathcal{A}$  and*

$$\langle \ell, \mathbb{X}_t \rangle \langle \varphi, \mathbb{X}_t \rangle = \langle \ell \sqcup \varphi, \mathbb{X}_t \rangle.$$

*Proof.* The proof is provided in Appendix C. □

### 3.1 The space $\mathcal{I}$ and Itô's formula

The aim of this subsection is the derivation of an Itô's formula in Theorem 3.3 below for infinite linear combinations of signature elements. In preparation, we define  $\mathcal{Q}_i(X)$ , the set of coordinates of  $X$  that have a non-zero quadratic covariation with  $X^i$ :

$$\mathcal{Q}_i(X) := \{ \mathbf{j} \in A_d : [X^{\mathbf{j}}, X^i]_t \neq 0, \text{ on a set of non-zero } dt \otimes d\mathbb{P} \text{ measure} \}, \quad \mathbf{i} \in A_d.$$

*Remark 3.4.*  $\mathcal{Q}_i(X)$  is introduced in order to get rid of unnecessary assumptions in Lemma 3.2, since we only need  $\ell|_{\mathbf{j}} \in \mathcal{A}$  for those  $\mathbf{j} \in \mathcal{Q}_i(X)$ . For example, in the particular case of  $X_t = (t, W_t)$ , with  $W$  a one-dimensional Brownian motion, assumptions are only needed for the terms  $\ell, \ell|_{\mathbf{1}}, \ell|_{\mathbf{2}}, \ell|_{\mathbf{22}} \in \mathcal{A}$ , but not for  $\ell|_{\mathbf{21}}$  and  $\ell|_{\mathbf{12}}$ . ■

**Lemma 3.2.** *Fix  $\mathbf{i} \in A_d$  and let  $\ell \in \mathcal{A}$  be such that*

$$\ell|_{\mathbf{i}} \in \mathcal{A}, \quad \int_0^T \left( \|\ell\|_t^{\mathcal{A}} \right)^2 d[X^{\mathbf{i}}, X^{\mathbf{i}}]_t < \infty \text{ and } \int_0^T \|\ell\|_t^{\mathcal{A}} d|F_t^{\mathbf{i}}| < \infty, \quad (3.1)$$

where  $F^{\mathbf{i}}$  is the finite variation part of  $X^{\mathbf{i}}$ . We also assume that for every  $\mathbf{j} \in \mathcal{Q}_i(X)$ ,

$$\ell|_{\mathbf{j}} \in \mathcal{A} \quad \text{and} \quad \int_0^T \|\ell|_{\mathbf{j}}\|_t^{\mathcal{A}} d[X^{\mathbf{j}}, X^{\mathbf{i}}]_t < \infty. \quad (3.2)$$

Then,

$$\langle \ell|_{\mathbf{i}}, \mathbb{X}_t \rangle = \int_0^t \langle \ell, \mathbb{X}_s \rangle dX_s^{\mathbf{i}} + \frac{1}{2} \sum_{\mathbf{j} \in A_d} \int_0^t \langle \ell|_{\mathbf{j}}, \mathbb{X}_s \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s, \quad \text{a.s.} \quad (3.3)$$

for every  $t \in [0, T]$ .

*Proof.* Fix  $t \in [0, T]$ . Clearly, under the assumptions (3.1)-(3.2) on  $\ell$ , the right side of equation (3.3) is well-defined. For every word  $\mathbf{v} \in V$ , we can write it as  $\mathbf{v} = \sum_{\mathbf{j} \in A_d} \mathbf{v}|_{\mathbf{j}}$ . By Lemma 2.7, we obtain:

$$\langle \mathbf{v}\mathbf{i}, \mathbb{X}_t \rangle = \int_0^t \langle \mathbf{v}, \mathbb{X}_s \rangle \circ dX_s^{\mathbf{i}} = \frac{1}{2} \sum_{\mathbf{j} \in A_d} \int_0^t \langle \mathbf{v}|_{\mathbf{j}}, \mathbb{X}_s \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s + \int_0^t \langle \mathbf{v}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}}.$$

Combined with the decomposition (2.2), it yields that

$$\langle \ell\mathbf{i}, \mathbb{X}_t \rangle = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \langle \mathbf{v}\mathbf{i}, \mathbb{X}_t \rangle = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \int_0^t \left( \frac{1}{2} \sum_{\mathbf{j} \in A_d} \langle \mathbf{v}|_{\mathbf{j}}, \mathbb{X}_s \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s + \langle \mathbf{v}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}} \right). \quad (3.4)$$

It remains to argue, using dominated convergence, that both terms appearing on the right-hand side of (3.4) correspond to the terms in (3.3).

- (i) For the first term, if  $\mathbf{j} \notin \mathcal{Q}_i(X)$ ,  $[X^{\mathbf{j}}, X^{\mathbf{i}}]_t = 0$  for all  $t$ , so we only care about the letters  $\mathbf{j} \in \mathcal{Q}_i(X)$ . Since  $\ell|_{\mathbf{j}} = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \mathbf{v}|_{\mathbf{j}} \in \mathcal{A}$  by assumption (3.2), we can apply the dominated convergence theorem, with  $\|\ell|_{\mathbf{j}}\|_t^{\mathcal{A}}$  as the dominated function, and get

$$\begin{aligned} \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \int_0^t \langle \mathbf{v}|_{\mathbf{j}}, \mathbb{X}_s \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s &= \int_0^t \left( \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \langle \mathbf{v}|_{\mathbf{j}}, \mathbb{X}_s \rangle \right) d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s \\ &= \int_0^t \langle \ell|_{\mathbf{j}}, \mathbb{X}_s \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s \end{aligned} \quad (3.5)$$

- (ii) For the second term, let us define

$$Y_t^N := \sum_{n=0}^N \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \int_0^t \langle \mathbf{v}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}}.$$

On the one hand, as  $\int_0^T \left( \|\ell\|_t^{\mathcal{A}} \right)^2 d[X^{\mathbf{i}}, X^{\mathbf{i}}]_t < \infty$  and  $\sum_{n=0}^N \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \langle \mathbf{v}, \mathbb{X}_t \rangle \leq \|\ell\|_t^{\mathcal{A}}$  for every  $N$  and assumption (3.1), by the dominated convergence theorem for stochastic integrals, the random variable  $Y^N$  converges in probability to

$$Y_t^{\infty} := \int_0^t \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \langle \mathbf{v}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}}$$

as  $N \rightarrow \infty$ .

On the other hand, since  $\ell\mathbf{i} \in \mathcal{A}$ , the left-hand side of (3.4) is a.s. finite, and since second term of its right-hand side has been proved to be finite, it follows that  $\sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \int_0^t \langle \mathbf{v}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}}$  exists a.s. and corresponds to an almost sure limit of  $Y^N$  as  $N \rightarrow \infty$ . Which yields

$$\sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \ell^{\mathbf{v}} \int_0^t \langle \mathbf{v}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}} = Y_t^{\infty} = \int_0^t \langle \ell, \mathbb{X}_s \rangle dX_s^{\mathbf{i}} \quad a.s. \quad (3.6)$$

Combining (3.4), (3.5) and (3.6) yields (3.3) and completes the proof.  $\square$

*Remark 3.5.* To obtain that equality (3.3) holds for every  $t$  almost surely, one could argue that  $\langle \ell\mathbf{i}, \mathbb{X}_t \rangle$  is a continuous process for instance. This requires additional conditions on  $\ell$ , see for instance Proposition 3.6-(iii) below.  $\blacksquare$

We are now in place to state the main result of this section which is an Itô's formula for infinite linear combinations of signature elements. For this, we need to define the following set

$$\mathcal{I}(X) := \left\{ \ell \in \mathcal{A} : \text{for every } \mathbf{i} \in A_d \text{ and } \mathbf{j} \in \mathcal{Q}_i(X), \ell_{|\mathbf{i}}, \ell_{|\mathbf{j}} \in \mathcal{A} \text{ and} \int_0^T \left( \|\ell_{|\mathbf{i}}\|_t^A d|F_t^{\mathbf{i}}| + \|\ell_{|\mathbf{j}}\|_t^A d[X^{\mathbf{j}}, X^{\mathbf{i}}]_t + \left( \|\ell_{|\mathbf{i}}\|_t^A \right)^2 d[X^{\mathbf{i}}, X^{\mathbf{i}}]_t \right) < \infty, \text{ a.s.} \right\},$$

which can be seen as the analog of the space  $C^2(\mathbb{R})$  when one applies Itô's formula on  $f(W)$ , see Example 3.6 below.

**Theorem 3.3** (Itô's formula). *Let  $\ell \in \mathcal{I}(X)$ , then  $(\langle \ell, \mathbb{X}_t \rangle)_{t \leq T}$  is an Itô process such that*

$$\langle \ell, \mathbb{X}_t \rangle = \ell^\circ + \sum_{\mathbf{i} \in A_d} \int_0^t \langle \ell_{|\mathbf{i}}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}} + \frac{1}{2} \sum_{\mathbf{i} \in A_d} \sum_{\mathbf{j} \in A_d} \int_0^t \langle \ell_{|\mathbf{j}}, \mathbb{X}_s \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s, \quad (3.7)$$

for all  $t \leq T$ . In particular, when we take  $X_t$  to be  $\widehat{W}_t = (t, W_t^2, \dots, W_t^d)$ , with  $W$  a  $(d-1)$ -Brownian motion then, the set  $\mathcal{I}$  reads

$$\mathcal{I}(\widehat{W}) = \left\{ \ell \in \mathcal{A} : \ell_{|\mathbf{1}}, \ell_{|\mathbf{j}}, \ell_{|\mathbf{jj}} \in \mathcal{A}, \mathbf{j} \in \{2, \dots, d\}, \text{ and} \int_0^T \|\ell_{|\mathbf{1}}\|_t^A dt + \sum_{\mathbf{j} \in \{2, \dots, d\}} \int_0^T \left( \|\ell_{|\mathbf{jj}}\|_t^A + \left( \|\ell_{|\mathbf{j}}\|_t^A \right)^2 \right) dt < \infty, \text{ a.s.} \right\},$$

and for every  $\ell \in \mathcal{I}(\widehat{W})$ :

$$\langle \ell, \widehat{\mathbb{W}}_t \rangle = \ell^\circ + \int_0^t \left\langle \ell_{|\mathbf{1}} + \frac{1}{2} \sum_{\mathbf{j} \in \{2, \dots, d\}} \ell_{|\mathbf{jj}}, \widehat{\mathbb{W}}_s \right\rangle ds + \sum_{\mathbf{j} \in \{2, \dots, d\}} \int_0^t \langle \ell_{|\mathbf{j}}, \widehat{\mathbb{W}}_s \rangle dW_s^{\mathbf{j}}. \quad (3.8)$$

*Proof.* The proof follows from an application of Lemma 3.2 combined with the decomposition (2.3).  $\square$

The following example highlights that Theorem 3.3 is indeed an extension of the usual Itô's formula on  $f(W)$ .

*Example 3.6.* Let  $d = 2$  and  $\widehat{W}_t = (t, W_t)$ , and fix an analytic function  $f(y) := \sum_{n \geq 0} a_n y^n$  with infinite radius of convergence which we apply to  $W$ . It is clear that

$$f(W_t) = \sum_{n \geq 0} a_n W_t^n = \langle \ell, \widehat{\mathbb{W}}_t \rangle,$$

where

$$\ell = \sum_{n \geq 0} a_n \mathbf{2}^{\llcorner n} = \sum_{n \geq 0} a_n n! \mathbf{2}^{\otimes n}.$$

In particular, the projections read:

$$\begin{aligned} \ell_{|\mathbf{1}} &= 0, \\ \ell_{|\mathbf{2}} &= \sum_{n \geq 0} a_{n+1} (n+1)! \mathbf{2}^{\otimes n} = \sum_{n \geq 0} a_{n+1} (n+1) \mathbf{2}^{\llcorner n}, \\ \ell_{|\mathbf{22}} &= \sum_{n \geq 0} a_{n+2} (n+2)! \mathbf{2}^{\otimes n} = \sum_{n \geq 0} a_{n+2} (n+1)(n+2) \mathbf{2}^{\llcorner n}. \end{aligned}$$

It is easy to verify that  $\ell, \ell_{|\mathbf{2}}, \ell_{|\mathbf{22}} \in \mathcal{A}$  since  $f$  has infinite radius of convergence. We can thus further derive that

$$f'(W_t) = \langle \ell_{|\mathbf{2}}, \widehat{\mathbb{W}}_t \rangle, \quad f''(W_t) = \langle \ell_{|\mathbf{22}}, \widehat{\mathbb{W}}_t \rangle.$$

On the other hand we can see that  $\|\ell_{|\mathbf{2}}\|_t^A = \sum_{n=0}^{\infty} (n+1) |a_{n+1}| \cdot |W_t|^n$ , and since  $g(x) := \sum_{n \geq 0} (n+1) |a_{n+1}| x^n$  is also analytic and  $\|\ell_{|\mathbf{2}}\|_t^A$  has continuous sample path almost surely, then

$$\sup_{t \in [0, T]} \|\ell_{|\mathbf{2}}\|_t^A < \infty.$$

With similar arguments we can show that  $\sup_{t \in [0, T]} \|\ell|_{\mathbf{22}}\|_t^A < \infty$ . This allows us to verify that  $\ell \in \mathcal{I}(\widehat{W})$ . An application of our Itô's formula (3.8) yields

$$\langle \ell, \widehat{W}_t \rangle = \ell^\circ + \int_0^t \langle \ell|_{\mathbf{2}}, \widehat{W}_s \rangle dW_s + \frac{1}{2} \int_0^t \langle \ell|_{\mathbf{22}}, \widehat{W}_s \rangle ds,$$

which translates into

$$f(W_t) = f(0) + \int_0^t f'(W_s) dW_s + \frac{1}{2} \int_0^t f''(W_s) ds,$$

the standard Itô's formula.  $\blacksquare$

*Remark 3.7.* It should be noted that the identity (3.7) only holds almost surely for every fixed  $t$ . However, as remark 3.5, if we additionally assume  $\langle \ell, \mathbb{X}_t \rangle$  is continuous, then (3.7) holds for every  $t \in [0, T]$  almost surely. In particular, when  $X_t = \widehat{W}_t = (t, W_t^2, \dots, W_t^d)$ , and  $\ell \in \mathcal{A}_h$  (where  $\mathcal{A}_h$  is introduced in Section 3.2 below), the continuity assumption is satisfied by Proposition (3.6)-(iii). Our applications of Theorem 3.3 in later section are all in this strong sense.  $\blacksquare$

We can easily extend Theorem 3.3 to time-dependent linear functionals, i.e.  $\ell : [0, T] \rightarrow T((\mathbb{R}^d))$ . For this, let us define

$$\mathcal{I}'(X) := \left\{ \ell : [0, T] \rightarrow \mathcal{A} : \begin{array}{l} \text{for every } \mathbf{v} \in V, \ell^\mathbf{v} \in C^1([0, T], \mathbb{R}) \text{ and} \\ \text{for every } \mathbf{i} \in A_d, \mathbf{j} \in \mathcal{Q}_i(X) \text{ and for all } t \in [0, T], \ell_t|_{\mathbf{i}}, \ell_t|_{\mathbf{j}}, \dot{\ell}_t \in \mathcal{A} \text{ and} \\ \int_0^T \left( \|\ell_t|_{\mathbf{i}}\|_t^A d|F_t^{\mathbf{i}}| + \|\ell_t|_{\mathbf{j}}\|_t^A d[X^{\mathbf{j}}, X^{\mathbf{i}}]_t + \left( \|\ell_t|_{\mathbf{i}}\|_t^A \right)^2 d[X^{\mathbf{i}}, X^{\mathbf{i}}]_t \right) < \infty \text{ a.s.} \end{array} \right\},$$

where  $\dot{\ell}_t = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \frac{d}{dt} \ell_t^\mathbf{v} \mathbf{v}$ .

**Corollary 3.4.** *Let  $\ell \in \mathcal{I}'(X)$ , then  $(\langle \ell_t, \mathbb{X}_t \rangle)_{t \leq T}$  is an Itô process such that*

$$\langle \ell_t, \mathbb{X}_t \rangle = \ell_t^\circ + \sum_{\mathbf{i} \in A_d} \int_0^t \langle \ell_s|_{\mathbf{i}}, \mathbb{X}_s \rangle dX_s^{\mathbf{i}} + \frac{1}{2} \sum_{\mathbf{j} \in A_d} \sum_{\mathbf{i} \in A_d} \int_0^t \langle \ell_s|_{\mathbf{j}}, \mathbb{X}_s \rangle d[X^{\mathbf{j}}, X^{\mathbf{i}}]_s + \int_0^t \langle \dot{\ell}_s, \mathbb{X}_s \rangle ds,$$

for all  $t \leq T$ .

*Proof.* The proof is similar to the proof of Theorem 3.3.  $\square$

*Remark 3.8.* The example of the fractional Riemann-Liouville Brownian motion, treated in Section 4.3, shows that, contrary to the finite case, not all infinite linear combinations of signature elements of  $\widehat{W}$  are semimartingales. In particular, Theorem 3.3 and Corollary 3.4 cannot be applied in such cases.  $\blacksquare$

## 3.2 The space $\mathcal{A}_h$ and moment estimates

In practice, it is not an easy task to verify directly whether a given  $\ell$  belongs to  $\mathcal{A}$ , even for the case of the time extended  $(d-1)$ -dimensional Brownian motion  $X_t = \widehat{W}_t = (t, W_t^2, \dots, W_t^d)$ . The aim of this subsection is to introduce a subset  $\mathcal{A}_h$  of  $\mathcal{A}$ , specific to such  $\widehat{W}$  with  $d \geq 2$ , by essentially controlling the expectation of its signature  $\widehat{W}$ . We will give a tractable criterion to prove that a given  $\ell$  belongs to  $\mathcal{A}_h$  and thus to  $\mathcal{A}$ .

In order to motivate the set  $\mathcal{A}_h$ , we will derive an upper bound for the moments of the signature elements of  $\widehat{W}$  with the help of Fawcett's formula [21] using a deterministic function  $h$ . For that and to ease the notations, for every  $\mathbf{v} \in V$  we let  $n(\mathbf{v})$  denote length of  $\mathbf{v}$ , i.e.  $\mathbf{v} \in V_{n(\mathbf{v})}$ , and  $x(\mathbf{v})$  the number of occurrences of the letter  $\mathbf{1}$  in  $\mathbf{v}$ . In the following, if it does not cause ambiguity, we will write  $n$  and  $x$  in place of  $n(\mathbf{v})$  and  $x(\mathbf{v})$  respectively for brevity.

*Example 3.9.* Let  $\mathbf{v} = \mathbf{111321}$ , then  $n = 6$  and  $x = 4$ .  $\blacksquare$

We first provide a tight bound for the second moment:

**Proposition 3.5.** *Fix  $t \in [0, T]$  and  $\mathbf{v} \in V$ . We have*

$$(i) \quad \mathbb{E} \left| \langle \mathbf{v}, \widehat{W}_t \rangle \right|^2 \leq \binom{2n}{n} \frac{t^{n+x}}{(n+x)!} 2^{x-n}, \quad (3.9)$$

(ii) and there exists a non-negative constant  $C \geq 2$  independent of  $\mathbf{v}$  and  $t$ , such that

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left| \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle \right|^2 \right] \leq h_t(\mathbf{v})^2, \quad (3.10)$$

where

$$h_t(\mathbf{v}) := \frac{C}{(n+1)^{\frac{1}{4}}} \frac{(2t)^{\frac{1}{2}(n+x)}}{\sqrt{(n+x-1)!}},$$

with the convention that  $n! = 1$  for  $n \leq 0$ . In particular,  $h$  satisfies the following sub-multiplicative property

$$h_t(\mathbf{uv}) \leq h_t(\mathbf{u})h_t(\mathbf{v}), \quad \mathbf{u}, \mathbf{v} \in V. \quad (3.11)$$

*Proof.* The proof is given in Appendix B.  $\square$

Based on Proposition 3.5, an application of Jensen's inequality yields

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left| \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle \right| \right] \leq h_t(\mathbf{v}). \quad (3.12)$$

We are now ready to introduce a norm on  $T((\mathbb{R}^d))$  using the function  $h$ :

$$\|\ell\|_t^{\mathcal{A}_h} := \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} |\ell^{\mathbf{v}}| h_t(\mathbf{v}), \quad \ell \in T((\mathbb{R}^d)).$$

Unlike  $\|\cdot\|^{\mathcal{A}}$ ,  $\|\cdot\|^{\mathcal{A}_h}$  is deterministic, thus simplifying the computation. Moreover, notice that  $\|\cdot\|_t^{\mathcal{A}_h}$  is an increasing function of  $t$ , which naturally leads to the definition of the normed vector space:

$$\mathcal{A}_h := \left\{ \ell \in T((\mathbb{R}^d)) : \|\ell\|_T^{\mathcal{A}_h} < \infty \right\}.$$

*Remark 3.10.* We can see that  $\mathcal{A}_h$  is a Banach space by simply constructing an isometry from  $\mathcal{A}_h$  to  $\{(a_i)_{i=0}^{\infty} : \sum_{i=0}^{\infty} |a_i| < \infty\}$ , and in particular it is a Köthe sequence space, see [34, 43].  $\blacksquare$

**Proposition 3.6.** Fix  $\ell \in \mathcal{A}_h$ . The following properties hold:

(i)  $\mathcal{A}_h \subset \mathcal{A}$ , and

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \|\ell\|_s^{\mathcal{A}} \right] \leq \|\ell\|_t^{\mathcal{A}_h}, \quad t \leq T.$$

(ii)  $\mathcal{A}_h$  induces a natural upper bound

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left\langle \ell, \widehat{\mathbb{W}}_s \right\rangle^2 \right] \leq \left( \|\ell\|_t^{\mathcal{A}_h} \right)^2, \quad t \leq T.$$

(iii)  $t \mapsto \left\langle \ell, \widehat{\mathbb{W}}_t \right\rangle$  is continuous and the strong error of convergence of the truncated linear combination of the signature elements satisfies

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left| \left\langle \ell, \widehat{\mathbb{W}}_t^{\leq M} \right\rangle - \left\langle \ell, \widehat{\mathbb{W}}_t \right\rangle \right| \right] \leq \sum_{n=M}^{\infty} \sum_{\mathbf{v} \in V_n} |\ell^{\mathbf{v}}| h_T(\mathbf{v}), \quad M \in \mathbb{N}. \quad (3.13)$$

*Proof.* (i) For all  $\ell \in \mathcal{A}_h$ , the inequality (3.12) yields

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \|\ell\|_s^{\mathcal{A}} \right] \leq \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} |\ell^{\mathbf{v}}| \mathbb{E} \left[ \sup_{s \in [0, t]} \left| \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle \right| \right] \leq \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} |\ell^{\mathbf{v}}| h_t(\mathbf{v}) \leq \|\ell\|_t^{\mathcal{A}_h} < \infty.$$

Consequently,  $\sup_{t \in [0, T]} \|\ell\|_t^{\mathcal{A}} < \infty$  a.s. and thus  $\ell \in \mathcal{A}(\widehat{\mathbb{W}})$ .

(ii) Starting with

$$\langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_s \rangle^2 = \left( \sum_{\mathbf{v} \in V} \boldsymbol{\ell}^{\mathbf{v}} \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle \right)^2 \leq \sum_{\mathbf{u} \in V} \sum_{\mathbf{w} \in V} |\boldsymbol{\ell}^{\mathbf{u}} \boldsymbol{\ell}^{\mathbf{w}}| \cdot \left| \langle \mathbf{u}, \widehat{\mathbb{W}}_s \rangle \langle \mathbf{w}, \widehat{\mathbb{W}}_s \rangle \right|,$$

we get that

$$\sup_{s \in [0, t]} \langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_s \rangle^2 \leq \sum_{\mathbf{u} \in V} \sum_{\mathbf{w} \in V} |\boldsymbol{\ell}^{\mathbf{u}} \boldsymbol{\ell}^{\mathbf{w}}| \cdot \sup_{s \in [0, t]} \left| \langle \mathbf{u}, \widehat{\mathbb{W}}_s \rangle \right| \cdot \sup_{s \in [0, t]} \left| \langle \mathbf{w}, \widehat{\mathbb{W}}_s \rangle \right|.$$

Taking the expectation and applying Cauchy-Schwarz inequality yields that

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \in [0, t]} \langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_s \rangle^2 \right] &\leq \sum_{\mathbf{u} \in V} \sum_{\mathbf{w} \in V} |\boldsymbol{\ell}^{\mathbf{u}} \boldsymbol{\ell}^{\mathbf{w}}| \sqrt{\mathbb{E} \sup_{s \in [0, t]} \langle \mathbf{u}, \widehat{\mathbb{W}}_s \rangle^2} \sqrt{\mathbb{E} \sup_{s \in [0, t]} \langle \mathbf{w}, \widehat{\mathbb{W}}_s \rangle^2} \\ &= \left( \sum_{\mathbf{v} \in V} |\boldsymbol{\ell}^{\mathbf{v}}| \sqrt{\mathbb{E} \sup_{s \in [0, t]} \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle^2} \right)^2 \\ &\leq \left( \sum_{\mathbf{v} \in V} |\boldsymbol{\ell}^{\mathbf{v}}| h_t(\mathbf{v}) \right)^2 \\ &= \left( \|\boldsymbol{\ell}\|_t^{\mathcal{A}_h} \right)^2. \end{aligned}$$

(iii) Define

$$X_t := \langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_t \rangle = \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \boldsymbol{\ell}^{\mathbf{v}} \widehat{\mathbb{W}}_t^{\mathbf{v}} \quad \text{and} \quad X_t^M := \langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_t^{\leq M} \rangle = \sum_{n=0}^M \sum_{\mathbf{v} \in V_n} \boldsymbol{\ell}^{\mathbf{v}} \widehat{\mathbb{W}}_t^{\mathbf{v}}.$$

It follows from inequality (3.12) that

$$\mathbb{E} \left[ \sup_{t \in [0, T]} |X_t - X_t^M| \right] \leq \sum_{n=M+1}^{\infty} \sum_{\mathbf{v} \in V_n} |\boldsymbol{\ell}^{\mathbf{v}}| \mathbb{E} \left[ \sup_{t \in [0, T]} \langle \mathbf{v}, \widehat{\mathbb{W}}_t \rangle \right] \leq \sum_{n=M+1}^{\infty} \sum_{\mathbf{v} \in V_n} |\boldsymbol{\ell}^{\mathbf{v}}| h_T(\mathbf{v}).$$

Moreover,  $\boldsymbol{\ell} \in \mathcal{A}_h$  implies that

$$\lim_{M \rightarrow \infty} \sum_{n=M+1}^{\infty} \sum_{\mathbf{v} \in V_n} |\boldsymbol{\ell}^{\mathbf{v}}| h_T(\mathbf{v}) = 0.$$

Thus

$$\mathbb{E} \left[ \sup_{t \in [0, T]} |X_t - X_t^M| \right] \rightarrow 0,$$

implying that

$$\sup_{t \in [0, T]} |X_t - X_t^M| \rightarrow 0 \quad \text{in probability as } M \rightarrow \infty.$$

So there exists a sub-sequence  $(M_k)_{k \in \mathbb{N}}$  such that  $\sup_{t \in [0, T]} |X_t - X_t^{M_k}| \rightarrow 0$  a.s. Moreover, as  $X^{M_k}$  is continuous,  $X$  is also continuous.  $\square$

Derived from (3.11), the next proposition shows that  $\|\cdot\|^{\mathcal{A}_h}$  is sub-multiplicative for the tensor product.

**Proposition 3.7.** *For every  $\boldsymbol{\ell}, \boldsymbol{\varphi} \in T((\mathbb{R}^d))$ ,*

$$\|\boldsymbol{\ell} \boldsymbol{\varphi}\|_t^{\mathcal{A}_h} \leq \|\boldsymbol{\ell}\|_t^{\mathcal{A}_h} \|\boldsymbol{\varphi}\|_t^{\mathcal{A}_h}.$$

*In particular, if  $\boldsymbol{\ell}, \boldsymbol{\varphi} \in \mathcal{A}_h$ , then  $\boldsymbol{\ell} \boldsymbol{\varphi} \in \mathcal{A}_h$ .*

*Proof.* First notice that for all  $k \in \mathbb{N}$  and  $\mathbf{w} \in V_k$ ,

$$\langle \ell \varphi, \mathbf{w} \rangle = \sum_{n=0}^k \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \langle \mathbf{u}\mathbf{v}, \mathbf{w} \rangle,$$

so that

$$\begin{aligned} \|\ell \varphi\|_t^{\mathcal{A}_h} &= \sum_{k=0}^{\infty} \sum_{\mathbf{w} \in V_k} |\langle \ell \varphi, \mathbf{w} \rangle| h_t(\mathbf{w}) \\ &= \sum_{k=0}^{\infty} \sum_{\mathbf{w} \in V_k} \left| \sum_{n=0}^k \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \langle \mathbf{u}\mathbf{v}, \mathbf{w} \rangle \right| h_t(\mathbf{w}) \\ &\leq \sum_{k=0}^{\infty} \sum_{n=0}^k \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \sum_{\mathbf{w} \in V_k} |\ell^{\mathbf{u}} \varphi^{\mathbf{v}}| \langle \mathbf{u}\mathbf{v}, \mathbf{w} \rangle h_t(\mathbf{w}) \end{aligned}$$

Then, using (3.11)

$$\begin{aligned} \|\ell \varphi\|_t^{\mathcal{A}_h} &\leq \sum_{k=0}^{\infty} \sum_{n=0}^k \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} |\ell^{\mathbf{u}} \varphi^{\mathbf{v}}| h_t(\mathbf{u}) h_t(\mathbf{v}) \sum_{\mathbf{w} \in V_k} \langle \mathbf{u}\mathbf{v}, \mathbf{w} \rangle \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_m} |\ell^{\mathbf{u}} \varphi^{\mathbf{v}}| h_t(\mathbf{u}) h_t(\mathbf{v}) \\ &= \|\ell\|_t^{\mathcal{A}_h} \|\varphi\|_t^{\mathcal{A}_h}. \end{aligned}$$

□

*Remark 3.11.* Compared to the literature, our estimate for the supremum norm (3.10) appears to be used for the first time and plays a crucial role in our convergence analysis. The closest related results appeared in Ben Arous [9] where moment estimates of  $\mathbb{E}|\langle \mathbf{v}, \widehat{\mathbb{W}}_t \rangle|^2$  and certain convergence criterion for infinite series of signatures of  $\widehat{W}$ , i.e. iterated sequence of Stratonovich integrals, are established. The estimate in [9, Lemma 2] has the same order as (3.9) given here, although the constant there is not as sharp as ours (since we make use of Fawcett's formula) but has the same order of magnitude (see the remark following [9, Lemma 3]). We also provide estimates for  $\mathbb{E}|\langle \mathbf{v}, \widehat{\mathbb{W}}_t \rangle|^p$  for every  $p \in \mathbb{N}_+$ , see proposition B.1. But most importantly, we provide an estimate for  $\mathbb{E}[\sup_{s \in [0, t]} |\langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle|^2]$  in (3.10), which is necessary to prove convergence for all  $t$  almost surely instead of simply for a fixed  $t$ , when compared to the convergence criterion in [9, Theorem 2 (1)]. Additionally, the second criteria [9, Theorem 2 (2)] requires the second order norm to be bounded, we only require the first order norm  $\|\cdot\|_t^{\mathcal{A}_h}$  to be bounded, which is hence milder. ■

### 3.3 Back to the linear algebraic equation

We now prove that the solution  $\ell$  to the linear algebraic equation (2.6) belongs to  $\mathcal{A}_h$  under suitable conditions on  $\mathbf{p}$  and  $\mathbf{q}$ , see Theorem 3.9. This is a crucial ingredient in obtaining convergence of our representations formula  $\langle \ell, \widehat{\mathbb{W}}_t \rangle$  in Section 4. For this, we start by introducing the following partial order:

**Definition 3.8.** Let  $\ell, \varphi \in T((\mathbb{R}^d))$ . We say that  $\ell$  is dominated by  $\varphi$  if for all  $\mathbf{v} \in V$ ,  $|\ell^{\mathbf{v}}| \leq \varphi^{\mathbf{v}}$ . We then write

$$\ell \preceq \varphi.$$

**Theorem 3.9.** Let  $\mathbf{p} \in \mathcal{A}_h$ , let  $\mathbf{q} \in T((\mathbb{R}^d))$  be such that  $|\mathbf{q}^{\emptyset}| < 1$  and assume there exists  $\varphi \in T((\mathbb{R}^d))$  such that

$$\mathbf{q} \preceq \varphi \quad \text{and} \quad (\emptyset - \varphi)^{-1} \in \mathcal{A}_h.$$

Then, the resolvent  $(\emptyset - \mathbf{q})^{-1}$  is well-defined and belongs to  $\mathcal{A}_h$ . Furthermore, the unique solution  $\ell \in T((\mathbb{R}^d))$  to the linear equation

$$\ell = \mathbf{p} + \ell \mathbf{q}$$

belongs to  $\mathcal{A}_h$ .

In order to prove Theorem 3.9, we will first show that the resolvent is order-preserving. That is:

**Lemma 3.10.** *Let  $\mathbf{q}, \varphi \in T((\mathbb{R}^d))$  such that  $|\mathbf{q}^\emptyset|, \varphi^\emptyset < 1$  and  $\mathbf{q} \preceq \varphi$ , then*

$$(\emptyset - \mathbf{q})^{-1} \preceq (\emptyset - \varphi)^{-1}.$$

*Proof.* We will first rewrite the equation of the resolvent (2.5) as

$$(\emptyset - \mathbf{q})^{-1} = \sum_{k=0}^{\infty} \mathbf{q}^{\otimes k} = \sum_{k=0}^{\infty} \left( \sum_{m=0}^{\infty} \sum_{\mathbf{v} \in V_m} \mathbf{q}^{\mathbf{v}\mathbf{v}} \right)^{\otimes k} = \sum_{m=0}^{\infty} \sum_{\mathbf{v} \in V_m} f_{\mathbf{v}}(\mathbf{q}^{\mathbf{u}}, n(\mathbf{u}) \leq m) \mathbf{v}$$

where  $f_{\mathbf{v}}$  is a polynomial function of degree  $n$  for all  $\mathbf{v} \in V_n$ , which is independent of  $\mathbf{q}$  and has  $\frac{d^{n+1}-1}{d-1}$  variables  $\{\mathbf{q}^{\mathbf{u}} : |\mathbf{u}| \leq |\mathbf{v}|\}$ . We also notice that the coefficients of  $f_{\mathbf{v}}$  are all non-negative integers. (For example,  $f_{\mathbf{11}}(\mathbf{q}^{\mathbf{u}}, |\mathbf{u}| \leq 2) = 2\mathbf{q}^\emptyset \mathbf{q}^{\mathbf{11}} + (\mathbf{q}^{\mathbf{1}})^2$  and  $f_{\mathbf{12}}(\mathbf{q}^{\mathbf{u}}, |\mathbf{u}| \leq 2) = 2\mathbf{q}^\emptyset \mathbf{q}^{\mathbf{12}} + 2\mathbf{q}^{\mathbf{1}} \mathbf{q}^{\mathbf{2}}$ .) Now recall the Definition 3.8 where  $\mathbf{q} \preceq \varphi$  implies  $|\mathbf{q}^{\mathbf{v}}| \leq \varphi^{\mathbf{v}}$  for all  $\mathbf{v} \in V$ . Combined this with the fact that  $f$  has non-negative coefficients, it is easy to see that, for all  $\mathbf{v} \in V_m$  we have

$$|f_{\mathbf{v}}(\mathbf{q}^{\mathbf{u}}, n(\mathbf{u}) \leq m)| \leq f_{\mathbf{v}}(\varphi^{\mathbf{u}}, n(\mathbf{u}) \leq m),$$

which ends the proof.  $\square$

We can now prove Theorem 3.9.

*Proof of Theorem 3.9.* By assumption on  $\mathbf{q}$  and by Lemma 3.10,  $(\emptyset - \mathbf{q})^{-1}$  is well-defined and belongs to  $\mathcal{A}_h$ . Moreover, by Proposition 2.2

$$\ell = \mathbf{p}(\emptyset - \mathbf{q})^{-1}.$$

Finally, by assumption on  $\mathbf{p}$  and Proposition 3.7, it is clear that  $\ell \in \mathcal{A}_h$ .  $\square$

### 3.4 The space $\mathcal{A}_{\text{exp}}$ and exponentially dominated functionals

This final subsection is dedicated to the use of shuffle exponentials to dominate linear functionals. Shuffle exponentials allow for straightforward computations and it is easy to show their inclusion in  $\mathcal{A}_h$ , see Lemma 3.11. Moreover, they allow for the definition of the space  $\mathcal{A}_{\text{exp}} \subset \mathcal{A}_h$  of linear functionals dominated by a shuffle exponential. This space holds interesting properties, among them,  $\mathcal{A}_{\text{exp}}$  is closed under the projection operation, allowing for the inclusion of  $\mathcal{A}_{\text{exp}}$  in  $\mathcal{I}(\widehat{W})$ , see Corollary 3.13, making it much easier to prove that a linear functional is an Itô process.

First, we prove that  $e^{\sqcup \mathbf{b}} \in \mathcal{A}_h$ :

**Lemma 3.11.** *Let  $\mathbf{b} = \sum_{\mathbf{i} \in A_d} \mathbf{b}^{\mathbf{i}} \mathbf{i}$  with  $\mathbf{b}^{\mathbf{i}} \in \mathbb{R}$ , then*

$$e^{\sqcup \mathbf{b}} = (\emptyset - \mathbf{b})^{-1} \in \mathcal{A}_h.$$

*Proof.* First, the equality is ensured by Proposition 2.3, so it suffices to prove  $(\emptyset - \mathbf{b})^{-1} \in \mathcal{A}_h$ . For simplicity, we provide a proof for the case where  $d = 2$ , the extension to the general case follows along the same lines. For a word  $\mathbf{v} \in V$ , we recall the notations  $n(\mathbf{v})$  and  $x(\mathbf{v})$  that denote the length of  $\mathbf{v}$  and the number of  $\mathbf{1}$  in  $\mathbf{v}$  respectively. We also recall that  $h_T(\mathbf{v})$  depends on  $\mathbf{v}$  only through  $n(\mathbf{v})$  and  $x(\mathbf{v})$ , meaning that for any  $\mathbf{u} \in V$ ,  $n(\mathbf{v}) = n(\mathbf{u})$  and  $x(\mathbf{v}) = x(\mathbf{u})$  imply  $h_T(\mathbf{v}) = h_T(\mathbf{u})$ . Therefore, assuming  $\mathbf{b} = a\mathbf{1} + b\mathbf{2}$  and

$\ell = (\emptyset - (a\mathbf{1} + b\mathbf{2}))^{-1}$ , we can decompose  $\|\ell\|_T^{A_h}$  as:

$$\begin{aligned}
\|\ell\|_T^{A_h} &= \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} |\ell^{\mathbf{v}}| h_T(\mathbf{v}) \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^n \binom{n}{k} |a|^k |b|^{n-k} h_T(\mathbf{1}^{\otimes k} \mathbf{2}^{\otimes n-k}) \\
&= \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \binom{k+m}{k} |a|^k |b|^m h_T(\mathbf{1}^{\otimes k} \mathbf{2}^{\otimes m}) \\
&= \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{(k+m)!}{k! m!} |a|^k |b|^m \frac{C}{(k+m+1)^{\frac{1}{4}}} \frac{(2T)^{k+\frac{m}{2}}}{\sqrt{(2k+m-1)!}} \\
&= C \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \frac{|2Ta|^k}{k!} \frac{|\sqrt{2Tb}|^m}{\sqrt{m!}} (k+m+1)^{-\frac{1}{4}} \sqrt{\frac{(k+m)!^2}{m!(2k+m-1)!}}.
\end{aligned}$$

After remarking that  $(k+m+1)^{-\frac{1}{4}} \leq 1$  for all  $k, m \in \mathbb{N}$ , we can shift our attention to the square root on the right-hand side:

$$\frac{(k+m)!^2}{m!(2k+m-1)!} = \frac{\prod_{j=1}^k (m+j)}{\prod_{j=1}^{k-1} (k+m+j)} \leq k+m+1 \leq (k+1)(m+1).$$

This gives

$$\|\ell\|_T^{A_h} \leq C \sum_{m=0}^{\infty} \left( \sqrt{\frac{m+1}{m!}} |\sqrt{2Tb}|^m \right) \sum_{k=0}^{\infty} \left( \frac{\sqrt{k+1}}{k!} |2Ta|^k \right).$$

Now, thanks to Stirling's formula,

$$\frac{\sqrt{k+1}}{k!} \sim \left(\frac{e}{k}\right)^k \quad \text{and} \quad \sqrt{\frac{m+1}{m!}} \sim m^{\frac{1}{4}} \left(\frac{e}{m}\right)^{\frac{m}{2}}$$

so it is clear that both  $\sum_{k \geq 0} \frac{\sqrt{k+1}}{k!} x^k$  and  $\sum_{m \geq 0} \sqrt{\frac{m+1}{m!}} y^m$  have an infinite radius of convergence, hence proving

$$\|\ell\|_T^{A_h} < \infty.$$

□

Naturally, we will define the set  $\mathcal{A}_{\text{exp}}$  of all linear functionals that are *exponentially dominated*:

$$\mathcal{A}_{\text{exp}} := \left\{ \ell : \text{there exists } C \in \mathbb{R}_+ \text{ such that } \ell \preceq C e^{\sqcup\sqcup C \sum_{i \in A_d} \mathbf{i}} \right\}.$$

In dimension 1,  $\mathcal{A}_{\text{exp}}$  represents the power series that have a factorial decay. Obviously  $\mathcal{A}_{\text{exp}} \subset \mathcal{A}_h$  thanks to Lemma 3.11.

**Proposition 3.12.**  $\mathcal{A}_{\text{exp}}$  is closed under linear operations, shuffle product, concatenation, resolvent and projection. That is, let  $\ell, \varphi \in \mathcal{A}_{\text{exp}}$ , then

- (i)  $\alpha\ell + \beta\varphi \in \mathcal{A}_{\text{exp}}$  for  $\alpha, \beta \in \mathbb{R}$ ,
- (ii)  $\ell|_{\mathbf{u}} \in \mathcal{A}_{\text{exp}}$  for all  $\mathbf{u} \in V$ ,
- (iii)  $\ell \sqcup\sqcup \varphi \in \mathcal{A}_{\text{exp}}$ ,
- (iv)  $\ell\varphi \in \mathcal{A}_{\text{exp}}$ ,
- (v) If in addition  $\ell^{\emptyset} \neq 1$ ,  $(\emptyset - \ell)^{-1} \in \mathcal{A}_{\text{exp}}$ .

*Proof.* To simplify notations, let us define  $\mathbf{b} := \sum_{i \in A_d} \mathbf{i}$ . Moreover, let  $\alpha, \beta \in \mathbb{R}$  and  $C_\ell, C_\varphi > 1$ , and let  $\ell, \varphi \in \mathcal{A}_{\text{exp}}$  be such that  $\ell \preceq C_\ell e^{\sqcup\sqcup C_\ell \mathbf{b}}$  and  $\varphi \preceq C_\varphi e^{\sqcup\sqcup C_\varphi \mathbf{b}}$ .

- (i) Clearly, one has  $\alpha\ell + \beta\varphi \preceq (|\alpha|C_\ell + |\beta|C_\varphi)e^{\sqcup(C_\ell \vee C_\varphi)\mathbf{b}}$ ,
- (ii) Using Proposition 2.3, one also has  $\ell|_{\mathbf{u}} \preceq (C_\ell)^{n+1}e^{\sqcup C_\ell \mathbf{b}}$  for  $\mathbf{u} \in V_n$  and  $n \in \mathbb{N}$ ,
- (iii) Using [17, Lemma 2.10] on the shuffle product of shuffle exponentials, one can deduce

$$\ell \sqcup \varphi \preceq C_\ell e^{\sqcup C_\ell \mathbf{b}} \sqcup C_\varphi e^{\sqcup C_\varphi \mathbf{b}} \preceq (C_\ell C_\varphi) e^{\sqcup(C_\ell + C_\varphi)\mathbf{b}},$$

- (iv) Seeing that the first (and last) terms in the shuffle product are simple concatenations of the two functionals, one can conveniently dominate a concatenation by a shuffle product, i.e.

$$\ell\varphi \preceq |\ell| |\varphi| \preceq |\ell| \sqcup |\varphi| \preceq (C_\ell C_\varphi) e^{\sqcup(C_\ell + C_\varphi)\mathbf{b}},$$

where  $|\ell| := \sum_{\mathbf{v} \in V} |\ell^{\mathbf{v}}| \mathbf{v}$ ,

- (v) We can first assume  $\ell^\emptyset = 0$ , as  $(\emptyset - \ell)^{-1} = \frac{1}{1-\ell^\emptyset} \left( \emptyset - \frac{1}{1-\ell^\emptyset} (\ell - \ell^\emptyset \emptyset) \right)^{-1}$ , and  $C_\ell, C_\varphi > 1$  without loss of generality. It implies  $\ell \preceq C_\ell (e^{\sqcup C_\ell \mathbf{b}} - \emptyset) = (C_\ell)^2 \mathbf{b} e^{\sqcup C_\ell \mathbf{b}}$ . Moreover, by Lemma 3.10,

$$(\emptyset - \ell)^{-1} \preceq (\emptyset - (C_\ell)^2 \mathbf{b} e^{\sqcup C_\ell \mathbf{b}})^{-1}.$$

There thus only remains to make sure that, for any  $\alpha, \beta \in \mathbb{R}$ ,  $(\emptyset - \alpha \mathbf{b} e^{\sqcup \beta \mathbf{b}})^{-1} \in \mathcal{A}_{\text{exp}}$ . For this purpose, recalling that  $(\emptyset - \beta \mathbf{b})^{-1} (\emptyset - \beta \mathbf{b}) = \emptyset$ , we have that

$$\begin{aligned} \emptyset - (\alpha + \beta)\mathbf{b} &= \emptyset - \alpha \mathbf{b} (\emptyset - \beta \mathbf{b})^{-1} (\emptyset - \beta \mathbf{b}) - \beta \mathbf{b} \\ &= \left( \emptyset - \alpha \mathbf{b} (\emptyset - \beta \mathbf{b})^{-1} \right) (\emptyset - \beta \mathbf{b}), \end{aligned}$$

which directly implies from Proposition 2.3 that

$$(\emptyset - \alpha \mathbf{b} e^{\sqcup \beta \mathbf{b}})^{-1} = (\emptyset - \beta \mathbf{b}) (\emptyset - (\alpha + \beta)\mathbf{b})^{-1} = (\emptyset - \beta \mathbf{b}) e^{\sqcup(\alpha + \beta)\mathbf{b}}.$$

Finally applying (iv), we clearly see that the *rhs* is dominated, ending the proof. □

**Corollary 3.13.** *Let  $\mathbf{p}, \mathbf{q} \in \mathcal{A}_{\text{exp}}$ ,  $\mathbf{q}^\emptyset \neq 1$ , then the unique solution  $\ell \in T((\mathbb{R}^d))$  to the linear equation*

$$\ell = \mathbf{p} + \ell \mathbf{q}$$

*belongs to  $\mathcal{A}_{\text{exp}}$ . Moreover, it also belongs to  $\mathcal{I}(\widehat{W})$ .*

*Proof.* First, applying Proposition 3.12 (iv)-(v), we obtain  $\ell = \mathbf{p}(\emptyset - \mathbf{q})^{-1} \in \mathcal{A}_{\text{exp}}$ . Then, using Proposition 3.6 (ii), together with Proposition 3.12 (ii), it follows that  $\mathcal{A}_{\text{exp}} \subset \mathcal{I}(\widehat{W})$ . □

Corollary 3.13 essentially implies that, provided  $\mathbf{p}$  and  $\mathbf{q}$  are exponentially dominated, an easily verifiable condition in practice,

$$X_t := \left\langle \mathbf{p}(\emptyset - \mathbf{q})^{-1}, \widehat{W}_t \right\rangle$$

is well-defined and is an Itô process as in (3.8).

## 4 Signature representations of path-dependent processes

In this section, we provide our main signature representation formulae for linear path-dependent equations. We consider three categories:

- Linear Volterra equations: exact, time-independent representations for smooth enough kernels, and an approximation result for more general kernels, in Section 4.1,
- Certain linear delay equations: exact, time-independent representations for a weighted sum of exponential kernels, in Section 4.2,
- Gaussian Volterra processes: exact, time-dependent representations for possibly singular kernels, including the Riemann-Liouville fractional Brownian motion for a Hurst index  $H \in (0, 1)$ , in Section 4.3

In the sequel, we will consider the 2-dimensional semimartingale  $\widehat{W}_t = (t, W_t)$ , where  $W$  is a one-dimensional Brownian motion, we will discuss extensions for a  $d$ -dimensional Brownian motion in Remarks 4.6 and 4.8.

## 4.1 Linear Volterra equations

The first class we consider is linear Volterra equations of the form

$$Y_t = y + \int_0^t K_1(t-s)(a_1 + b_1 Y_s) ds + \int_0^t K_2(t-s)(a_2 + b_2 Y_s) dW_s, \quad (4.1)$$

for some real coefficients  $y, a_1, b_1, a_2, b_2$  and locally square-integrable kernels  $K_1, K_2$ . Since the drift and volatility coefficients are linear in  $Y$  and hence Lipschitz continuous, the stochastic integral equation (4.1) admits a unique strong solution  $Y$ , see for instance [4, Theorem 3.3].

In our next theorem, we provide an explicit infinite linear representation for the solution  $Y_t = \langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_t \rangle$  in terms of the signature elements of  $\widehat{W}$  and time-independent coefficients  $\boldsymbol{\ell}$ . For this, we will need the following structure on the kernels  $K_1$  and  $K_2$ :

$$K_1(u) = \int_{[0, \infty)} e^{-xu} \mu_1(dx) \quad \text{and} \quad K_2(u) = \int_{[0, \infty)} e^{-xu} \mu_2(dx), \quad (4.2)$$

for finite measures  $\mu_1$  and  $\mu_2$  and such that

$$\int_{[0, \infty)} x^n \mu_1(dx) + \int_{[0, \infty)} x^n \mu_2(dx) < M^n, \quad n \in \mathbb{N}, \quad (4.3)$$

for some constant  $M > 0$ .

In particular, this implies that  $K_1, K_2$  are infinitely continuously differentiable on  $[0, T]$ . For such smooth coefficients  $K_1, K_2$  the solution  $Y$  is even a semimartingale.

*Example 4.1.* Let us consider  $\mu(dx) = f(x)dx$  for a positive bounded measurable function  $f$  with compact support. It is easy to see that for such  $\mu$ , the assumptions (4.2)-(4.3) are satisfied. Another example is the case  $\mu(dx) = \sum_{k=1}^n w_k^n \delta_{x_k^n}(dx)$  and  $K(u) = \sum_{k=1}^n w_k^n e^{-x_k^n u}$  for  $w_k^n, x_k^n \in \mathbb{R}$ . ■

Before we state the theorem, we need to clarify how one can make sense of  $\int_{[0, \infty)} \boldsymbol{\ell}(x) \mu(dx)$  when  $\boldsymbol{\ell} : x \mapsto \boldsymbol{\ell}(x)$  is a function taking values in  $T((\mathbb{R}^d))$ .

**Definition 4.1** (Integral of parameterized linear functionals). *Fix a Borel measure  $\mu$  on  $[0, \infty)$ . We define  $L^1(T((\mathbb{R}^d)), \mu)$  to be the set of all the weakly integrable maps from  $([0, \infty), \mu)$  to the locally convex topological vector space  $T((\mathbb{R}^d))$ , i.e.*

$$L^1(T((\mathbb{R}^d)), \mu) := \left\{ \boldsymbol{\ell} : [0, \infty) \longrightarrow T((\mathbb{R}^d)) \text{ s.t. } \int_{[0, \infty)} |\boldsymbol{\ell}^{\mathbf{v}}(x)| \mu(dx) < \infty, \text{ for all } \mathbf{v} \in V \right\}.$$

Furthermore, we also define

$$L^1(\mathcal{A}, \mu) := \left\{ \boldsymbol{\ell} \in L^1(T((\mathbb{R}^d)), \mu) : \int_{[0, \infty)} \boldsymbol{\ell}(x) \mu(dx) \in \mathcal{A} \right\}.$$

*Example 4.2.* Take  $\boldsymbol{\ell}(x) = e^{\llbracket -x \mathbf{1} \rrbracket}$  and  $\mu(dx) = e^{-\frac{x^2}{2}} dx$ , then  $\boldsymbol{\ell} \in L^1(T((\mathbb{R}^d)), \mu)$  and thus  $\int_{[0, \infty)} \boldsymbol{\ell}(x) \mu(dx)$  is a well-defined element of  $T((\mathbb{R}^d))$ . Furthermore we can check that  $\int_{[0, \infty)} \boldsymbol{\ell}(x) \mu(dx) \in \mathcal{A}$ , showing that  $\boldsymbol{\ell} \in L^1(\mathcal{A}, \mu)$ . Now fixing  $\boldsymbol{\ell}(x) = e^{\llbracket -x \mathbf{1} \rrbracket}$ , for a general measure  $\mu$ , we can easily see that  $\boldsymbol{\ell} \in L^1(\mathcal{A}, \mu)$  if and only if  $\sum_{n=0}^{\infty} \int_{[0, \infty)} \frac{x^n}{n!} \mu(dx) < \infty$ . This partially explains why the assumption (4.3) arises. ■

**Theorem 4.2.** *Fix  $K_1, K_2$  satisfying (4.2)-(4.3). The solution  $Y$  to the linear Volterra equation (4.1) admits the time-independent signature representation*

$$Y_t = \left\langle \boldsymbol{\ell}^{\text{VOL}}, \widehat{\mathbb{W}}_t \right\rangle \text{ for every } t \leq T, \text{ a.s., with } \boldsymbol{\ell}^{\text{VOL}} = \mathbf{p}^{\text{VOL}} (\boldsymbol{\phi} - \mathbf{q}^{\text{VOL}})^{-1} \quad (4.4)$$

and

$$\begin{aligned}\mathbf{p}^{\text{VOL}} &:= a_1 \mathbf{1} \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_1(dx) + a_2 \left( \mathbf{2} - \frac{1}{2} b_2 K_2(0) \mathbf{1} \right) \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_2(dx) + y \mathfrak{o}, \\ \mathbf{q}^{\text{VOL}} &:= b_1 \mathbf{1} \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_1(dx) + b_2 \left( \mathbf{2} - \frac{1}{2} b_2 K_2(0) \mathbf{1} \right) \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_2(dx),\end{aligned}$$

and are such that  $\mathbf{p}^{\text{VOL}}, \mathbf{q}^{\text{VOL}}, \boldsymbol{\ell}^{\text{VOL}} \in \mathcal{A}_{\text{exp}}$ . In particular, they all belong to  $\mathcal{A}$  and  $\mathcal{I}(\widehat{W})$ .

*Proof.* The proof is given in Section 6. □

For  $b_2 = 0$  and  $K_1 = K_2 \equiv 1$ , one recovers from the representation in Theorem 4.2 a time-independent representation for the Ornstein-Uhlenbeck process as shown in the following example. For a time-dependent representation of the same process, we refer to Example 4.10 below.

*Example 4.3* (Ornstein-Uhlenbeck). Set  $b_2 = 0$ , and  $\mu_1(dx) = \mu_2(dx) = \delta_0(dx)$ , where  $\delta_0$  is the Dirac mass at 0. In this case,  $K_1(u) = K_2(u) = 1$  and equation (4.1) reads

$$dY_t = (a_1 + b_1 Y_t)dt + a_2 dW_t, \quad Y_0 = y \in \mathbb{R}.$$

meaning that  $Y$  is an Ornstein-Uhlenbeck process. Moreover, in this case,  $\int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_1(dx) = \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_2(dx) = \mathfrak{o}$ . So that

$$\mathbf{q}^{\text{VOL}} = b_1 \mathbf{1}, \quad \mathbf{p}^{\text{VOL}} = y \mathfrak{o} + a_1 \mathbf{1} + a_2 \mathbf{2}.$$

An application of Theorem 4.2 yields that the Ornstein-Uhlenbeck process  $Y$  can be written as

$$Y_t = \left\langle \boldsymbol{\ell}^{\text{OU}}, \widehat{\mathbb{W}}_t \right\rangle$$

with

$$\boldsymbol{\ell}^{\text{OU}} = (y \mathfrak{o} + a_1 \mathbf{1} + a_2 \mathbf{2}) (\mathfrak{o} - b_1 \mathbf{1})^{-1} = (y \mathfrak{o} + a_1 \mathbf{1} + a_2 \mathbf{2}) e^{\llbracket b_1 \mathbf{1} \rrbracket},$$

where the second equality follows from Proposition 2.3. ■

Furthermore, for  $a_1 = a_2 = 0$  and  $K_1 = K_2 \equiv 1$ , one recovers from the representation in Theorem 4.2 the explicit solution of the geometric Brownian motion:

*Example 4.4* (Geometric Brownian motion). Take  $a_1 = a_2 = 0$ , and  $\mu_1(dx) = \mu_2(dx) = \delta_0(dx)$ , where  $\delta_0$  is the Dirac mass at 0. In this case,  $K_1(u) = K_2(u) = 1$  and  $\int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_1(dx) = \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rrbracket} \mu_2(dx) = \mathfrak{o}$ .

Therefore  $\mathbf{q}^{\text{VOL}} = (b_1 - \frac{1}{2} b_2^2) \mathbf{1} + b_2 \mathbf{2}$ ,  $\mathbf{p}^{\text{VOL}} = y \mathfrak{o}$  and thus, using Proposition 2.3,

$$\boldsymbol{\ell}^{\text{VOL}} = y (\mathfrak{o} - (b_1 - \frac{1}{2} b_2^2) \mathbf{1} - b_2 \mathbf{2})^{-1} = y e^{\llbracket (b_1 - \frac{1}{2} b_2^2) \mathbf{1} + b_2 \mathbf{2} \rrbracket}.$$

It follows from the shuffle property 3.1 and the definition of the shuffle exponential in (2.7), that

$$Y_t = y \left\langle e^{\llbracket (b_1 - \frac{1}{2} b_2^2) \mathbf{1} + b_2 \mathbf{2} \rrbracket}, \widehat{\mathbb{W}}_t \right\rangle = y e^{\left\langle (b_1 - \frac{1}{2} b_2^2) \mathbf{1} + b_2 \mathbf{2}, \widehat{\mathbb{W}}_t \right\rangle} = y e^{(b_1 - \frac{1}{2} b_2^2)t + b_2 W_t},$$

which is the geometric Brownian motion, solution to

$$\frac{dY_t}{Y_t} = b_1 dt + b_2 dW_t, \quad Y_0 = y \in \mathbb{R}.$$

Upon initial inspection, our assumptions regarding  $K_1$  and  $K_2$  may appear limiting. However, the stability results of stochastic Volterra equations allow us to approximate solutions for broader classes of kernels. For example, we can approximate solutions for the singular fractional kernel  $K(t) = t^{H-1/2}$  with  $H \in (0, 1/2)$  using linear combinations of the signature elements, as demonstrated in the following corollary. ■

**Corollary 4.3.** Let  $K_1, K_2$  be locally square-integrable kernels and denote by  $Y$  the solution of (4.1). For  $n \in \mathbb{N}$  and  $i = 1, 2$ , let  $\mu_i^n(dx) = \sum_{k=1}^n w_k^{n,i} \delta_{x_k^{n,i}}(dx)$  for some  $w_k^{n,i}, x_k^{n,i} \in \mathbb{R}$ , and set  $K_1^n$  and  $K_2^n$  as in (4.2) and  $Y_t^n = \langle \ell_n^{\text{VOL}}, \widehat{\mathbb{W}}_t \rangle$  as in (4.4) both with  $\mu_1^n$  and  $\mu_2^n$  instead of  $\mu_1$  and  $\mu_2$ . Assume that

$$\int_0^T |K_1^n(s) - K_1(s)|^2 ds + \int_0^T |K_2^n(s) - K_2(s)|^2 ds \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (4.5)$$

Then,

$$\sup_{t \in [0, T]} \mathbb{E} \left[ \left| Y_t - \langle \ell_n^{\text{VOL}}, \widehat{\mathbb{W}}_t \rangle \right|^p \right] \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad p \in \mathbb{N}.$$

*Proof.* For  $n \in \mathbb{N}$ , it follows from Example 4.1 that  $\mu_1^n$  and  $\mu_2^n$  satisfy the assumptions (4.3). Hence, an application of Theorem 4.2 yields that  $Y_t^n = \langle \ell_n^{\text{VOL}}, \widehat{\mathbb{W}}_t \rangle$  solves the stochastic Volterra equation

$$Y_t^n = y + \int_0^t K_1^n(t-s) (a_1 + b_1 Y_s) ds + \int_0^t K_2^n(t-s) (a_2 + b_2 Y_s) dW_s.$$

Routine applications of Jensen's and Grownall's convolution inequalities, see for instance the proof of [4, Lemma 2.4] and [28, Lemma 9.8.2], yield that

$$\sup_{t \in [0, T]} \mathbb{E} [|Y_t - Y_t^n|^p] \leq C_p \left( \left( \int_0^T |K_1^n(s) - K_1(s)|^2 ds \right)^{\frac{p}{2}} + \left( \int_0^T |K_2^n(s) - K_2(s)|^2 ds \right)^{\frac{p}{2}} \right), \quad p \in \mathbb{N},$$

for some  $C_p > 0$ , which ends the proof.  $\square$

*Remark 4.5.* Different choices exist in the literature on Volterra processes for the parameters  $w_i^k, x_i^k$  to ensure the convergence of the weighted sum of exponentials towards the initial kernels in (4.5) whenever  $K_1$  and  $K_2$  are completely monotone functions on  $(0, \infty)$ , i.e. infinitely differentiable on  $(0, \infty)$  such that  $(-1)^n K^{(n)} \geq 0$ . For instance, for a Hurst index  $H \in (0, 1/2)$ , the singular fractional kernel  $t^{H-1/2}$  is completely monotone and admits the following representation (following Bernstein's theorem) as a Laplace transform:

$$\frac{t^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} = \int_{[0, \infty)} e^{-xt} \mu(dx) \quad \text{and} \quad \mu(dx) = \frac{x^{-H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})\Gamma(\frac{1}{2}-H)} dx.$$

For  $n \geq 1$  and  $r_n > 1$ , such that

$$r_n \downarrow 1 \quad \text{and} \quad n \ln r_n \rightarrow \infty, \quad \text{as } n \rightarrow \infty,$$

with the following parametrization for the weights and the mean reversion

$$w_k^n = \frac{(r_n^{1-\alpha} - 1) r_n^{(\alpha-1)(1+\frac{\alpha}{2})}}{\Gamma(\alpha)\Gamma(2-\alpha)} r_n^{(1-\alpha)k} \quad \text{and} \quad x_k^n = \frac{1-\alpha}{2-\alpha} \frac{r_n^{2-\alpha} - 1}{r_n^{1-\alpha} - 1} r_n^{k-1-\frac{\alpha}{2}}, \quad \text{for } k = 1, \dots, n,$$

where  $\alpha := H + 1/2$  for some  $H \in (0, 1/2)$ , one can obtain the convergence of  $\sum_{k=1}^n c_k^n e^{-x_k^n t}$  towards the fractional kernel in the sense of (4.5), see [1, 2] for more details.  $\blacksquare$

*Remark 4.6.* Theorem 4.2 can be readily extended to the case where the Volterra equation is driven by a  $(d-1)$ -dimensional Brownian motion  $W$  of the form

$$Y_t = y + \int_0^t K_{\mathbf{1}}(t-s) (a_{\mathbf{1}} + b_{\mathbf{1}} Y_s) ds + \sum_{\mathbf{j} \in \{\mathbf{2}, \dots, \mathbf{d}\}} \int_0^t K_{\mathbf{j}}(t-s) (a_{\mathbf{j}} + b_{\mathbf{j}} Y_s) dW_s^{\mathbf{j}}.$$

Here each  $K_{\mathbf{i}}$  is of the form  $K_{\mathbf{i}}(u) = \int_{[0, \infty)} e^{-xu} \mu_{\mathbf{i}}(dx)$  with a finite measure  $\mu_{\mathbf{i}}$  satisfying (4.3) with the same constant  $M$ , for  $\mathbf{i} \in \{\mathbf{1}, \dots, \mathbf{d}\}$ . Then, similar to Theorem 4.2, we obtain the representation

$$Y_t = \langle \ell^{\text{VOL}}, \widehat{\mathbb{W}}_t \rangle, \quad \ell^{\text{VOL}} = \mathbf{p}^{\text{VOL}} (\boldsymbol{\phi} - \mathbf{q}^{\text{VOL}})^{-1},$$

while  $\mathbf{p}^{\text{VOL}}$  and  $\mathbf{q}^{\text{VOL}}$  are updated to:

$$\begin{aligned}\mathbf{p}^{\text{VOL}} &:= \sum_{\mathbf{i} \in \{1, \dots, \mathbf{d}\}} a_{\mathbf{i}} \mathbf{i} \int_{[0, \infty)} e^{\llbracket \mathbf{i} \rrbracket - x \mathbf{1}} \mu_{\mathbf{i}}(dx) - \frac{1}{2} \sum_{\mathbf{j} \in \{2, \dots, \mathbf{d}\}} a_{\mathbf{j}} b_{\mathbf{j}} K_{\mathbf{j}}(0) \mathbf{1} \int_{[0, \infty)} e^{\llbracket \mathbf{j} \rrbracket - x \mathbf{1}} \mu_{\mathbf{j}}(dx) + y \emptyset, \\ \mathbf{q}^{\text{VOL}} &:= \sum_{\mathbf{i} \in \{1, \dots, \mathbf{d}\}} b_{\mathbf{i}} \mathbf{i} \int_{[0, \infty)} e^{\llbracket \mathbf{i} \rrbracket - x \mathbf{1}} \mu_{\mathbf{i}}(dx) - \frac{1}{2} \sum_{\mathbf{j} \in \{2, \dots, \mathbf{d}\}} b_{\mathbf{j}}^2 K_{\mathbf{j}}(0) \mathbf{1} \int_{[0, \infty)} e^{\llbracket \mathbf{j} \rrbracket - x \mathbf{1}} \mu_{\mathbf{j}}(dx).\end{aligned}$$

■

## 4.2 Linear delay equations

A second equation that we consider is a delay equation with linear delays both in the drift and in the volatility parts. To keep notations simple and highlight the main ideas, we restrict to weighted sum of exponential kernels:

$$\begin{aligned}dZ_t &= \left( a_1 + b_1 Z_t + \int_0^t K_1(t-s) Z_s ds \right) dt + \left( a_2 + b_2 Z_t + \int_0^t K_2(t-s) Z_s ds \right) dW_t, \\ Z_0 &= z \in \mathbb{R},\end{aligned}\tag{4.6}$$

where

$$K_i(t) = \sum_{k=1}^{n_i} w_i^k e^{x_i^k t},$$

with  $a_i, b_i, w_i^k, x_i^k \in \mathbb{R}, n_i \in \mathbb{N}$  for  $i = 1, 2$ . Again, it is straightforward to obtain the existence and uniqueness of a strong solution  $Z$  since the coefficients are linear. We stress that  $Z$  cannot be written as a Volterra equation in the sense of (4.1) unless  $K_2$  is zero.

**Theorem 4.4.** *The solution  $Z$  to the linear delay equation (4.6) admits the time-independent signature representation*

$$Z_t = \left\langle \boldsymbol{\ell}^{\text{DE}}, \widehat{\mathbb{W}}_t \right\rangle \text{ for every } t \leq T, \text{ a.s., with } \boldsymbol{\ell}^{\text{DE}} = \mathbf{p}^{\text{DE}} (\boldsymbol{\phi} - \mathbf{q}^{\text{DE}})^{-1},\tag{4.7}$$

and

$$\begin{aligned}\mathbf{p}^{\text{DE}} &:= (a_1 - \frac{1}{2} b_2 a_2) \mathbf{1} + a_2 \mathbf{2} + z \emptyset, \\ \mathbf{q}^{\text{DE}} &:= (b_1 - \frac{1}{2} b_2^2) \mathbf{1} + b_2 \mathbf{2} + \left( \sum_{k=1}^{n_1} w_1^k \mathbf{1} e^{\llbracket \mathbf{1} \rrbracket x_1^k \mathbf{1}} - \frac{1}{2} b_2 \sum_{k=1}^{n_2} w_2^k \mathbf{1} e^{\llbracket \mathbf{1} \rrbracket x_2^k \mathbf{1}} \right) \mathbf{1} + \left( \sum_{k=1}^{n_2} w_2^k \mathbf{1} e^{\llbracket \mathbf{1} \rrbracket x_2^k \mathbf{1}} \right) \mathbf{2},\end{aligned}$$

and are such that  $\mathbf{p}^{\text{DE}}, \mathbf{q}^{\text{DE}}, \boldsymbol{\ell}^{\text{DE}} \in \mathcal{A}_{\text{exp}}$ . In particular, they all belong to  $\mathcal{A}$  and  $\mathcal{I}(\widehat{W})$ .

*Proof.* The proof is given in Section 7. □

*Remark 4.7.* For more general kernels  $K_1, K_2$  a similar stability argument as in Corollary 4.3 allows to approximate the solution to the delay equation by linear combination of the signatures. ■

*Remark 4.8.* Similar to Remark 4.6, Theorem 4.4 can also be extended to the case where the Delayed equation is driven by a  $(d-1)$ -dimensional Brownian motion  $W$  of the form:

$$\begin{aligned}dZ_t &= \left( a_{\mathbf{1}} + b_{\mathbf{1}} Z_t + \int_0^t K_{\mathbf{1}}(t-s) Z_s ds \right) dt + \sum_{\mathbf{j} \in \{2, \dots, \mathbf{d}\}} \left( a_{\mathbf{j}} + b_{\mathbf{j}} Z_t + \int_0^t K_{\mathbf{j}}(t-s) Z_s ds \right) dW_t^{\mathbf{j}}, \\ Z_0 &= z \in \mathbb{R},\end{aligned}$$

where

$$K_{\mathbf{i}}(t) = \sum_{k=1}^{n_{\mathbf{i}}} w_{\mathbf{i}}^k e^{x_{\mathbf{i}}^k t}.$$

Then similar to Theorem 4.4, we have the representation:

$$Z_t = \left\langle \boldsymbol{\ell}^{\text{DE}}, \widehat{\mathbb{W}}_t \right\rangle, \quad \boldsymbol{\ell}^{\text{DE}} = \mathbf{p}^{\text{DE}} (\boldsymbol{\phi} - \mathbf{q}^{\text{DE}})^{-1},$$

while  $\mathbf{p}^{\text{DE}}$  and  $\mathbf{q}^{\text{DE}}$  are updated to:

$$\begin{aligned}\mathbf{p}^{\text{DE}} &= \sum_{\mathbf{i} \in \{1, \dots, \mathbf{d}\}} a_{\mathbf{i}} \mathbf{i} - \frac{1}{2} \sum_{\mathbf{j} \in \{2, \dots, \mathbf{d}\}} a_{\mathbf{j}} b_{\mathbf{j}} \mathbf{1} + z \emptyset, \\ \mathbf{q}^{\text{DE}} &= \sum_{\mathbf{i} \in \{1, \dots, \mathbf{d}\}} b_{\mathbf{i}} \mathbf{i} - \frac{1}{2} \sum_{\mathbf{j} \in \{2, \dots, \mathbf{d}\}} b_{\mathbf{j}}^2 \mathbf{1} + \sum_{\mathbf{i} \in \{1, \dots, \mathbf{d}\}} \sum_{k=1}^{n_{\mathbf{i}}} w_{\mathbf{i}}^k \mathbf{1} e^{\llbracket \mathbf{x}_{\mathbf{i}}^k \rrbracket} \mathbf{i} - \frac{1}{2} \sum_{\mathbf{j} \in \{2, \dots, \mathbf{d}\}} b_{\mathbf{j}} \sum_{k=1}^{n_{\mathbf{j}}} w_{\mathbf{j}}^k \mathbf{1} e^{\llbracket \mathbf{x}_{\mathbf{j}}^k \rrbracket} \mathbf{1}.\end{aligned}$$

■

### 4.3 Gaussian Volterra processes

The final class that we consider are continuous Gaussian Volterra processes of the form

$$X_t = \int_0^t K(t-s) dW_s, \quad (4.8)$$

for a locally square-integrable deterministic kernel  $K$  that is infinitely differentiable on  $(0, T]$  and satisfies

$$\left\{ \begin{array}{l} \sum_{n=0}^{\infty} |K^{(n)}(t)| \frac{t^{n+\frac{1}{2}}}{n! \sqrt{2n+1}} < \infty \quad \text{and} \quad \int_0^t \left( \sum_{n=0}^{\infty} |K^{(n)}(t)| \frac{s^n}{n!} \right)^2 ds < \infty, \quad t \in (0, T], \\ \int_0^T \left( \sum_{n=0}^{\infty} |K^{(n)}(t)| \frac{t^{n+\frac{1}{2}}}{n! \sqrt{2n+1}} \right) dt < \infty, \end{array} \right. \quad (4.9)$$

where  $K^{(n)}$  denotes the  $n$ -th derivative on  $(0, T]$ .

We note that here, and in contrast with Theorem 4.2,  $K$  can have a singularity at 0, as shown in Example 4.11 for the Riemann-Liouville fractional Brownian motion. For such kernels, Corollary 4.3 already provides an approximation result by infinite linear combinations of signature elements of  $\widehat{W}$  with time-independent coefficients  $\ell$ . More interestingly, the following theorem shows that exact infinite linear representation of  $X$  are possible but with time-dependent coefficients  $\ell_t$ . Time-dependent representations seemed crucial for us to obtain exact representations for singular kernels.

**Theorem 4.5.** *Fix  $K : (0, T] \rightarrow \mathbb{R}$  satisfying (4.9), the Gaussian Volterra process  $X$  defined in (4.8) admits a time-dependent signature representation*

$$X_t = \left\langle \ell_t^{\text{GV}}, \widehat{W}_t \right\rangle, \quad \text{a.s.}$$

for every  $t \leq T$ , with  $\ell_t^{\text{GV}}$  given by

$$\ell_t^{\text{GV}} = \begin{cases} \sum_{n=0}^{\infty} K^{(n)}(t) (-1)^n \mathbf{1}^{\otimes n} \mathbf{2} & \text{if } t \in (0, T], \\ 0 & \text{else.} \end{cases} \quad (4.10)$$

*Proof.* The proof is given in Section 8. □

*Remark 4.9.* Unlike the representations in Theorems 4.2 and 4.4, here  $\left\langle \ell_t^{\text{GV}}, \widehat{W}_t \right\rangle$  is a modification of  $X_t$  and it is not necessarily indistinguishable from the process  $X$ . ■

The following example shows that the exponential kernel  $K(t) = e^{-\kappa t}$  for  $\kappa > 0$  satisfies the assumption of Theorem 4.5. This provides a time-dependent representation of the Ornstein-Uhlenbeck process, recall Example 4.3 for the time-independent representation.

*Example 4.10* (Ornstein-Uhlenbeck process). Let

$$X_t = \int_0^t e^{-\kappa(t-s)} dW_s = \int_0^t K(t-s) dW_s$$

with  $K(t) = e^{-\kappa t}$  and  $\kappa \in \mathbb{R}$ . We first check that  $K$  satisfies (4.9). Clearly  $K$  is infinitely differentiable on  $[0, T]$  such that  $|K^{(n)}(t)| = |\kappa|^n K(t)$ . Therefore,

$$\int_0^t \left( \sum_{n=0}^{\infty} \frac{|K^{(n)}(t)| s^n}{n!} \right)^2 ds = \int_0^t \left( K(t) \sum_{n=0}^{\infty} \frac{(|\kappa| s)^n}{n!} \right)^2 ds = K(t)^2 \int_0^t e^{2|\kappa|s} ds < \infty.$$

Moreover

$$\sum_{n=0}^{\infty} \frac{|K^{(n)}(t)|}{n! \sqrt{2n+1}} t^{n+\frac{1}{2}} = \sqrt{t} e^{-\kappa t} \sum_{n=0}^{\infty} \frac{(|\kappa|t)^n}{n!} \frac{1}{\sqrt{2n+1}} \leq \sqrt{T} e^{|\kappa|T} \sum_{n=0}^{\infty} \frac{(|\kappa|t)^n}{n!} = \sqrt{T} e^{2|\kappa|T} < \infty.$$

Thus

$$\int_0^T \left( \sum_{n=0}^{\infty} \frac{|K^{(n)}(t)|}{n! \sqrt{2n+1}} t^{n+\frac{1}{2}} \right) dt \leq T^{\frac{3}{2}} e^{2|\kappa|T} < \infty.$$

Finally, Theorem 4.5 yields the signature representation  $X_t = \langle \ell_t^{\text{OU}}, \widehat{\mathbb{W}}_t \rangle$ , with

$$\ell_t^{\text{OU}} = e^{-\kappa t} \sum_{n=0}^{\infty} \kappa^n \mathbf{1}^{\otimes n} \mathbf{2}.$$

Note that we can more compactly write it as

$$\ell_t^{\text{OU}} = e^{\llbracket -\kappa(t\mathbf{e}-1) \rrbracket} \mathbf{2},$$

which is the algebraic translation of the probabilistic expression

$$X_t = \int_0^t e^{-\kappa(t-s)} dW_s.$$

The following example shows that the fractional kernel  $K(t) = t^{H-1/2}$  for  $H \in (0, 1)$ , with singularity at 0 for  $H < 1/2$ , satisfies the assumption of Theorem 4.5. This provides a representation of the Riemann-Liouville fractional Brownian motion, and shows that infinite linear combinations of signature elements are not always semimartingales, recall Remark 3.8. ■

*Example 4.11* (Riemann-Liouville fractional Brownian motion). Let

$$X_t = \frac{1}{\Gamma(H + \frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} dW_s = \int_0^t K(t-s) dW_s,$$

with  $K(t) = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})}$ . We first check that  $K$  satisfies (4.9). Clearly  $K$  is analytical on  $(0, T]$  for  $H \in (0, 1)$  :

$$K^{(n)}(t) = K(t) t^{-n} (H - \frac{1}{2})^{\underline{n}} = K(t) (-t)^{-n} (\frac{1}{2} - H)^{\bar{n}}$$

where  $x^{\underline{n}} = \prod_{k=0}^{n-1} (x-k)$  and  $x^{\bar{n}} = \prod_{k=0}^{n-1} (x+k)$  denote the falling and rising factorials respectively.

Furthermore, one can then remark that

$$|K^{(n)}(t)| = \begin{cases} K^{(n)}(t) (-1)^n & \text{if } H < \frac{1}{2}, \\ K^{(n)}(t) (-1)^{n+1} & \text{else.} \end{cases}$$

Therefore

$$\int_0^t \left( \sum_{n=0}^{\infty} \frac{|K^{(n)}(t)| s^n}{n!} \right)^2 ds = \int_0^t \left( K(t) \sum_{n=0}^{\infty} (H - \frac{1}{2})^{\underline{n}} \frac{1}{n!} \left( \frac{-s}{t} \right)^n \right)^2 ds = \int_0^t K(t-s)^2 ds < \infty,$$

where the last equality comes from the Taylor expansion of  $(1-x)^{H-1/2}$  which converges for  $|x| < 1$ . Moreover we have

$$|K^{(n)}(t)| = K(t) \frac{\Gamma(n + \frac{1}{2} - H)}{\Gamma(\frac{1}{2} - H)} t^{-n} = O\left( \binom{n}{t}^{\frac{1}{2}-H} (n-1)! t^{-n} \right), \quad n > 0,$$

therefore

$$\sum_{n=0}^{\infty} \frac{|K^{(n)}(t)|}{n! \sqrt{2n+1}} t^{n+\frac{1}{2}} \leq C t^H \sum_{n=1}^{\infty} n^{-1-H} < \infty.$$

And

$$\int_0^T \left( \sum_{n=0}^{\infty} \frac{|K^{(n)}(t)|}{n! \sqrt{2n+1}} t^{n+\frac{1}{2}} \right) dt \leq \frac{CT^{H+1}}{H+1} \sum_{n=1}^{\infty} n^{-1-H} < \infty.$$

Finally, Theorem 4.5 yields the signature representation  $X_t = \langle \boldsymbol{\ell}_t^{\text{RL}}, \widehat{\mathbb{W}}_t \rangle$ , with

$$\boldsymbol{\ell}_t^{\text{RL}} = \frac{t^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \sum_{n=0}^{\infty} t^{-n} \left(\frac{1}{2} - H\right)^{\bar{n}} \mathbf{1}^{\otimes n} \mathbf{2}, \quad (4.11)$$

where  $(\cdot)^{\bar{n}}$  is the rising factorial. Remarking that  $(\frac{1}{2} - H)^{\bar{n}} = 0$  for all  $n > 0$  when  $H = \frac{1}{2}$  gives  $\boldsymbol{\ell}_t = \mathbf{2}$ , the coordinate of  $W_t$  in  $\widehat{\mathbb{W}}_t$ . Note that we can also formally write it as

$$\boldsymbol{\ell}_t^{\text{RL}} = \frac{1}{\Gamma(H+\frac{1}{2})} (t\boldsymbol{\emptyset} - \mathbf{1})^{\sqcup H - \frac{1}{2}} \mathbf{2},$$

which is the algebraic translation of the probabilistic expression

$$X_t = \frac{1}{\Gamma(H+\frac{1}{2})} \int_0^t (t-s)^{H-\frac{1}{2}} dW_s.$$

■

## 4.4 Numerical illustrations

In this subsection, we provide numerical implementations of the truncated form of the signature processes  $X^{\leq M}$  defined by

$$X_t^{\leq M} := \langle \boldsymbol{\ell}, \widehat{\mathbb{W}}_t^{\leq M} \rangle, \quad M \geq 0,$$

where  $\boldsymbol{\ell} \in T(\mathbb{R}^2)$  and  $\widehat{\mathbb{W}}_t^{\leq M} : [0, T] \rightarrow T^M(\mathbb{R}^2)$  is defined in (2.11), i.e. its  $M$  first levels coincide with  $\widehat{\mathbb{W}}_t$  and everything else is set to 0.

### 4.4.1 Shifted Riemann-Liouville fractional Brownian motion

Figure 1 and Table 1 display simulations of a shifted Riemann-Liouville fractional Brownian motion  $X^\varepsilon$  together with mean-squared errors, i.e.

$$X_t^\varepsilon = \frac{1}{\Gamma(H+\frac{1}{2})} \int_0^t (t+\varepsilon-s)^{H-\frac{1}{2}} dW_s,$$

and its linear representation counterpart  $\boldsymbol{\ell}_{t+\varepsilon}^{\text{RL}}$ , following from (4.11), that is

$$\boldsymbol{\ell}_{t+\varepsilon}^{\text{RL}} = \frac{(t+\varepsilon)^{H-\frac{1}{2}}}{\Gamma(H+\frac{1}{2})} \sum_{n=0}^{\infty} (t+\varepsilon)^{-n} \left(\frac{1}{2} - H\right)^{\bar{n}} \mathbf{1}^{\otimes n} \mathbf{2} = \frac{1}{\Gamma(H+\frac{1}{2})} ((t+\varepsilon)\boldsymbol{\emptyset} - \mathbf{1})^{\sqcup H - \frac{1}{2}} \mathbf{2},$$

for several truncation orders  $M \in \mathbb{N}$ . The reason for shifting the kernel is to speed up the convergence as it can be quite slow for small values of  $H$  when  $\varepsilon = 0$ .

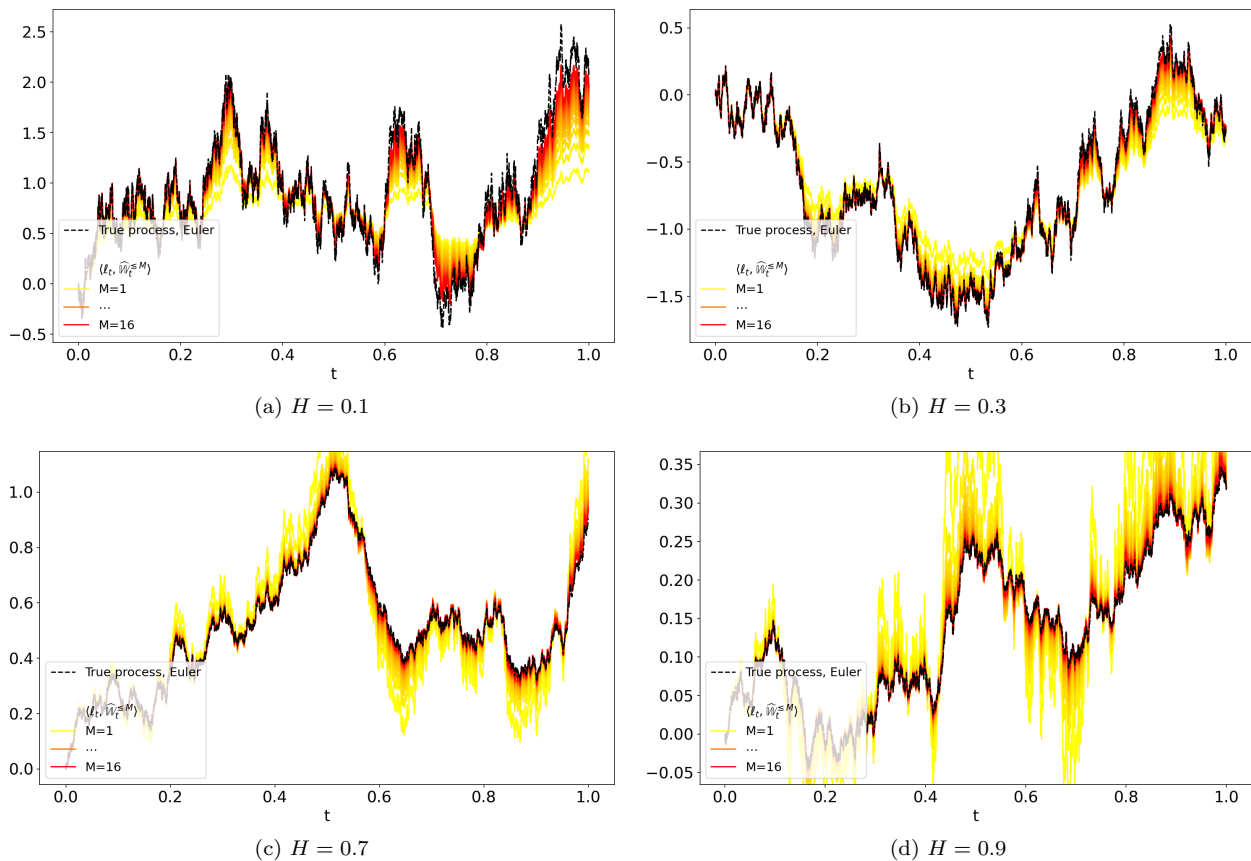


Figure 1: Trajectories of a shifted Riemann-Liouville fractional Brownian motion (black) against their truncated time-dependent linear representation (4.11), i.e.  $\langle \ell_{t+\varepsilon}^{\text{RL}}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$ , for several truncation orders  $M$  and  $\varepsilon = 1/52$ .

	H=0.1	H=0.3	H=0.7	H=0.9
M=2	9.915e-02	1.600e-02	4.015e-03	6.217e-03
M=4	4.735e-02	6.197e-03	9.022e-04	9.931e-04
M=8	1.851e-02	1.992e-03	1.819e-04	1.528e-04
M=16	5.298e-03	4.712e-04	2.793e-05	1.848e-05

Table 1: Mean squared error between the shifted Riemann-Liouville fractional Brownian motion and its truncated time-dependent linear representation (4.11), i.e.  $\langle \ell_{t+\varepsilon}^{\text{RL}}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$ , for several truncation orders  $M$  and  $\varepsilon = 1/52$ , averaged across 100,000 simulations over 1000 time-steps.

We can see in Table 1 and Figure 1 the clear convergence of the truncated linear representation for several values of  $H$ . The smaller  $H$  is, the slower the convergence. We can also see that the fit deteriorates with the time variable  $t$ .

#### 4.4.2 Linear delay process

Figure 2 and Table 2 display simulations of a delay equation and its truncated representations together with mean-squared errors following Theorem 4.4. This convergence is shown for two sets of parameters: (a) has  $b_1 = 0$  and  $x_i > 0$  whereas (b) has  $b_2 = 0$  and  $x_i < 0$ .

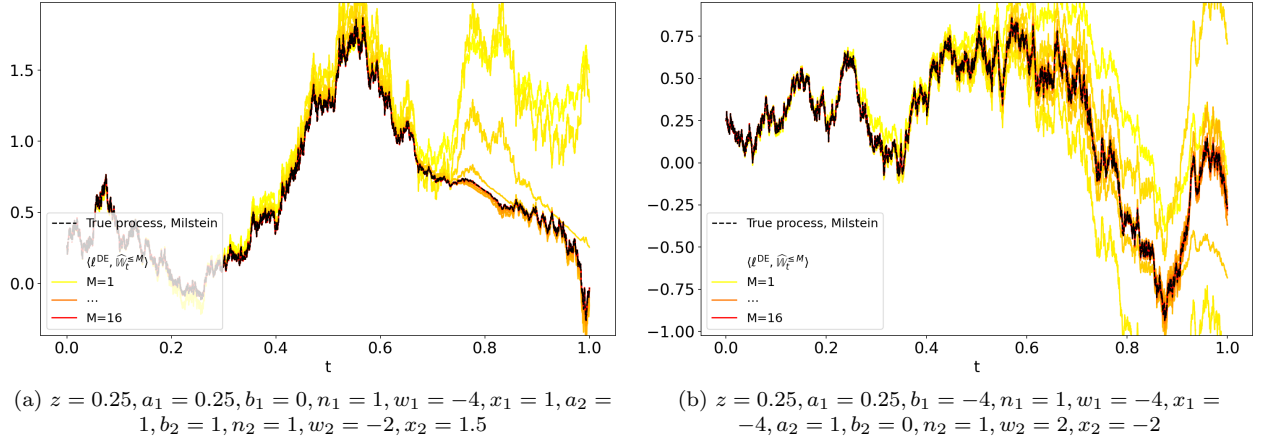


Figure 2: Trajectories of a delay process (black) against their truncated time-independent linear representation (4.7), i.e.  $\langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$ , for several truncation orders  $M$ .

	(a)	(b)
M=2	4.808e-01	4.405e-01
M=4	4.015e-02	2.795e-01
M=8	4.153e-04	2.511e-02
M=16	7.584e-07	2.739e-07

Table 2: Mean squared error between the delay process and its truncated time-independent linear representation (4.7), i.e.  $\langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$ , for several truncation orders  $M$ , averaged across 100,000 simulations over 1000 time-steps.

We can see in Table 2 and Figure 2 the clear convergence of the truncated linear representation for the delay equation process.

Recall that here since  $\ell^{\text{DE}} \in \mathcal{A}_h$ , the inequality (3.13) could in principle yield a control of the strong error, i.e. for every  $\epsilon > 0$ , there exists  $M_\epsilon \in \mathbb{N}$  such that

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left| \langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_t^{\leq M_\epsilon} \rangle - \langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_t \rangle \right| \right] \leq \epsilon.$$

For related numerical illustrations with other processes, we refer to Abi Jaber and Gérard [3].

#### 4.4.3 Linear Volterra process

For  $n \in \mathbb{N}$ , we consider a linear Volterra process  $Y^n$  solution to the Volterra equation (4.1), where the kernel is given by a sum of  $n$  weighted exponentials

$$K_n^{\alpha, \epsilon}(t) = \sum_{i=1}^n w_i e^{-x_i t}. \quad (4.12)$$

We denote by  $\ell_n^{\text{VOL}}$  the exact coefficients given from the signature representation in Theorem 4.2 and we consider  $\langle \ell_n^{\text{VOL}}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$  its truncated exact representation at level  $M \geq 1$ . The parameters  $(x_i, w_i)_{i=1, \dots, n}$  are chosen in such a way to optimally approximate the  $\epsilon$ -shifted fractional kernel

$$K^{\alpha, \epsilon}(t) := \frac{(t + \epsilon)^{\alpha-1}}{\Gamma(\alpha)} = \int_0^\infty e^{yt} \mu^{\alpha, \epsilon}(dy), \quad \text{with} \quad \mu^{\alpha, \epsilon}(dy) := \frac{y^{-\alpha} e^{-y\epsilon}}{\Gamma(\alpha)\Gamma(1-\alpha)} dy.$$

Given a partition  $(\eta_i)_{0 \leq i \leq n}$  where  $0 = \eta_0 < \eta_1 < \dots < \eta_n < \infty$ , we can readily compute its corresponding points  $x_i$  and weights  $w_i$

$$w_i = \int_{\eta_{i-1}}^{\eta_i} \mu^{\alpha, \varepsilon}(dy) = \varepsilon^{\alpha-1} \frac{\Gamma(1-\alpha; \varepsilon \eta_{i-1}, \varepsilon \eta_i)}{\Gamma(\alpha)\Gamma(1-\alpha)},$$

$$x_i = \frac{1}{w_i} \int_{\eta_{i-1}}^{\eta_i} y \mu^{\alpha, \varepsilon}(dy) = \frac{\varepsilon^{\alpha-2}}{w_i} \frac{\Gamma(2-\alpha; \varepsilon \eta_{i-1}, \varepsilon \eta_i)}{\Gamma(\alpha)\Gamma(1-\alpha)},$$

where  $\Gamma(z; a, b) = \int_a^b t^{z-1} e^{-t} dt$  is the incomplete gamma function. We chose a geometric partition  $\eta_i = (r_n)^{i-n/2}$  for some  $L^2$ -optimal  $r_n > 0$  as in [1, 2, 13], i.e.

$$r_n = \arg \min_r \|K_n^{\alpha, \varepsilon} - K^{\alpha, \varepsilon}\|_{L^2(\tau, T)}$$

$$= \arg \min_r \left( \sum_{i,j=1}^n w_i w_j \frac{e^{-(x_i+x_j)\tau} - e^{-(x_i+x_j)T}}{x_i + x_j} - 2 \sum_{i=1}^n w_i (x_i)^{-\alpha} e^{x_i \varepsilon} \frac{\Gamma(\alpha; (\tau + \varepsilon)x_i, (T + \varepsilon)x_i)}{\Gamma(\alpha)} \right).$$

In Figure 3, and later in Figure 4, we compare the truncated signature representation  $\langle \ell_n^{\text{VOL}}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$  of the multifactor approximation of the Volterra process  $Y^n$  with the kernel (4.12) to the Euler discretization of the (true) Volterra process  $Y$ , i.e.

$$Y_t = \int_0^t K^{\alpha, \varepsilon}(t-s) ((a_1 + b_1 Y_s) ds + ((a_2 + b_2 Y_s) dW_s)).$$

In particular, this provides a numerical illustration of the approximation result given in Corollary 4.3. Figure 3 shows that for different values of  $\alpha$ , the truncated signature of the multifactor approximation with  $n = 10$  factors yields very accurate sample paths, at least for short time horizons.

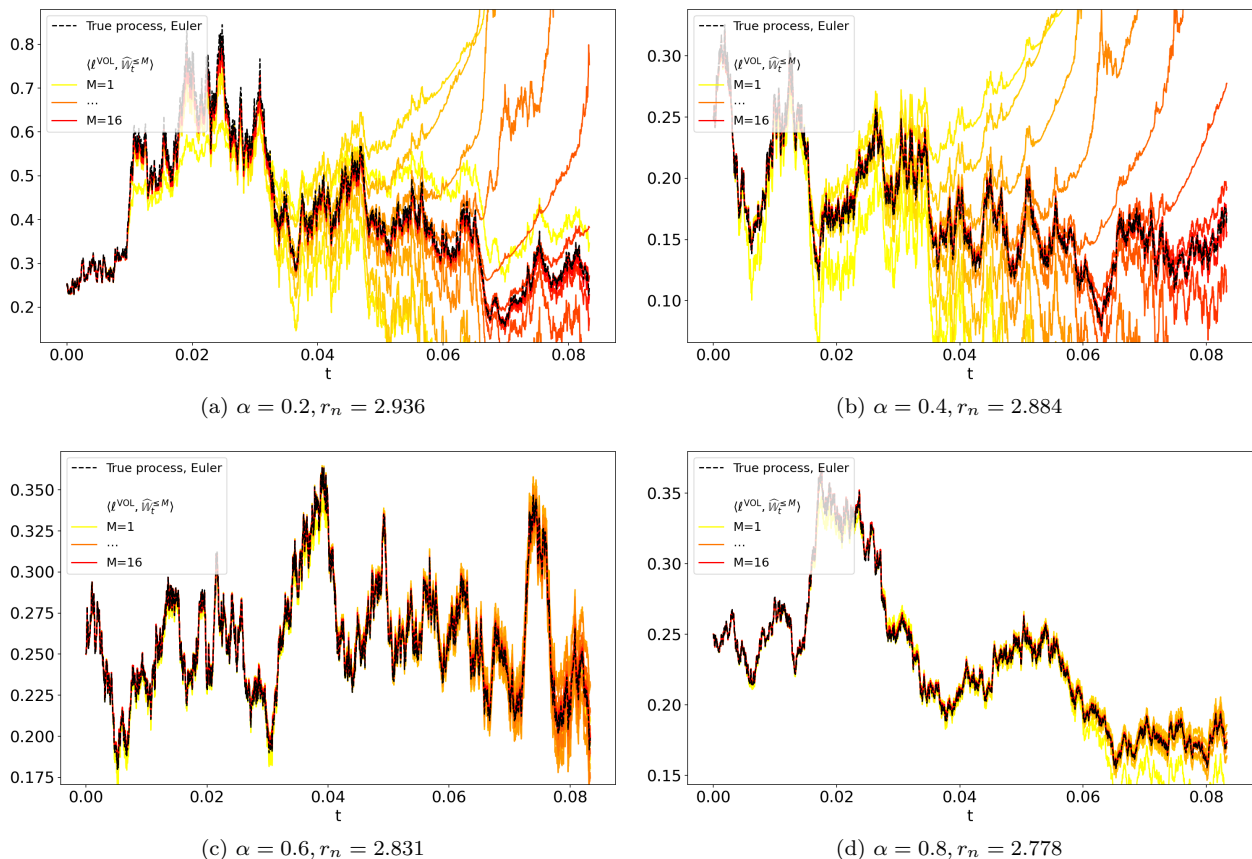


Figure 3: Trajectories of the shifted fractional Volterra process (black) against their truncated time-independent linear multifactor approximation representation (4.4), i.e.  $\langle \ell_n^{\text{VOL}}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$ , for several truncation orders  $M$  and  $y = 0.25, a_1 = 0.25, b_1 = -1, a_2 = 0, b_2 = 1, \varepsilon = 1/52, T = 1/12, \tau = 1/106, n = 10$ .

## 5 Conditional moments

As a key application of Section 4.1 and Section 4.2, our representation formulas provide explicit expressions for the conditional moments of Volterra and delay processes, which can be efficiently computed numerically without relying on Monte Carlo simulations.

The computation of unconditional and conditional moments is a central but notoriously difficult problem for non-Markovian processes. Our approach leverages the Markovianity of the signature of  $\widehat{\mathbb{W}}$  to make even conditional moments straightforward to obtain. The linear Volterra process (4.1) is a particular case of a polynomial Volterra process studied in [5], where unconditional moments were derived in terms of deterministic integral equations, without any treatment of conditional moments or numerical methods. To the best of our knowledge, this is the first time that explicit formulas are used to implement efficient numerics for both unconditional and conditional moments in this setting. Moreover, the delayed case (4.6) with nonzero  $b_2$  or  $K_2$  lies outside the polynomial Volterra family and thus cannot be recovered from [5], further underlining the significance of our representation formulas.

Theorem 5.2 establishes the conditional moment formulas combining our representation formulas in (4.4) and (4.7) with the celebrated Fawcett formula [21]:

$$\mathbb{E}[\widehat{\mathbb{W}}_t] = \mathcal{E}_t, \quad t \leq T,$$

with

$$\mathcal{E}_t := e^{\otimes (1 + \frac{1}{2} \mathbf{22})t}.$$

Before stating Theorem 5.2, we will generalize the projection defined in (2.3) against linear functionals of the form

$$\varphi = e^{\otimes a_1 \mathbf{1} + b_2 \mathbf{2}} = \sum_{m=0}^{\infty} \frac{1}{m!} (a_1 \mathbf{1} + a_2 \mathbf{2})^{\otimes m},$$

and show that under this framework it is the adjoint concatenation operation.

**Lemma 5.1.** *Let  $\ell \in \mathcal{A}_{\text{exp}}$  and  $\varphi = e^{\otimes a_1 \mathbf{1} + \sum_{i \in \{2, \dots, d\}} a_i \mathbf{i}}$  for some constants  $a_1, a_2, \dots, a_d \in \mathbb{R}$ . Then*

$$\ell|_{\varphi} := \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \varphi^{\mathbf{v}} \ell|_{\mathbf{v}} \in \mathcal{A}_{\text{exp}}$$

and

$$\langle \ell|_{\varphi}, \widehat{\mathbb{W}}_t \rangle = \langle \ell, \widehat{\mathbb{W}}_t \otimes \varphi \rangle, \quad t \leq T.$$

*Proof.* Without loss of generality, we will assume  $d = 2$ . Now, recall that for all  $C > 0$  and  $\mathbf{v} \in V_n$  with  $n \in \mathbb{N}$ ,  $e^{\sqcup C(\mathbf{1}+\mathbf{2})}|_{\mathbf{v}} \preceq C^n e^{\sqcup C(\mathbf{1}+\mathbf{2})}$ . This essentially means that we are only interested in the length  $n$  of  $\mathbf{v}$  when dominating projections of elements of  $\mathcal{A}_{\text{exp}}$ . Then, since by assumption on  $\ell$ , there exists  $C > 1$  such that  $\ell \preceq C e^{\sqcup C(\mathbf{1}+\mathbf{2})}$ , we can write

$$\begin{aligned} \ell|_{\varphi} &\preceq \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k=0}^m \binom{m}{k} |a_1|^{m-k} |a_2|^k C e^{\sqcup C(\mathbf{1}+\mathbf{2})}|_{\mathbf{1}^{\otimes m-k} \mathbf{2}^{\otimes 2k}} \\ &\preceq C e^{\sqcup C(\mathbf{1}+\mathbf{2})} \sum_{m=0}^{\infty} \sum_{k=0}^m \frac{|a_1|^{m-k}}{(m-k)!} \frac{|a_2|^k}{k!} C^{m+k} \\ &\preceq C e^{\sqcup C(\mathbf{1}+\mathbf{2})} e^{(C|a_1|+C^2|a_2|)}, \end{aligned}$$

which clearly shows  $\ell|_{\varphi} \in \mathcal{A}_{\text{exp}}$ . We can now decompose  $\langle \ell|_{\varphi}, \widehat{\mathbb{W}}_t \rangle$  as follows

$$\begin{aligned} \langle \ell|_{\varphi}, \widehat{\mathbb{W}}_t \rangle &= \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \varphi^{\mathbf{v}} \langle \ell|_{\mathbf{v}}, \widehat{\mathbb{W}}_t \rangle \\ &= \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \varphi^{\mathbf{v}} \sum_{m=0}^{\infty} \sum_{\mathbf{u} \in V_m} \ell^{\mathbf{u}\mathbf{v}} \widehat{\mathbb{W}}_t^{\mathbf{u}} \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{\mathbf{w} \in V_{m+n}} \ell^{\mathbf{w}} \sum_{\mathbf{v} \in V_n} \sum_{\mathbf{u} \in V_m} \widehat{\mathbb{W}}_t^{\mathbf{u}} \varphi^{\mathbf{v}} \mathbf{1}_{\{\mathbf{w}=\mathbf{u}\mathbf{v}\}} \\ &= \sum_{k=0}^{\infty} \sum_{n=0}^k \sum_{\mathbf{w} \in V_k} \ell^{\mathbf{w}} \sum_{\mathbf{v} \in V_n} \sum_{\mathbf{u} \in V_{k-n}} \widehat{\mathbb{W}}_t^{\mathbf{u}} \varphi^{\mathbf{v}} \mathbf{1}_{\{\mathbf{w}=\mathbf{u}\mathbf{v}\}} \\ &= \sum_{k=0}^{\infty} \sum_{\mathbf{w} \in V_k} \ell^{\mathbf{w}} \left( \widehat{\mathbb{W}}_t \otimes \varphi \right)^{\mathbf{w}} \\ &= \langle \ell, \widehat{\mathbb{W}}_t \otimes \varphi \rangle, \end{aligned}$$

thereby completing the proof.  $\square$

**Theorem 5.2.** *Consider the representation formulas for  $Y$  and  $Z$  given by (4.4) and (4.7). Let  $m \in \mathbb{N}$ , then*

$$\mathbb{E}[Y_T^m | \mathcal{F}_t] = \langle (\ell^{\text{VOL}})^{\sqcup m} |_{\mathcal{E}_{T-t}}, \widehat{\mathbb{W}}_t \rangle, \quad \mathbb{E}[Z_T^m | \mathcal{F}_t] = \langle (\ell^{\text{DE}})^{\sqcup m} |_{\mathcal{E}_{T-t}}, \widehat{\mathbb{W}}_t \rangle, \quad t \leq T.$$

*Proof.* By Theorems 4.2 and 4.4, we have  $\ell^{\text{VOL}}, \ell^{\text{DE}} \in \mathcal{A}_{\text{exp}}$ . Proposition 3.12 (iii) thus makes it clear that  $(\ell^{\text{VOL}})^{\sqcup m}, (\ell^{\text{DE}})^{\sqcup m} \in \mathcal{A}_{\text{exp}}$ . We then apply the shuffle property of Proposition 3.1 to write

$$Y_T^m = \langle (\ell^{\text{VOL}})^{\sqcup m}, \widehat{\mathbb{W}}_T \rangle, \quad Z_T^m = \langle (\ell^{\text{DE}})^{\sqcup m}, \widehat{\mathbb{W}}_T \rangle.$$

Applying the dominated convergence theorem, we get

$$\mathbb{E} \left[ \left\langle \ell^{\perp\perp m}, \widehat{\mathbb{W}}_T \right\rangle \mid \mathcal{F}_t \right] = \left\langle \ell^{\perp\perp m}, \mathbb{E} \left[ \widehat{\mathbb{W}}_T \mid \mathcal{F}_t \right] \right\rangle,$$

for  $\ell \in \{\ell^{\text{VOL}}, \ell^{\text{DE}}\}$ . Using Chen's identity [12], together with Fawcett's formula [21], we have

$$\mathbb{E} \left[ \widehat{\mathbb{W}}_T \mid \mathcal{F}_t \right] = \widehat{\mathbb{W}}_t \otimes \mathbb{E} \left[ \widehat{\mathbb{W}}_{T-t} \mid \mathcal{F}_t \right] = \widehat{\mathbb{W}}_t \otimes \mathcal{E}_{T-t}.$$

Finally, we apply Lemma 5.1 which completes the proof.  $\square$

Figure 4 displays the unconditional and conditional third and fourth moments of the linear Volterra process as in Subsection 4.4.3 computed using Theorem 5.2, illustrating the convergence as the truncation level  $M$  increases for relatively short maturities. The reference value and the confidence intervals in black are computed using Monte Carlo simulation with 50,000 simulations and 500 time steps. We recall that, as was the case for Figure 3, the moments of the true process  $Y$  are estimated with the true kernel  $K^{\alpha, \epsilon}$ , and the signature representations are computed using the multifactor approximation. This illustrates once again the usefulness of Corollary 4.3.

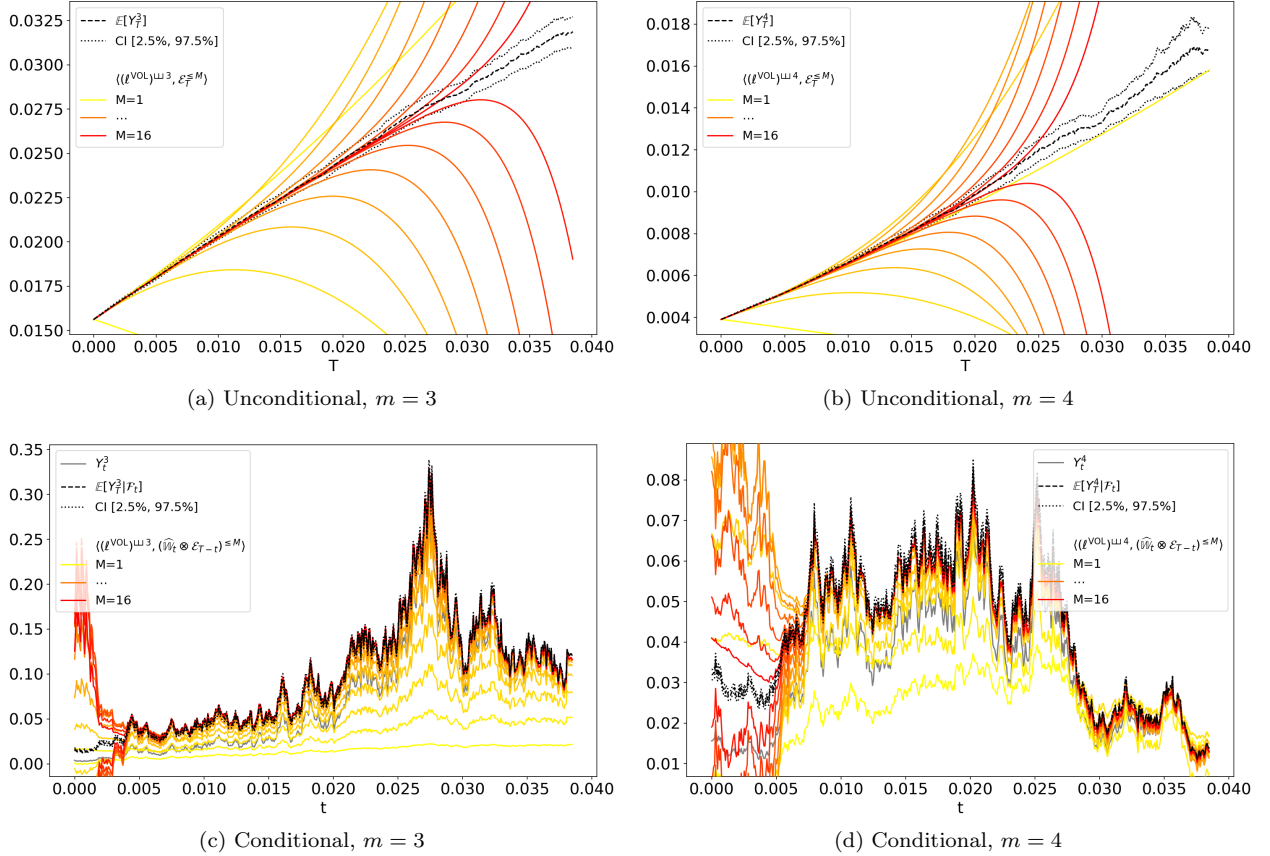


Figure 4:  $m^{\text{th}}$  unconditional, (a) and (b), and conditional, (c) and (d), moments of the shifted fractional Volterra process (black) against their truncated time-independent linear multifactor approximation representation (4.4) to the  $m^{\text{th}}$  (shuffle) power, i.e.  $\left\langle (\ell_n^{\text{VOL}})^{\perp\perp m}, \widehat{\mathbb{W}}_t^{\leq M} \right\rangle$ , for several truncation orders  $M$  and  $y = 0.25, a_1 = 0.25, b_1 = -1, a_2 = -0.1, b_2 = 1, \alpha = 0.6, \epsilon = 1/52, T = 1/26, \tau = 1/106, n = 10$ .

Finally, Figure 5 displays the unconditional and conditional third and fourth moments of the linear Delayed equation process as in Section 4.2 computed using Theorem 5.2, illustrating the convergence as the truncation level  $M$  increases for longer maturities with this choice of parameters. The reference value and the confidence intervals in black are also computed using Monte Carlo simulation with 50,000 simulations and 500 time steps.

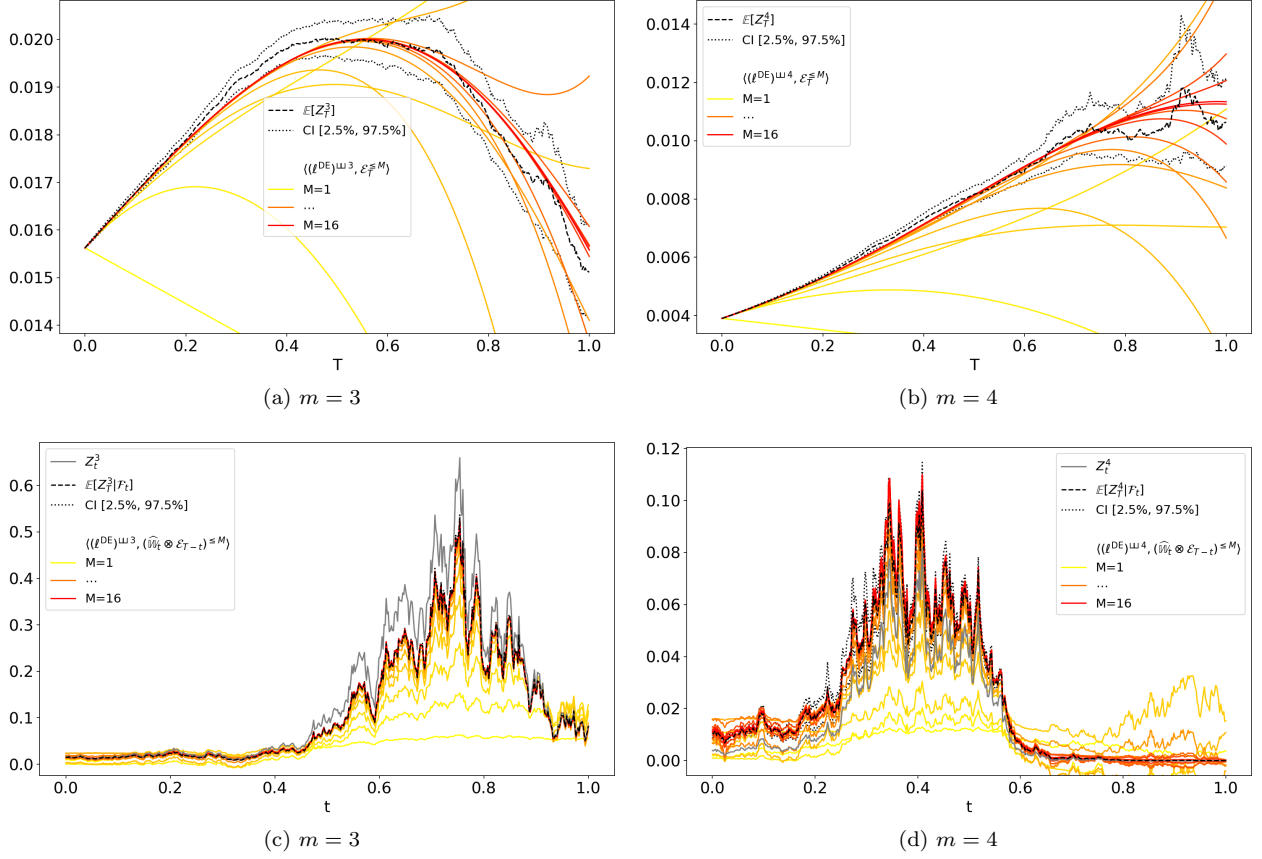


Figure 5:  $m^{\text{th}}$  unconditional, (a) and (b), and conditional (c) and (d) moments of the Delayed equation process (4.6) to the  $m^{\text{th}}$  (shuffle) power, i.e.  $\langle (\ell^{\text{DE}})^{\llcorner m}, \widehat{\mathbb{W}}_t^{\leq M} \rangle$ , for several truncation orders  $M$  and  $z = 0.25, a_1 = 0.125, b_1 = -0.5, c_1 = -0.5, \alpha_1 = 1, a_2 = 0, b_2 = 0.5, c_2 = 0.5, \alpha_2 = -1$ .

## 6 Proof of Theorem 4.2

In a first easy step, we will show that  $\mathbf{p}^{\text{VOL}}$  and  $\mathbf{q}^{\text{VOL}}$  both belong to  $\mathcal{A}_{\text{exp}}$ . Recall the definition of the kernels  $K_1$  and  $K_2$  (4.2) and the assumption (4.3) on them. We can now readily exponentially dominate them:

$$\begin{aligned}
\mathbf{p}^{\text{VOL}} &= a_1 \mathbf{1} \int_{[0, \infty)} e^{\llcorner x} \mu_1(dx) + a_2 \left( \mathbf{2} - \frac{1}{2} K_2(0) b_2 \mathbf{1} \right) \int_{[0, \infty)} e^{\llcorner x} \mu_2(dx) + y \mathfrak{O} \\
&\preceq \left( |a_1| + |a_2| + \frac{1}{2} |a_2 b_2 K_2(0)| \right) (\mathbf{1} + \mathbf{2}) \sum_{n=0}^{\infty} \mathbf{1}^{\otimes n} \int_{[0, \infty)} x^n (\mu_1(dx) + \mu_2(dx)) \\
&\preceq \left( |a_1| + |a_2| + \frac{1}{2} |a_2 b_2 K_2(0)| \right) (\mathbf{1} + \mathbf{2}) e^{\llcorner M} + |y| \mathfrak{O}, \\
\mathbf{q}^{\text{VOL}} &= b_1 \mathbf{1} \int_{[0, \infty)} e^{\llcorner x} \mu_1(dx) + b_2 \left( \mathbf{2} - \frac{1}{2} K_2(0) b_2 \mathbf{1} \right) \int_{[0, \infty)} e^{\llcorner x} \mu_2(dx) \\
&\preceq \left( |b_1| + |b_2| + \frac{1}{2} |b_2^2 K_2(0)| \right) (\mathbf{1} + \mathbf{2}) e^{\llcorner M}.
\end{aligned}$$

By Corollary 3.13,  $\ell^{\text{VOL}}$  belongs to  $\mathcal{A}_{\text{exp}}$  and in particular to  $\mathcal{I}(\widehat{W})$ . The process

$$\bar{Y}_t := \langle \ell^{\text{VOL}}, \widehat{\mathbb{W}}_t \rangle$$

is then well-defined and admits the Itô decomposition

$$\bar{Y}_t = (\ell^{\text{VOL}})^{\circ} + \int_0^t \left\langle \ell^{\text{VOL}}|_{\mathbf{1}} + \frac{1}{2} \ell^{\text{VOL}}|_{\mathbf{22}}, \widehat{\mathbb{W}}_s \right\rangle ds + \int_0^t \left\langle \ell^{\text{VOL}}|_{\mathbf{2}}, \widehat{\mathbb{W}}_s \right\rangle dW_s.$$

It then remains to prove that  $\bar{Y}$  solves the equation

$$\begin{aligned} d\bar{Y}_t &= K_1(0) (a_1 + b_1 \bar{Y}_t) dt + \int_0^t K_1'(t-s) (a_1 + b_1 \bar{Y}_s) ds dt \\ &\quad + K_2(0) (a_2 + b_2 \bar{Y}_t) dW_t + \int_0^t K_2'(t-s) (a_2 + b_2 \bar{Y}_s) dW_s dt, \\ \bar{Y}_0 &= y, \end{aligned}$$

which is the differential version of (4.1) obtained from Itô's lemma, since the kernels  $K_1, K_2$  are smooth. Then, we immediately deduce that  $Y_t = \bar{Y}_t = \left\langle \ell^{\text{VOL}}, \widehat{\mathbb{W}}_t \right\rangle$  by uniqueness of the strong solution, which will complete the proof of Theorem 4.2.

**Lemma 6.1.** *We have that*

$$\begin{aligned} \left\langle \ell^{\text{VOL}}|_{\mathbf{1}} + \frac{1}{2} \ell^{\text{VOL}}|_{\mathbf{22}}, \widehat{\mathbb{W}}_t \right\rangle &= K_1(0) (a_1 + b_1 \bar{Y}_t) \\ &\quad + \int_0^t K_1'(t-s) (a_1 + b_1 \bar{Y}_s) ds + \int_0^t K_2'(t-s) (a_2 + b_2 \bar{Y}_s) dW_s, \\ \left\langle \ell^{\text{VOL}}|_{\mathbf{2}}, \widehat{\mathbb{W}}_t \right\rangle &= K_2(0) (a_2 + b_2 \bar{Y}_t). \end{aligned}$$

## 6.1 Proof of Lemma 6.1

Now, as we will be using integral of elements of  $T((\mathbb{R}^2))$ , recall the Definition 4.1, we need to make sure its bracket with the signature and concatenation make sense. This is the goal of the next two lemmas:

**Lemma 6.2.** *Let  $\mathbf{p} \in L^1(T((\mathbb{R}^2)), \mu)$ , then  $\ell \mathbf{p}, \mathbf{p} \ell \in L^1(T((\mathbb{R}^2)), \mu)$  for every  $\ell \in T((\mathbb{R}^2))$ ,*

$$\int_{[0, \infty)} \ell \mathbf{p}(x) \mu(dx) = \ell \left( \int_{[0, \infty)} \mathbf{p}(x) \mu(dx) \right)$$

and

$$\int_{[0, \infty)} \mathbf{p}(x) \ell \mu(dx) = \left( \int_{[0, \infty)} \mathbf{p}(x) \mu(dx) \right) \ell.$$

*Proof.* Obvious. □

**Lemma 6.3.** *Let  $\mathbf{p} \in L^1(T((\mathbb{R}^2)), \mu)$  such that  $\int_{[0, \infty)} \|\mathbf{p}(x)\|_t^A \mu(dx) < \infty$  for every  $t \in [0, T]$ , then*

$$\left\langle \int_{[0, \infty)} \mathbf{p}(x) \mu(dx), \widehat{\mathbb{W}}_t \right\rangle = \int_{[0, \infty)} \left\langle \mathbf{p}(x), \widehat{\mathbb{W}}_t \right\rangle \mu(dx).$$

*Proof.* By assumption

$$\int_{[0, \infty)} \|\mathbf{p}(x)\|_t^A \mu(dx) = \int_{[0, \infty)} \sum_{n=0}^{\infty} \left| \sum_{\mathbf{v} \in V_n} \mathbf{p}^{\mathbf{v}}(x) \widehat{\mathbb{W}}_t^{\mathbf{v}} \right| \mu(dx) < \infty.$$

This implies that  $\mathbf{p}(x) \in \mathcal{A}$   $\mu$ -a.e., so by the dominated convergence theorem

$$\begin{aligned} \left\langle \int_{[0, \infty)} \mathbf{p}(x) \mu(dx), \widehat{\mathbb{W}}_t \right\rangle &= \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \left( \int_{[0, \infty)} \mathbf{p}^{\mathbf{v}}(x) \mu(dx) \right) \widehat{\mathbb{W}}_t^{\mathbf{v}} \\ &= \int_{[0, \infty)} \sum_{n=0}^{\infty} \sum_{\mathbf{v} \in V_n} \mathbf{p}^{\mathbf{v}}(x) \widehat{\mathbb{W}}_t^{\mathbf{v}} \mu(dx) \\ &= \int_{[0, \infty)} \left\langle \mathbf{p}(x), \widehat{\mathbb{W}}_t \right\rangle \mu(dx). \end{aligned}$$

□

We now introduce two important propositions, which establish the commutativity of the kernel integral and the bracket.

**Proposition 6.4.** *Let  $\ell \in \mathcal{A}_h$ . Assume that  $\mu$  is a measure on  $[0, \infty)$  satisfying  $\int_{[0, \infty)} x^n \mu(dx) < M^n$  for every  $n \in \mathbb{N}$  and a positive constant  $M$ . We then have for  $\mathbf{i} = \mathbf{1}, \mathbf{2}$*

$$\left\langle \ell \mathbf{i} \left( \int_{[0, \infty)} x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \mu(dx) \right), \widehat{\mathbb{W}}_t \right\rangle = \left\langle \int_{[0, \infty)} \ell \mathbf{i} x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \mu(dx), \widehat{\mathbb{W}}_t \right\rangle \quad (6.1)$$

$$= \int_{[0, \infty)} \left\langle \ell \mathbf{i} x e^{\lfloor \cdot \rfloor - x \mathbf{1}}, \widehat{\mathbb{W}}_t \right\rangle \mu(dx). \quad (6.2)$$

*Proof.* We start by proving the first equality (6.1). Fix  $\mathbf{i} \in \{\mathbf{1}, \mathbf{2}\}$ . By assumption of  $\mu$ , it is easy to show that  $\int_{[0, \infty)} x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \mu(dx) \in \mathcal{A}_h$  by direct computation. In particular  $x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \in L^1(\mathcal{A}, \mu)$ . A direct application of Lemma 6.2 shows that

$$\ell \mathbf{i} \left( \int_{[0, \infty)} x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \mu(dx) \right) = \int_{[0, \infty)} \ell \mathbf{i} x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \mu(dx) \in T((\mathbb{R}^2)).$$

Moreover, Proposition 3.7 ensures that it actually belongs to  $\mathcal{A}_h$  so that the bracket is well-defined, which proves the first equality.

For the second equality (6.2), it suffices to verify that  $\int_{[0, \infty)} \|\ell \mathbf{i} x e^{\lfloor \cdot \rfloor - x \mathbf{1}}\|_t^{\mathcal{A}} \mu(dx) < \infty$  for every  $t \in [0, T]$ :

$$\int_{[0, \infty)} \|\ell \mathbf{i} x e^{\lfloor \cdot \rfloor - x \mathbf{1}}\|_t^{\mathcal{A}} \mu(dx) \leq \int_{[0, \infty)} \|\ell \mathbf{i} x e^{\lfloor \cdot \rfloor - x \mathbf{1}}\|_t^{\mathcal{A}_h} \mu(dx) \leq \int_{[0, \infty)} \|\ell\|_t^{\mathcal{A}_h} \|\mathbf{i}\|_t^{\mathcal{A}_h} \|x e^{\lfloor \cdot \rfloor - x \mathbf{1}}\|_t^{\mathcal{A}_h} \mu(dx),$$

where the second equality comes from Proposition 3.7. It then follows

$$\begin{aligned} \int_{[0, \infty)} \|\ell \mathbf{i} x e^{\lfloor \cdot \rfloor - x \mathbf{1}}\|_t^{\mathcal{A}} \mu(dx) &\leq \|\ell\|_t^{\mathcal{A}_h} \|\mathbf{i}\|_t^{\mathcal{A}_h} \int_{[0, \infty)} \|x e^{\lfloor \cdot \rfloor - x \mathbf{1}}\|_t^{\mathcal{A}_h} \mu(dx) \\ &\leq C \|\ell\|_t^{\mathcal{A}_h} \|\mathbf{i}\|_t^{\mathcal{A}_h} \int_{[0, \infty)} \left( \sum_{n=0}^{\infty} x^{n+1} (n+1)^{-\frac{1}{4}} \frac{(2t)^n}{\sqrt{(2n-1)!}} \right) \mu(dx) \\ &\leq C \|\ell\|_t^{\mathcal{A}_h} \|\mathbf{i}\|_t^{\mathcal{A}_h} \sum_{n=0}^{\infty} M^{n+1} (n+1)^{-\frac{1}{4}} \frac{(2t)^n}{\sqrt{(2n-1)!}} \\ &\leq C \|\ell\|_t^{\mathcal{A}_h} \|\mathbf{i}\|_t^{\mathcal{A}_h} M \left( \sum_{n=1}^{\infty} M^n \frac{(2t)^n}{(n-1)!} + 1 \right) \\ &= C \|\ell\|_t^{\mathcal{A}_h} \|\mathbf{i}\|_t^{\mathcal{A}_h} M (2tM e^{2tM} + 1) \\ &< \infty. \end{aligned}$$

An application of Lemma 6.3 completes the proof.  $\square$

**Proposition 6.5.**  *$\mu$  is a measure on  $[0, \infty)$  satisfying  $\int_{[0, \infty)} x^n \mu(dx) < M^n$  for every  $n \in \mathbb{N}$  and a positive constant  $M$ , then*

$$\int_{[0, \infty)} \left\langle \ell^{\text{VOL}} \mathbf{1} x e^{\lfloor \cdot \rfloor - x \mathbf{1}}, \widehat{\mathbb{W}}_t \right\rangle \mu(dx) = - \int_0^t J'(t-s) \bar{Y}_s ds \quad (6.3)$$

and

$$\int_{[0, \infty)} \left\langle \ell^{\text{VOL}} \mathbf{2} x e^{\lfloor \cdot \rfloor - x \mathbf{1}}, \widehat{\mathbb{W}}_t \right\rangle \mu(dx) = - \int_0^t J'(t-s) \bar{Y}_s dW_s - \frac{1}{2} K_2(0) \int_0^t J'(t-s) (a_2 + b_2 \bar{Y}_s) ds$$

where  $J(t) = \int_{[0, \infty)} e^{-xt} \mu(dx)$ .

*Proof.* By Lemma A.1 and the shuffle property, we derive

$$\langle \ell^{\text{VOL}} \mathbf{1} x e^{\lfloor x \rfloor}, \widehat{\mathbb{W}}_t \rangle = x \langle e^{\lfloor x \rfloor}, \widehat{\mathbb{W}}_t \rangle \int_0^t \langle e^{\lfloor x \rfloor} \lfloor \ell^{\text{VOL}} \rfloor, \widehat{\mathbb{W}}_s \rangle ds = x \int_0^t e^{-x(t-s)} \bar{Y}_s ds, \quad (6.4)$$

Similarly, using Lemma A.1 and the shuffle property,

$$\begin{aligned} \langle \ell^{\text{VOL}} \mathbf{2} x e^{\lfloor x \rfloor}, \widehat{\mathbb{W}}_t \rangle &= x \langle e^{\lfloor x \rfloor}, \widehat{\mathbb{W}}_t \rangle \langle (e^{\lfloor x \rfloor} \lfloor \ell^{\text{VOL}} \rfloor) \mathbf{2}, \widehat{\mathbb{W}}_t \rangle \\ &= x \langle e^{\lfloor x \rfloor}, \widehat{\mathbb{W}}_t \rangle \int_0^t \langle e^{\lfloor x \rfloor} \lfloor \ell^{\text{VOL}} \rfloor, \widehat{\mathbb{W}}_s \rangle dW_s \\ &\quad + \frac{1}{2} x \langle e^{\lfloor x \rfloor}, \widehat{\mathbb{W}}_t \rangle \int_0^t \langle e^{\lfloor x \rfloor} \lfloor \ell^{\text{VOL}} \rfloor \mathbf{2}, \widehat{\mathbb{W}}_s \rangle ds \\ &= x \int_0^t e^{-x(t-s)} \bar{Y}_s dW_s + \frac{1}{2} K_2(0) x \int_0^t e^{-x(t-s)} (a_2 + b_2 \bar{Y}_t) ds, \end{aligned}$$

where the second and third equalities come from Theorem 3.2 and (6.8) respectively.

We will now show that we can apply the deterministic and stochastic Fubini theorem. First, noticing that for all  $t \in [0, T]$ ,

$$\mathbb{E} |\bar{Y}_t| = \mathbb{E} \left| \langle \ell^{\text{VOL}}, \widehat{\mathbb{W}}_t \rangle \right| \leq \| \ell^{\text{VOL}} \|_t^{\mathcal{A}_h},$$

and thus

$$\sup_{t \in [0, T]} \mathbb{E} |\bar{Y}_t| < \infty.$$

We can now apply Tonelli's theorem:

$$\begin{aligned} \mathbb{E} \left[ \int_{[0, \infty)} \int_0^t e^{-x(t-s)} |\bar{Y}_s| ds \mu(dx) \right] &= \int_{[0, \infty)} \int_0^t e^{-x(t-s)} \mathbb{E} |\bar{Y}_s| ds \mu(dx) \\ &\leq \int_{[0, \infty)} \int_0^t \mathbb{E} |\bar{Y}_s| ds \mu(dx) \\ &\leq \mu([0, \infty)) \int_0^t \mathbb{E} |\bar{Y}_s| ds \\ &\leq \mu([0, \infty)) T \sup_{t \in [0, T]} \mathbb{E} |\bar{Y}_t| < \infty. \end{aligned}$$

This leads to

$$\int_{[0, \infty)} \int_0^t e^{-x(t-s)} |\bar{Y}_s| ds \mu(dx) < \infty \text{ a.s..}$$

Now we can apply Fubini theorem on equation (6.4) and derive:

$$\int_{[0, \infty)} \langle \ell^{\text{VOL}} \mathbf{1} x e^{\lfloor x \rfloor}, \widehat{\mathbb{W}}_t \rangle \mu(dx) = \int_0^t \left( \int_{[0, \infty)} x e^{-x(t-s)} \mu(dx) \right) \bar{Y}_s ds = - \int_0^t J'(t-s) \bar{Y}_s ds,$$

thus proving (6.3). On the other hand, by Proposition 3.6 (ii),

$$\sup_{t \in [0, T]} \mathbb{E} [\bar{Y}_t^2] \leq \left( \| \ell^{\text{VOL}} \|_T^{\mathcal{A}_h} \right)^2 < \infty$$

and

$$\begin{aligned} \mathbb{E} \left[ \int_{[0, \infty)} \sqrt{\int_0^t e^{-2x(t-s)} |\bar{Y}_s|^2 ds} \mu(dx) \right] &\leq \mu([0, \infty)) \mathbb{E} \left[ \sqrt{\int_0^t |\bar{Y}_s|^2 ds} \right] \\ &\leq \mu([0, \infty)) \sqrt{\mathbb{E} \left[ \int_0^t |\bar{Y}_s|^2 ds \right]} \\ &\leq \mu([0, \infty)) \sqrt{T \sup_{t \in [0, T]} \mathbb{E} |\bar{Y}_t|^2} < \infty, \end{aligned}$$

which leads to

$$\int_{[0,\infty)} \sqrt{\int_0^t e^{-2x(t-s)} |\bar{Y}_s|^2 ds} \mu(dx) < \infty \text{ a.s.} \quad (6.5)$$

Note that (6.5) is actually the condition for stochastic Fubini theorem to hold, see [45, Theorem 2.2]. Applying it, we derive

$$\int_{[0,\infty)} \langle \ell^{\text{VOL}} \mathbf{2} x e^{\llbracket -x \mathbf{1} \rceil}, \widehat{\mathbb{W}}_t \rangle \mu(dx) = - \int_0^t J'(t-s) \bar{Y}_s dW_s - \frac{1}{2} K_2(0) \int_0^t J'(t-s) (a_2 + b_2 \bar{Y}_s) ds,$$

thus completing the proof.  $\square$

Now, taking  $\mu = \mu_1, \mu_2$ , we combine Proposition 6.4 with Proposition 6.5 and derive, for  $i = 1, 2$ :

$$\left\langle \ell^{\text{VOL}} \mathbf{1} \int_{[0,\infty)} x e^{\llbracket -x \mathbf{1} \rceil} \mu_i(dx), \widehat{\mathbb{W}}_t \right\rangle = - \int_0^t K'_i(t-s) \bar{Y}_s ds, \quad (6.6)$$

$$\left\langle \ell^{\text{VOL}} \mathbf{2} \int_{[0,\infty)} x e^{\llbracket -x \mathbf{1} \rceil} \mu_i(dx), \widehat{\mathbb{W}}_t \right\rangle = - \int_0^t K'_i(t-s) \bar{Y}_s dW_s - \frac{1}{2} K_2(0) \int_0^t K'_i(t-s) (a_2 + b_2 \bar{Y}_s) ds. \quad (6.7)$$

Finally, we can now compute the projections  $\ell^{\text{VOL}}|_{\mathbf{1}}, \ell^{\text{VOL}}|_{\mathbf{2}}, \ell^{\text{VOL}}|_{\mathbf{22}}$ . First recall

$$\begin{aligned} \mathbf{p}^{\text{VOL}} &= a_1 \mathbf{1} \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rceil} \mu_1(dx) + a_2 \left( \mathbf{2} - \frac{1}{2} K_2(0) b_2 \mathbf{1} \right) \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rceil} \mu_2(dx) + y \mathbf{\emptyset}, \\ \mathbf{q}^{\text{VOL}} &= b_1 \mathbf{1} \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rceil} \mu_1(dx) + b_2 \left( \mathbf{2} - \frac{1}{2} K_2(0) b_2 \mathbf{1} \right) \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rceil} \mu_2(dx). \end{aligned}$$

We then have

$$\begin{aligned} \mathbf{p}^{\text{VOL}}|_{\mathbf{1}} &= a_1 \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rceil} \mu_1(dx) - a_2 \int_{[0,\infty)} \left( x \mathbf{2} + \frac{1}{2} K_2(0) b_2 \mathbf{\emptyset} \right) e^{\llbracket -x \mathbf{1} \rceil} \mu_2(dx), \\ \mathbf{q}^{\text{VOL}}|_{\mathbf{1}} &= b_1 \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rceil} \mu_1(dx) - b_2 \int_{[0,\infty)} \left( x \mathbf{2} + \frac{1}{2} K_2(0) b_2 \mathbf{\emptyset} \right) e^{\llbracket -x \mathbf{1} \rceil} \mu_2(dx), \\ \mathbf{p}^{\text{VOL}}|_{\mathbf{2}} &= \int_{[0,\infty)} \mu_2(dx) a_2 \mathbf{\emptyset} = K_2(0) a_2 \mathbf{\emptyset}, \quad \mathbf{p}^{\text{VOL}}|_{\mathbf{22}} = 0, \\ \mathbf{q}^{\text{VOL}}|_{\mathbf{2}} &= \int_{[0,\infty)} \mu_2(dx) b_2 \mathbf{\emptyset} = K_2(0) b_2 \mathbf{\emptyset}, \quad \mathbf{q}^{\text{VOL}}|_{\mathbf{22}} = 0. \end{aligned}$$

Then, using Lemma A.2 and noticing  $(\mathbf{q}^{\text{VOL}})^{\mathbf{\emptyset}} = 0$ , it immediately follows that

$$\ell^{\text{VOL}}|_{\mathbf{2}} = \mathbf{p}^{\text{VOL}}|_{\mathbf{2}} + \ell^{\text{VOL}}(\mathbf{q}^{\text{VOL}}|_{\mathbf{2}}) = K_2(0)(a_2 \mathbf{\emptyset} + b_2 \ell^{\text{VOL}}), \quad (6.8)$$

$$\ell^{\text{VOL}}|_{\mathbf{22}} = K_2(0) b_2 \ell^{\text{VOL}}|_{\mathbf{2}} = K_2^2(0) b_2 (a_2 \mathbf{\emptyset} + b_2 \ell^{\text{VOL}}), \quad (6.9)$$

and that

$$\begin{aligned} \ell^{\text{VOL}}|_{\mathbf{1}} &= \mathbf{p}^{\text{VOL}}|_{\mathbf{1}} + \ell^{\text{VOL}}(\mathbf{q}^{\text{VOL}}|_{\mathbf{1}}) \\ &= (a_1 \mathbf{\emptyset} + b_1 \ell^{\text{VOL}}) \int_{[0,\infty)} e^{\llbracket -x \mathbf{1} \rceil} \mu_1(dx) - (a_2 \mathbf{\emptyset} + b_2 \ell^{\text{VOL}}) \int_{[0,\infty)} \left( x \mathbf{2} + \frac{1}{2} K_2(0) b_2 \mathbf{\emptyset} \right) e^{\llbracket -x \mathbf{1} \rceil} \mu_2(dx) \\ &= K_1(0) (a_1 \mathbf{\emptyset} + b_1 \ell^{\text{VOL}}) - \frac{1}{2} K_2^2(0) b_2 (a_2 \mathbf{\emptyset} + b_2 \ell^{\text{VOL}}) \\ &\quad + (a_1 \mathbf{\emptyset} + b_1 \ell^{\text{VOL}}) \mathbf{1} \int_{[0,\infty)} -x e^{\llbracket -x \mathbf{1} \rceil} \mu_1(dx) \\ &\quad + (a_2 \mathbf{\emptyset} + b_2 \ell^{\text{VOL}}) \left( \mathbf{2} - \frac{1}{2} K_2(0) b_2 \mathbf{1} \right) \int_{[0,\infty)} -x e^{\llbracket -x \mathbf{1} \rceil} \mu_2(dx). \end{aligned} \quad (6.10)$$

*Proof of Lemma 6.1.* Following directly from (6.8)-(6.9), we get

$$\langle \ell^{\text{VOL}}|_{\mathbf{2}}, \widehat{\mathbb{W}}_t \rangle = K_2(0)(a_2 + b_2 \bar{Y}_t) \quad \text{and} \quad \langle \ell^{\text{VOL}}|_{\mathbf{22}}, \widehat{\mathbb{W}}_t \rangle = K_2^2(0)b_2(a_2 + b_2 \bar{Y}_t).$$

It now only remains to compute  $\langle \ell^{\text{VOL}}|_{\mathbf{1}}, \widehat{\mathbb{W}}_t \rangle$ . For this, recall (6.10)

$$\begin{aligned} \ell^{\text{VOL}}|_{\mathbf{1}} &= K_1(0) (a_1 \boldsymbol{\phi} + b_1 \ell^{\text{VOL}}) - \frac{1}{2} K_2^2(0) b_2 (a_2 \boldsymbol{\phi} + b_2 \ell^{\text{VOL}}) \\ &\quad + (a_1 \boldsymbol{\phi} + b_1 \ell^{\text{VOL}}) \mathbf{1} \int_{[0, \infty)} -x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \mu_1(dx) \\ &\quad + (a_2 \boldsymbol{\phi} + b_2 \ell^{\text{VOL}}) \left( \mathbf{2} - \frac{1}{2} K_2(0) b_2 \mathbf{1} \right) \int_{[0, \infty)} -x e^{\lfloor \cdot \rfloor - x \mathbf{1}} \mu_2(dx). \end{aligned}$$

A direct application of equation (6.6) and (6.7) yields:

$$\begin{aligned} \langle \ell^{\text{VOL}}|_{\mathbf{1}}, \widehat{\mathbb{W}}_t \rangle &= K_1(0) (a_1 + b_1 \bar{Y}_t) - \frac{1}{2} K_2^2(0) b_2 (a_2 + b_2 \bar{Y}_t) \\ &\quad + \int_0^t K_1'(t-s) (a_1 + b_1 \bar{Y}_s) ds + \int_0^t K_2'(t-s) (a_2 + b_2 \bar{Y}_s) dW_s, \end{aligned}$$

which concludes the proof of Lemma 6.1 and thus of Theorem 4.2.  $\square$

## 7 Proof of Theorem 4.4

It is straightforward to see that  $\mathbf{p}^{\text{DE}} \in \mathcal{A}_{\text{exp}}$  as

$$\mathbf{p}^{\text{DE}} = (a_1 - \frac{1}{2} b_2 a_2) \mathbf{1} + a_2 \mathbf{2} + z \boldsymbol{\phi}.$$

Using Proposition 3.12 (iv)-(i) we can easily see  $\mathbf{q}^{\text{DE}} \in \mathcal{A}_{\text{exp}}$  as

$$\mathbf{q}^{\text{DE}} = \left( (b_1 - \frac{1}{2} b_2^2) \boldsymbol{\phi} + \sum_{m=1}^{k_1} c_1^m \mathbf{1} e^{\lfloor \cdot \rfloor \alpha_1^m \mathbf{1}} - \frac{1}{2} b_2 \sum_{m=1}^{k_2} c_2^m \mathbf{1} e^{\lfloor \cdot \rfloor \alpha_2^m \mathbf{1}} \right) \mathbf{1} + \left( b_2 \boldsymbol{\phi} + \sum_{m=1}^{k_2} c_2^m \mathbf{1} e^{\lfloor \cdot \rfloor \alpha_2^m \mathbf{1}} \right) \mathbf{2}.$$

Then applying Corollary 3.13 we have  $\ell^{\text{DE}} \in \mathcal{A}_{\text{exp}}$  and in particular  $\ell^{\text{DE}} \in \mathcal{I}(\widehat{W})$ . Now recalling Lemma A.2, i.e. for all  $\mathbf{p}, \mathbf{q} \in T((\mathbb{R}^d))$  such that  $\mathbf{q}^\boldsymbol{\phi} = 0$ ,

$$\left( \mathbf{p}(\boldsymbol{\phi} - \mathbf{q})^{-1} \right) |_{\mathbf{i}} = \mathbf{p}|_{\mathbf{i}} + \mathbf{p}(\boldsymbol{\phi} - \mathbf{q})^{-1}(\mathbf{q}|_{\mathbf{i}}),$$

it directly follows that

$$\begin{aligned} \ell^{\text{DE}}|_{\mathbf{1}} &= \ell^{\text{DE}} \left( (b_1 - \frac{1}{2} b_2^2) \boldsymbol{\phi} + \sum_{m=1}^{k_1} c_1^m \mathbf{1} e^{\lfloor \cdot \rfloor \alpha_1^m \mathbf{1}} - \frac{1}{2} b_2 \sum_{m=1}^{k_2} c_2^m \mathbf{1} e^{\lfloor \cdot \rfloor \alpha_2^m \mathbf{1}} \right) + (a_1 - \frac{1}{2} b_2 a_2) \boldsymbol{\phi}, \\ \ell^{\text{DE}}|_{\mathbf{2}} &= \ell^{\text{DE}} \left( b_2 \boldsymbol{\phi} + \sum_{m=1}^{k_2} c_2^m \mathbf{1} e^{\lfloor \cdot \rfloor \alpha_2^m \mathbf{1}} \right) + a_2 \boldsymbol{\phi}, \\ \ell^{\text{DE}}|_{\mathbf{22}} &= b_2 \ell^{\text{DE}}|_{\mathbf{2}}. \end{aligned}$$

Finally,

$$\ell^{\text{DE}}|_{\mathbf{1}} + \frac{1}{2} \ell^{\text{DE}}|_{\mathbf{22}} = \ell^{\text{DE}} \left( b_1 \boldsymbol{\phi} + \sum_{m=1}^{k_1} c_1^m \mathbf{1} e^{\lfloor \cdot \rfloor \alpha_1^m \mathbf{1}} \right) + a_1 \boldsymbol{\phi}.$$

Now ready to apply Theorem 3.3 on  $\ell^{\text{DE}}$ , we get

$$\begin{aligned}
\langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_t \rangle &= (\ell^{\text{DE}})^\circ + \int_0^t \langle \ell^{\text{DE}} | \mathbf{1} + \frac{1}{2} \ell^{\text{DE}} | \mathbf{22}, \widehat{\mathbb{W}}_s \rangle ds + \int_0^t \langle \ell^{\text{DE}} | \mathbf{2}, \widehat{\mathbb{W}}_s \rangle dW_s \\
&= z + \int_0^t \left\langle \ell^{\text{DE}} \left( b_1 \circ + \sum_{m=1}^{k_1} c_1^m \mathbf{1} e^{\llcorner \alpha_1^m \mathbf{1}} \right) + a_1 \circ, \widehat{\mathbb{W}}_s \right\rangle ds \\
&\quad + \int_0^t \left\langle \ell^{\text{DE}} \left( b_2 \circ + \sum_{m=1}^{k_2} c_2^m \mathbf{1} e^{\llcorner \alpha_2^m \mathbf{1}} \right) + a_2 \circ, \widehat{\mathbb{W}}_s \right\rangle dW_s \\
&= z + \int_0^t \left( a_1 + b_1 \langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_s \rangle + \sum_{m=1}^{k_1} c_1^m \langle \ell^{\text{DE}} \mathbf{1} e^{\llcorner \alpha_1^m \mathbf{1}}, \widehat{\mathbb{W}}_s \rangle \right) ds \\
&\quad + \int_0^t \left( a_2 + b_2 \langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_s \rangle + \sum_{m=1}^{k_2} c_2^m \langle \ell^{\text{DE}} \mathbf{1} e^{\llcorner \alpha_2^m \mathbf{1}}, \widehat{\mathbb{W}}_s \rangle \right) dW_s.
\end{aligned}$$

Finally, using Lemma A.1, i.e.  $\ell \mathbf{i} e^{\llcorner \gamma \mathbf{j}} = e^{\llcorner \gamma \mathbf{j}} \llcorner \llcorner ((e^{\llcorner -\gamma \mathbf{j}} \llcorner \llcorner \ell) \mathbf{i})$  for all  $\gamma \in \mathbb{R}$  and  $\mathbf{i}, \mathbf{j} \in A_d$ , and the fact that  $\ell^{\text{DE}} \mathbf{1} e^{\llcorner \alpha_i^m \mathbf{1}} \in \mathcal{A}$  for all  $m$  and  $i = 1, 2$ , it is then straightforward to see that

$$\langle \ell^{\text{DE}} \mathbf{1} e^{\llcorner \alpha_i^m \mathbf{1}}, \widehat{\mathbb{W}}_t \rangle = \int_0^t e^{\alpha_i^m (t-s)} \langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_s \rangle ds.$$

This concludes the proof of Theorem 4.4 that  $\langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_t \rangle$  is solution to (4.6) and hence  $Z_t = \langle \ell^{\text{DE}}, \widehat{\mathbb{W}}_t \rangle$  by strong uniqueness.

## 8 Proof of Theorem 4.5

We first illustrate that  $\langle \ell_t^{\text{GV}}, \widehat{\mathbb{W}}_t \rangle$  with  $\ell_t^{\text{GV}}$  defined in (4.10) is a well-defined stochastic process.

$$\begin{aligned}
\mathbb{E} \left[ \int_0^T \sum_{n=0}^{\infty} |K^{(n)}(t) \langle \mathbf{1}^{\otimes n} \mathbf{2}, \widehat{\mathbb{W}}_t \rangle| dt \right] &= \int_0^T \sum_{n=0}^{\infty} |K^{(n)}(t) \mathbb{E} \left[ \langle \mathbf{1}^{\otimes n} \mathbf{2}, \widehat{\mathbb{W}}_t \rangle \right]| dt \\
&= \int_0^T \sum_{n=0}^{\infty} |K^{(n)}(t) \mathbb{E} \left[ \int_0^t \frac{s^n}{n!} dW_s \right]| dt \\
&\leq \int_0^T \sum_{n=0}^{\infty} |K^{(n)}(t)| \frac{t^{n+\frac{1}{2}}}{n! \sqrt{2n+1}} dt
\end{aligned}$$

The last inequality is derived from Jensen inequality. Then by assumption we know that

$$\mathbb{E} \left[ \int_0^T \sum_{n=0}^{\infty} |K^{(n)}(t) \langle \mathbf{1}^{\otimes n} \mathbf{2}, \widehat{\mathbb{W}}_t \rangle| dt \right] < \infty,$$

thus

$$\sum_{n=0}^{\infty} |K^{(n)}(t) \langle \mathbf{1}^{\otimes n} \mathbf{2}, \widehat{\mathbb{W}}_t \rangle| < \infty, \mathbb{P}(d\omega) \otimes dt - a.e.$$

Consequently,  $\sum_{n=0}^{\infty} K^{(n)}(t) \langle \mathbf{1}^{\otimes n} \mathbf{2}, \widehat{\mathbb{W}}_t \rangle$  converges  $\mathbb{P}(d\omega) \otimes dt - a.e.$ , ensuring  $\langle \ell_t^{\text{GV}}, \widehat{\mathbb{W}}_t \rangle$  to be a  $\mathbb{P}(d\omega) \otimes dt - a.e.$ -well-defined map. Notice that for every  $N \in \mathbb{N}$ ,  $\sum_{n=0}^N K^{(n)}(t) \langle \mathbf{1}^{\otimes n} \mathbf{2}, \widehat{\mathbb{W}}_t \rangle$  is progressively measurable, so  $\langle \ell_t^{\text{GV}}, \widehat{\mathbb{W}}_t \rangle$  is a also progressively measurable. It remains to prove that for every fixed  $t$ ,

$$\langle \ell_t^{\text{GV}}, \widehat{\mathbb{W}}_t \rangle = \int_0^t K(t-s) dW_s. \tag{8.1}$$

For this we write

$$\langle \ell_t^{\text{GV}}, \widehat{\mathbb{W}}_t \rangle = \sum_{n=0}^{\infty} K^{(n)}(t) (-1)^n \langle \mathbf{1}^{\otimes n} \mathbf{2}, \widehat{\mathbb{W}}_t \rangle = \sum_{n=0}^{\infty} K^{(n)}(t) \int_0^t \frac{(-s)^n}{n!} dW_s.$$

Since  $\int_0^t (\sum_{n=0}^{\infty} |K^{(n)}(t)| \frac{s^n}{n!})^2 ds < \infty$  by assumption (4.9), the dominated convergence theorem of stochastic integrals yields

$$\sum_{n=0}^N K^{(n)}(t) \int_0^t \frac{(-s)^n}{n!} dW_s \longrightarrow \int_0^t \sum_{n=0}^{\infty} K^{(n)}(t) \frac{(-s)^n}{n!} dW_s = \int_0^t K(t-s) dW_s,$$

in probability, as  $N \rightarrow \infty$ . While the previous analysis ensures that

$$\sum_{n=0}^N K^{(n)}(t) \int_0^t \frac{(-s)^n}{n!} dW_s \xrightarrow{\text{a.s.}} \sum_{n=0}^{\infty} K^{(n)}(t) \int_0^t \frac{(-s)^n}{n!} dW_s = \langle \ell_t^{\text{GV}}, \widehat{\mathbb{W}}_t \rangle,$$

which proves (8.1) and completes the proof.

## A Properties of the resolvent

We start by deriving a transformation formula, which is simple but crucial to simplify the algebraic expressions.

**Lemma A.1.** *Let  $\ell \in T((\mathbb{R}^d))$ ,  $\mathbf{i} \in A_d$  and  $\mathbf{b} := \sum_{\mathbf{j} \in A_d} \mathbf{b}^{\mathbf{j}} \mathbf{j}$  for  $\mathbf{b}^{\mathbf{j}} \in \mathbb{R}$ , then*

$$\ell \mathbf{i} e^{\sqcup \mathbf{b}} = e^{\sqcup \mathbf{b}} \sqcup ((e^{\sqcup -\mathbf{b}} \sqcup \ell) \mathbf{i}).$$

*Proof.* Let  $\gamma := \ell \mathbf{i} e^{\sqcup \mathbf{b}} - e^{\sqcup \mathbf{b}} \sqcup ((e^{\sqcup -\mathbf{b}} \sqcup \ell) \mathbf{i})$ , then

$$\begin{aligned} \gamma &= \ell \mathbf{i} (\emptyset + e^{\sqcup \mathbf{b}} \mathbf{b}) - (\emptyset + e^{\sqcup \mathbf{b}} \mathbf{b}) \sqcup ((e^{\sqcup -\mathbf{b}} \sqcup \ell) \mathbf{i}) \\ &= \ell \mathbf{i} + \ell \mathbf{i} e^{\sqcup \mathbf{b}} \mathbf{b} - [\emptyset \sqcup (e^{\sqcup -\mathbf{b}} \sqcup \ell)] \mathbf{i} - [(e^{\sqcup \mathbf{b}} \mathbf{b}) \sqcup (e^{\sqcup -\mathbf{b}} \sqcup \ell)] \mathbf{i} - [e^{\sqcup \mathbf{b}} \sqcup (e^{\sqcup -\mathbf{b}} \sqcup \ell)] \mathbf{i} \mathbf{b} \\ &= [\ell - (e^{\sqcup \mathbf{b}} \sqcup e^{\sqcup -\mathbf{b}} \sqcup \ell)] \mathbf{i} + [\ell \mathbf{i} e^{\sqcup \mathbf{b}} - e^{\sqcup \mathbf{b}} \sqcup ((e^{\sqcup -\mathbf{b}} \sqcup \ell) \mathbf{i})] \mathbf{b} \\ &= 0 + \gamma \mathbf{b}, \end{aligned}$$

implying that  $\gamma = 0$ . □

We then provide the useful decomposition of the unique solution of the linear equation  $\ell = \mathbf{p} + \ell \mathbf{q}$ .

**Lemma A.2.** *Let  $\mathbf{p}, \mathbf{q} \in T((\mathbb{R}^d))$  such that  $\mathbf{q}^{\emptyset} \neq 1$  and  $\mathbf{i} \in A_d$ . Then*

$$\left( \mathbf{p} (\emptyset - \mathbf{q})^{-1} \right) \mathbf{i} = \frac{1}{1 - \mathbf{q}^{\emptyset}} \left[ \mathbf{p} \mathbf{i} + \mathbf{p} (\emptyset - \mathbf{q})^{-1} (\mathbf{q} \mathbf{i}) \right].$$

*Proof.* First recall the decomposition in Remark 2.1, i.e.

$$\mathbf{q} = \mathbf{q}^{\emptyset} \emptyset + \sum_{\mathbf{i} \in A_d} \mathbf{q} \mathbf{i},$$

and Proposition 2.2, i.e.

$$(\emptyset - \mathbf{q})^{-1} = \emptyset + (\emptyset - \mathbf{q})^{-1} \mathbf{q}.$$

One can then easily get that

$$(\emptyset - \mathbf{q})^{-1} = \emptyset + (\emptyset - \mathbf{q})^{-1} \left( \mathbf{q}^{\emptyset} \emptyset + \sum_{\mathbf{i} \in A_d} \mathbf{q} \mathbf{i} \right) = \frac{1}{1 - \mathbf{q}^{\emptyset}} \left( \emptyset + (\emptyset - \mathbf{q})^{-1} \sum_{\mathbf{i} \in A_d} \mathbf{q} \mathbf{i} \right).$$

It then follows that

$$\mathbf{p} (\emptyset - \mathbf{q})^{-1} = \frac{1}{1 - \mathbf{q}^{\emptyset}} \left( \mathbf{p} + \mathbf{p} (\emptyset - \mathbf{q})^{-1} \sum_{\mathbf{i} \in A_d} \mathbf{q} \mathbf{i} \right).$$

Taking the projection ends the proof. □

## B Proof of Proposition 3.5

In this section, we are specifically interested in the case  $X_t = \widehat{W}_t = (t, W_t^2, \dots, W_t^d)$  where  $W = (W^2, \dots, W^d)$  is a  $(d-1)$ -dimensional Brownian motion. We will denote  $n(\mathbf{v})$  the size of  $\mathbf{v}$ , i.e. its number of letters, and  $x(\mathbf{v})$  the number of **1**s specifically. If it does not cause ambiguity, we will write  $n$  for  $n(\mathbf{v})$  and  $x$  for  $x(\mathbf{v})$ .

Let  $g^p : [0, T] \times V \rightarrow \mathbb{R}$  be defined for all even  $p \in \mathbb{N}^*$  by

$$g_t^p(\mathbf{v}) = \left[ \prod_{j=2}^p \binom{jn}{n} \right] \frac{t^{\frac{p}{2}(n+x)}}{\left(\frac{p}{2}(n+x)\right)!} 2^{\frac{p}{2}(x-n)}$$

and for all odd  $p \in \mathbb{N}$  by

$$g_t^p(\mathbf{v}) = \left( g_t^{p+1}(\mathbf{v}) \right)^{\frac{p}{p+1}},$$

for all  $t \in [0, T]$  and all  $\mathbf{v} \in V$ . Notice that  $g_t^p(\mathbf{v})$  only depends on  $\mathbf{v}$  through its size and the number of its **1**s, i.e.  $n(\mathbf{v})$  and  $x(\mathbf{v})$ .

**Proposition B.1.** *Fix  $t \in [0, T]$ . We have*

$$\mathbb{E} \left[ \left| \langle \mathbf{v}, \widehat{W}_t \rangle \right|^p \right] \leq g_t^p(\mathbf{v}), \quad \mathbf{v} \in V, \quad p \in \mathbb{N}.$$

*Proof.* We first use Fawcett's formula [37, Proposition 4.10] to write

$$\mathbb{E} \left[ \widehat{W}_t \right] = \sum_{m=0}^{\infty} \frac{t^m}{m!} \left( \mathbf{1} + \frac{1}{2} \sum_{\mathbf{j} \in \{2, \dots, d\}} \mathbf{j} \mathbf{j} \right)^{\otimes m}.$$

We then define for all  $n \in \mathbb{N}$

$$\widehat{V}_n := \{ \mathbf{v} \in V_n : \exists j \in \mathbb{N} \text{ such that } \mathbf{v} = \mathbf{u}_1 \cdots \mathbf{u}_j \text{ for some } \mathbf{u}_k \in \{ \mathbf{1}, \mathbf{22}, \dots, \mathbf{dd} \} \text{ with } k = 1, 2, \dots, j \}.$$

This set corresponds to the coordinate of every signature element with non-zero expectation, i.e. for every  $n \in \mathbb{N}$  and  $\mathbf{v} \in V_n$ ,

$$\mathbb{E} \left[ \langle \mathbf{v}, \widehat{W}_t \rangle \right] = \begin{cases} \frac{t^{\frac{1}{2}(n+x)}}{\left(\frac{1}{2}(n+x)\right)!} \cdot 2^{\frac{1}{2}(x-n)} & \text{if } \mathbf{v} \in \widehat{V}_n, \\ 0 & \text{else.} \end{cases}$$

We are now ready to compute even moments of every signature element. First, remark that for all  $\mathbf{u} \in V_k$  and  $\mathbf{v} \in V_m$ ,  $\mathbf{u} \sqcup \mathbf{v}$  is a sum of  $\binom{k+m}{k}$  elements  $\mathbf{w} \in V_{k+m}$  of length  $n(\mathbf{w}) = k+m$  each, having exactly  $x(\mathbf{w}) = x(\mathbf{u}) + x(\mathbf{v})$  number of **1**s, so that

$$\mathbb{E} \left[ \langle \mathbf{u} \sqcup \mathbf{v}, \widehat{W}_t \rangle \right] \leq \binom{k+m}{k} \max_{\substack{\mathbf{w} \in V_{k+m} \\ \text{s.t. } x(\mathbf{w}) = x(\mathbf{u}) + x(\mathbf{v})}} \mathbb{E} \langle \mathbf{w}, \widehat{W}_t \rangle.$$

Consequently, for every  $p \in \mathbb{N}$ , we can iteratively apply the shuffle product  $p$ -times, and obtain the claimed upper bound for the even powers of the signature:

$$\mathbb{E} \left[ \langle \mathbf{v}, \widehat{W}_t \rangle^{2p} \right] = \mathbb{E} \left[ \langle \mathbf{v} \sqcup \mathbf{v}, \widehat{W}_t \rangle^{2p} \right] \leq \left[ \prod_{j=2}^{2p} \binom{jn}{n} \right] \max_{\substack{\mathbf{w} \in V_{2pn} \\ \text{s.t. } x(\mathbf{w}) = 2px(\mathbf{v})}} \mathbb{E} \langle \mathbf{w}, \widehat{W}_t \rangle = \left[ \prod_{j=2}^{2p} \binom{jn}{n} \right] \frac{t^{p(n+x)}}{(p(n+x))!} 2^{p(x-n)}.$$

We can finally treat the odd moments by applying Jensen's inequality on the convex function  $y \rightarrow y^{\frac{2p}{2p-1}}$  on  $\mathbb{R}_+$ :

$$\mathbb{E} \left[ \left| \langle \mathbf{v}, \widehat{W}_t \rangle \right|^{2p-1} \right]^{\frac{1}{2p-1}} \leq \mathbb{E} \left[ \langle \mathbf{v}, \widehat{W}_t \rangle^{2p} \right]^{\frac{1}{2p}} \leq \left( h_t^{2p}(\mathbf{v}) \right)^{\frac{1}{2p}} = \left( h_t^{2p-1}(\mathbf{v}) \right)^{\frac{1}{2p-1}},$$

concluding the proof of the proposition.  $\square$

We are now ready to prove Proposition 3.5:

*Proof of Proposition 3.5.* Proposition 3.5-(i) follows directly from an application of the above Proposition B.1. In order to prove (3.10) we will consider two cases: either when word  $\mathbf{v}$  ends with a  $\mathbf{1}$  or it ends with a  $\mathbf{j}$  in  $\{\mathbf{2}, \dots, \mathbf{d}\}$ .

(1) **Case  $\mathbf{v} = \mathbf{u}\mathbf{1}$  for some word  $\mathbf{u} \in V$ .** We apply Cauchy-Schwarz inequality to get

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \in [0, t]} \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle^2 \right] &= \mathbb{E} \left[ \sup_{s \in [0, t]} \left( \int_0^s \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle du \right)^2 \right] \\ &\leq \mathbb{E} \left[ \left( \int_0^t |\langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle| du \right)^2 \right] \\ &\leq \mathbb{E} \left[ t \int_0^t \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle^2 du \right] \\ &= t \int_0^t \mathbb{E} \left[ \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle^2 \right] du. \end{aligned}$$

Thus, using (3.9) with  $n = n(\mathbf{v}) = n(\mathbf{u}) + 1$  and  $x = x(\mathbf{v}) = x(\mathbf{u}) + 1$ , we have

$$\begin{aligned} \mathbb{E} \left[ \sup_{s \in [0, t]} \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle^2 \right] &\leq \binom{2(n-1)}{n-1} \frac{t}{2^{n-x}(n+x-2)!} \int_0^t u^{n+x-2} du \\ &= \binom{2(n-1)}{n-1} \frac{t^{n+x}}{2^{n-x}(n+x-1)!} \\ &\leq \frac{2^{n+x}}{\sqrt{n}} \frac{t^{n+x}}{(n+x-1)!}, \end{aligned} \tag{B.1}$$

where the last inequality comes from an application of Stirling's approximation that reads  $\binom{2m}{m} \leq \frac{2^{2m}}{\sqrt{m+1}}$  for all  $m \in \mathbb{N}$ .

(2) **Case  $\mathbf{v} = \mathbf{u}\mathbf{j}$  for some word  $\mathbf{u} \in V$  and some letter  $\mathbf{j} \in \{\mathbf{2}, \dots, \mathbf{d}\}$ .** First we apply Theorem 3.3 to get

$$\begin{aligned} \langle \mathbf{v}, \widehat{\mathbb{W}}_s \rangle^2 &= \left( \int_0^s \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle dW_u^{\mathbf{j}} + \frac{1}{2} \int_0^s \langle \mathbf{u}\mathbf{j}, \widehat{\mathbb{W}}_u \rangle du \right)^2 \\ &\leq 2 \left( \int_0^s \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle dW_u^{\mathbf{j}} \right)^2 + \frac{1}{2} \left( \int_0^s \langle \mathbf{u}\mathbf{j}, \widehat{\mathbb{W}}_u \rangle du \right)^2. \end{aligned} \tag{B.2}$$

Now, since  $\int_0^s \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle dW_u^{\mathbf{j}}$  is a true martingale, we apply Doob's inequality followed by the previous result (B.1) with  $n = n(\mathbf{v}) = n(\mathbf{u}) + 1$  and  $x = x(\mathbf{v}) = x(\mathbf{u})$  and get

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left( \int_0^s \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle dW_u^{\mathbf{j}} \right)^2 \right] \leq 4 \mathbb{E} \left[ \int_0^t \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle^2 du \right] = 4 \int_0^t \mathbb{E} \left[ \langle \mathbf{u}, \widehat{\mathbb{W}}_u \rangle^2 \right] du \leq 4 \frac{2^{n+x}}{\sqrt{n}} \frac{t^{n+x}}{(n+x)!}.$$

Then, again applying (B.1) to the right-most element in (B.2), we get

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left( \int_0^s \langle \mathbf{u}\mathbf{j}, \widehat{\mathbb{W}}_u \rangle du \right)^2 \right] = \mathbb{E} \left[ \sup_{s \in [0, t]} \langle \mathbf{u}\mathbf{j}\mathbf{1}, \widehat{\mathbb{W}}_s \rangle^2 \right] \leq \frac{2^{n+x}}{\sqrt{n}} \frac{t^{n+x}}{(n+x-1)!}.$$

Note that if  $\mathbf{u}\mathbf{j} = 0$ , the left-hand side is 0 and the equation still holds.

Combining the above inequalities, we derive

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left\langle \mathbf{v}, \widehat{\mathbb{W}}_s \right\rangle^2 \right] \leq C_1 \frac{2^{n+x-1}}{\sqrt{n}} \frac{t^{n+x}}{(n+x-1)!}, \quad (\text{B.3})$$

for some  $C_1 \in \mathbb{R}$  independent of  $t$  and  $\mathbf{v}$ .

Combining the two equalities (B.1)-(B.3) above, we have that for every  $\mathbf{v} \in V_n$ ,  $n \geq 1$ ,

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left| \left\langle \mathbf{v}, \widehat{\mathbb{W}}_s \right\rangle \right|^2 \right] \leq C_2 \frac{2^{n+x-1}}{\sqrt{n}} \frac{t^{n+x}}{(n+x-1)!}.$$

for some  $C_2 \in \mathbb{R}$  independent of  $t$  and  $\mathbf{v}$ .

Moreover, as  $n+1 \sim n$ , there exists a constant  $C$  such that

$$\mathbb{E} \left[ \sup_{s \in [0, t]} \left| \left\langle \mathbf{v}, \widehat{\mathbb{W}}_s \right\rangle \right|^2 \right] \leq C \frac{2^{n+x-1}}{\sqrt{n+1}} \frac{t^{n+x}}{(n+x-1)!}.$$

This includes the case  $\mathbf{v} = \emptyset \in V_0$  and proves (3.10).

It remains to prove the sub-multiplicative property (3.11). First, in the case  $\mathbf{u} = \emptyset$  or  $\mathbf{w} = \emptyset$ , (3.11) holds trivially. Now we can easily see that

$$\left( \frac{h_t(\mathbf{u})h_t(\mathbf{v})}{h_t(\mathbf{uv})} \right)^2 = C^2 A(\mathbf{u}, \mathbf{v})B(\mathbf{u}, \mathbf{v}),$$

where

$$A(\mathbf{u}, \mathbf{v}) := \frac{(n(\mathbf{u}) + x(\mathbf{u}) + n(\mathbf{v}) + x(\mathbf{v}) - 1)!}{(n(\mathbf{u}) + x(\mathbf{u}) - 1)!(n(\mathbf{v}) + x(\mathbf{v}) - 1)!} \quad \text{and} \quad B(\mathbf{u}, \mathbf{v}) := \sqrt{\frac{n(\mathbf{u}) + n(\mathbf{v}) + 1}{(n(\mathbf{u}) + 1)(n(\mathbf{v}) + 1)}}.$$

We now only need to show that  $C^2 A(\mathbf{u}, \mathbf{v})B(\mathbf{u}, \mathbf{v}) \geq 1$  for  $\mathbf{u}, \mathbf{v} \neq \emptyset$ . Without loss of generality, we assume  $\mathbf{u} \neq \emptyset$ , i.e.  $n(\mathbf{u}) > 0$ , otherwise we could simply swap  $\mathbf{u}$  and  $\mathbf{v}$  in the following equations. Let us then remark that

$$A(\mathbf{u}, \mathbf{v}) = \binom{n(\mathbf{u}) + x(\mathbf{u}) + n(\mathbf{v}) + x(\mathbf{v}) - 1}{n(\mathbf{u}) + x(\mathbf{u})} (n(\mathbf{u}) + x(\mathbf{u}))$$

and that

$$\begin{aligned} \sqrt{n(\mathbf{u}) + 1} &\leq C^2 (n(\mathbf{u}) + x(\mathbf{u})) \\ \sqrt{n(\mathbf{v}) + 1} \sqrt{n(\mathbf{u}) + 1} &\leq C^2 \sqrt{n(\mathbf{v}) + 1} (n(\mathbf{u}) + x(\mathbf{u})) \leq C^2 \sqrt{n(\mathbf{u}) + n(\mathbf{v}) + 1} (n(\mathbf{u}) + x(\mathbf{u})), \end{aligned}$$

which implies

$$B(\mathbf{u}, \mathbf{v}) \geq \frac{1}{C^2 (n(\mathbf{u}) + x(\mathbf{u}))}$$

and thus

$$\left( \frac{h_t(\mathbf{u})h_t(\mathbf{v})}{h_t(\mathbf{uv})} \right)^2 \geq \binom{n(\mathbf{u}) + x(\mathbf{u}) + n(\mathbf{v}) + x(\mathbf{v}) - 1}{n(\mathbf{u}) + x(\mathbf{u})} \geq 1.$$

□

## C Proof of Proposition 3.1

Let  $n \leq k \in \mathbb{N}$ , then for all  $\mathbf{u} \in V_n$ ,  $\mathbf{v} \in V_{k-n}$ , each element of  $\mathbf{u} \sqcup \mathbf{v}$  is of size  $k$ . This leads to the fact that  $\langle \mathbf{u} \sqcup \mathbf{v}, \mathbf{w} \rangle = 0$  whenever  $\mathbf{w} \notin V_k$ , i.e.

$$\langle \mathbf{u} \sqcup \mathbf{v}, \mathbb{X}_t \rangle = \sum_{\mathbf{w} \in V_k} \langle \mathbf{u} \sqcup \mathbf{v}, \mathbf{w} \rangle \mathbb{X}_t^{\mathbf{w}}.$$

Moreover, by bilinearity of the shuffle product and for  $\ell, \varphi \in T((\mathbb{R}^d))$ ,

$$\langle \ell \sqcup \varphi, \mathbf{w} \rangle = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_m} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \langle \mathbf{u} \sqcup \mathbf{v}, \mathbf{w} \rangle = \sum_{n=0}^k \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \langle \mathbf{u} \sqcup \mathbf{v}, \mathbf{w} \rangle.$$

Hence, it is easy to see that for all  $k \in \mathbb{N}$

$$\sum_{\mathbf{w} \in V_k} \langle \ell \sqcup \varphi, \mathbf{w} \rangle \mathbb{X}_t^{\mathbf{w}} = \sum_{n=0}^k \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \langle \mathbf{u} \sqcup \mathbf{v}, \mathbb{X}_t \rangle.$$

It then follows that the semi-norm of  $\ell \sqcup \varphi$  is of the form

$$\begin{aligned} \|\ell \sqcup \varphi\|_t^A &= \sum_{k=0}^{\infty} \left| \sum_{\mathbf{w} \in V_k} \langle \ell \sqcup \varphi, \mathbf{w} \rangle \mathbb{X}_t^{\mathbf{w}} \right| \\ &= \sum_{k=0}^{\infty} \left| \sum_{n=0}^k \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \langle \mathbf{u} \sqcup \mathbf{v}, \mathbb{X}_t \rangle \right| \\ &\leq \sum_{k=0}^{\infty} \sum_{n=0}^k \left| \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \langle \mathbf{u} \sqcup \mathbf{v}, \mathbb{X}_t \rangle \right|. \end{aligned}$$

Finally, using the shuffle product property on the right-hand side, recall Lemma 2.5, the semi-norm can be bounded

$$\|\ell \sqcup \varphi\|_t^A \leq \sum_{k=0}^{\infty} \sum_{n=0}^k \left| \sum_{\mathbf{u} \in V_n} \sum_{\mathbf{v} \in V_{k-n}} \ell^{\mathbf{u}} \varphi^{\mathbf{v}} \mathbb{X}_t^{\mathbf{u}} \mathbb{X}_t^{\mathbf{v}} \right| \leq \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left| \sum_{\mathbf{u} \in V_n} \ell^{\mathbf{u}} \mathbb{X}_t^{\mathbf{u}} \sum_{\mathbf{v} \in V_m} \varphi^{\mathbf{v}} \mathbb{X}_t^{\mathbf{v}} \right| \leq \|\ell\|_t^A \cdot \|\varphi\|_t^A < \infty.$$

It can then be argued very similarly that  $\langle \ell, \mathbb{X}_t \rangle \langle \varphi, \mathbb{X}_t \rangle = \langle \ell \sqcup \varphi, \mathbb{X}_t \rangle$ , ending the proof.

## References

- [1] Eduardo Abi Jaber. Lifting the heston model. *Quantitative finance*, 19(12):1995–2013, 2019.
- [2] Eduardo Abi Jaber and Omar El Euch. Multifactor approximation of rough volatility models. *SIAM journal on financial mathematics*, 10(2):309–349, 2019.
- [3] Eduardo Abi Jaber and Louis-Amand Gérard. Signature volatility models: Pricing and hedging with fourier. *SIAM Journal on Financial Mathematics*, 16(2):606–642, may 2025. ISSN 1945-497X. doi: 10.1137/24m1636952. URL <http://dx.doi.org/10.1137/24M1636952>.
- [4] Eduardo Abi Jaber, Martin Larsson, and Sergio Pulido. Affine Volterra processes. *The Annals of Applied Probability*, 29(5):3155–3200, 2019.
- [5] Eduardo Abi Jaber, Christa Cuchiero, Luca Pelizzari, Sergio Pulido, and Sara Svaluto-Ferro. Polynomial volterra processes. *Electronic Journal of Probability*, 29:1–37, 2024.
- [6] Christian Bayer, Peter K. Friz, Paul Gassiat, Jorg Martin, and Benjamin Stemper. A regularity structure for rough volatility. *Mathematical Finance*, 30(3):782–832, 2020. doi: 10.1111/mafi.12233. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12233>.

- [7] Christian Bayer, Paul P Hager, Sebastian Riedel, and John Schoenmakers. Optimal stopping with signatures. *The Annals of Applied Probability*, 33(1):238–273, 2023.
- [8] Christian Bayer, Luca Pelizzari, and John Schoenmakers. Primal and dual optimal stopping with signatures. *Finance and Stochastics*, 29(4):981–1014, October 2025. ISSN 1432-1122. doi: 10.1007/s00780-025-00570-8. URL <https://doi.org/10.1007/s00780-025-00570-8>.
- [9] Gérard Ben Arous. Flots et séries de taylor stochastiques. *Probability Theory and Related Fields*, 81: 29–77, 1989.
- [10] Yvain Bruned and Foivos Katsetiadis. Ramification of Volterra-type rough paths. *Electronic Journal of Probability*, 28:1 – 25, 2023. doi: 10.1214/22-EJP890. URL <https://doi.org/10.1214/22-EJP890>.
- [11] Alvaro Cartea, Imanol Perez Arribas, and Leandro Sanchez-Betancourt. Optimal execution of foreign securities: A double-execution problem, 2020. URL <https://ssrn.com/abstract=3562251>.
- [12] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Annals of Mathematics*, 65(1):163–178, 1957.
- [13] Etienne Chevalier, Sergio Pulido, and Elizabeth Zúñiga. American options in the volterra heston model. *SIAM Journal on Financial Mathematics*, 13(2):426–458, 2022. doi: 10.1137/21M140674X. URL <https://doi.org/10.1137/21M140674X>.
- [14] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. *arXiv preprint arXiv:1603.03788*, 2016.
- [15] Ilya Chevyrev and Harald Oberhauser. Signature moments to characterize laws of stochastic processes. *Journal of Machine Learning Research*, 23(176):1–42, 2022. URL <http://jmlr.org/papers/v23/20-1466.html>.
- [16] Christa Cuchiero, Guido Gazzani, Janka Möller, and Sara Svaluto-Ferro. Joint calibration to spx and vix options with signature-based models. *Mathematical Finance*, 35(1):161–213, 2025. doi: 10.1111/mafi.12442. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mafi.12442>.
- [17] Christa Cuchiero, Sara Svaluto-Ferro, and Josef Teichmann. Signature sdes from an affine and polynomial perspective, 2025. URL <https://arxiv.org/abs/2302.01362>.
- [18] Aurélien Deya and Samy Tindel. Rough volterra equations 1: The algebraic integration setting. *Stochastics and Dynamics*, 09(03):437–477, 2009. doi: 10.1142/S0219493709002737. URL <https://doi.org/10.1142/S0219493709002737>.
- [19] Aurélien Deya and Samy Tindel. Rough volterra equations 2: Convolutional generalized integrals. *Stochastic Processes and their Applications*, 121(8):1864–1899, 2011. ISSN 0304-4149. doi: 10.1016/j.spa.2011.05.003. URL <https://www.sciencedirect.com/science/article/pii/S0304414911001049>.
- [20] Bruno Dupire and Valentin Tisserand-Daguette. Functional expansions. *arXiv preprint arXiv:2212.13628*, 2022.
- [21] Thomas Fawcett. *Problems in stochastic analysis. Connections between rough paths and non-commutative harmonic analysis*. PhD thesis, University of Oxford, 2003.
- [22] Adeline Fermanian. Embedding and learning with signatures. *Computational Statistics & Data Analysis*, 157:107148, 2021.
- [23] Michel Fliess. On the concept of derivatives and taylor expansions for nonlinear input-output systems. In *The 22nd IEEE Conference on Decision and Control*, pages 643–646. IEEE, 1983.
- [24] Michel Fliess. Vers une notion de dérivation fonctionnelle causale. In *Annales de l’Institut Henri Poincaré C, Analyse non linéaire*, volume 3, pages 67–76. Elsevier, 1986.
- [25] Peter K. Friz and Nicolas B. Victoir. *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010. doi: 10.1017/CBO9780511845079.

- [26] J. G. Gaines. The algebra of iterated stochastic integrals. *Stochastics and Stochastic Reports*, 49(3-4): 169–179, 1994. doi: 10.1080/17442509408833918. URL <https://doi.org/10.1080/17442509408833918>.
- [27] Chad Giusti, Darrick Lee, Vidit Nanda, and Harald Oberhauser. A topological approach to mapping space signatures. *Advances in Applied Mathematics*, 163:102787, 2025. ISSN 0196-8858. doi: 10.1016/j.aam.2024.102787. URL <https://www.sciencedirect.com/science/article/pii/S0196885824001192>.
- [28] Gustaf Gripenberg, Stig-Olof Londen, and Olof Staffans. *Volterra integral and functional equations*. Cambridge University Press, 1990.
- [29] Yushi Hamaguchi. Variation of constants formulae for forward and backward stochastic volterra integral equations. *Journal of Differential Equations*, 343:332–389, 2023.
- [30] Fabian Harang, Samy Tindel, and Xiaohua Wang. Volterra equations driven by rough signals 3: Probabilistic construction of the volterra rough path for fractional brownian motions. *Journal of Theoretical Probability*, 37(1):307–351, March 2024. ISSN 1572-9230. doi: 10.1007/s10959-023-01251-y. URL <https://doi.org/10.1007/s10959-023-01251-y>.
- [31] Fabian A. Harang and Samy Tindel. Volterra equations driven by rough signals. *Stochastic Processes and their Applications*, 142:34–78, 2021. ISSN 0304-4149. doi: 10.1016/j.spa.2021.08.001. URL <https://www.sciencedirect.com/science/article/pii/S0304414921001253>.
- [32] Yaozhong Hu. Multiple integrals and expansion of solutions of differential equations driven by rough paths and by fractional brownian motions. *Stochastics*, 85(5):859–916, 2013. doi: 10.1080/17442508.2012.673615. URL <https://doi.org/10.1080/17442508.2012.673615>.
- [33] Peter E Kloeden, Eckhard Platen, Peter E Kloeden, and Eckhard Platen. *Stochastic differential equations*. Springer, 1992.
- [34] Gottfried Köthe. *Topological Vector Spaces*, pages 123–201. Springer Berlin Heidelberg, Berlin, Heidelberg, 1983. ISBN 978-3-642-64988-2. doi: 10.1007/978-3-642-64988-2\_3. URL [https://doi.org/10.1007/978-3-642-64988-2\\_3](https://doi.org/10.1007/978-3-642-64988-2_3).
- [35] Christian Litterer and Harald Oberhauser. On a chen–fliess approximation for diffusion functionals. *Monatshefte für Mathematik*, 175:577–593, 2014.
- [36] Terry Lyons. Rough paths, signatures and the modelling of functions on streams. *arXiv preprint arXiv:1405.4537*, 2014.
- [37] Terry Lyons and Nicolas Victoir. Cubature on wiener space. *Proceedings: Mathematical, Physical and Engineering Sciences*, 460(2041):169–198, 2004. ISSN 13645021. URL <http://www.jstor.org/stable/4143098>.
- [38] Terry Lyons, Sina Nejad, and Imanol Perez Arribas. Non-parametric pricing and hedging of exotic derivatives. *Applied Mathematical Finance*, 27(6):457–494, 2020. doi: 10.1080/1350486X.2021.1891555. URL <https://doi.org/10.1080/1350486X.2021.1891555>.
- [39] Terry J. Lyons. Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310, 1998. URL <http://eudml.org/doc/39555>.
- [40] A. Neuenkirch, I. Nourdin, A. Röbler, and S. Tindel. Trees and asymptotic expansions for fractional stochastic differential equations. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 45(1): 157 – 174, 2009. doi: 10.1214/07-AIHP159. URL <https://doi.org/10.1214/07-AIHP159>.
- [41] Andreas Neuenkirch, Ivan Nourdin, and Samy Tindel. Delay equations driven by rough paths. *Electronic Journal of Probability* [electronic only], 13:2031–2068, 2008.
- [42] Imanol Perez Arribas, Guy M Goodwin, John R Geddes, Terry Lyons, and Kate EA Saunders. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. *Translational psychiatry*, 8(1):274, 2018.

- [43] Julio C. Ramos-Fernández and Margot Salas-Brown. On multiplication operators acting on köthe sequence spaces. *Afrika Matematika*, 28(3):661–667, 2017. ISSN 2190-7668. doi: 10.1007/s13370-016-0475-3. URL <https://doi.org/10.1007/s13370-016-0475-3>.
- [44] Rimhak Ree. Lie elements and an algebra associated with shuffles. *Annals of Mathematics*, 68(2):210–220, 1958. ISSN 0003486X. URL <http://www.jstor.org/stable/1970243>.
- [45] Mark Veraar. The stochastic fubini theorem revisited. *Stochastics An International Journal of Probability and Stochastic Processes*, 84(4):543–551, 2012.