DISCRETE SIGNATURE VARIETIES

CARLO BELLINGERI, RAUL PENAGUIAO

TU Berlin, Max Planck Institute for the Mathematics in Sciences Leipzig

ABSTRACT. Discrete signatures are invariants extracted from a discretized version of paths that resembles the iterated integral signature of rough paths. In this paper we study the image of these discrete signatures, the discrete signature variety, and begin a classification of the *primary* signatures, elements that play a crutial role in computing the dimension of the associated Lie algebra. We also present some generators of this variety and the results of some *Macaulay2* code to that effect.

1. INTRODUCTION

Signatures of smooth curves $x : [0,1] \to \mathbb{R}$ were introduced in [Che57], leading to revolutionary applications on stochastic analysis with contributions by Lyons, Friz, Hairer and others, see [LCL07, FV10, FH20].

The discrete counterpart of this invariant is the **discrete signature** of a time-series, or iterated sum signature, introduced by [DET20]. When focusing on data that is naturally discrete, several interesting applications of this new invariant arise, of which we remark signal compression [BBSK⁺17]. These applications arise from the fact that discrete signatures are invariant under time-warping. That is, data is allowed to "stutter", leading nonetheless to the same signature.

The discrete signature is defined for a times series of finite vectors in \mathbb{R}^d , and results in a signature, *i.e.* an element in the tensor series space $T((\mathfrak{A}_d))$. This tensor space can be described by an infinite sum indexed on words of monomials. In this way, for a word $p_1 \bullet \cdots \bullet p_l$ of non-constant monomials p_i in $\mathbb{K}[X_1, \ldots, X_d]$ we can define the corresponding coefficient in the signature of a time-series $x = (x_1, \ldots, x_N)$ as follows:

(1)
$$S_{p_1 \bullet \dots \bullet p_l}(x) = \sum_{1 \le i_1 < \dots < i_l < N} \prod_{j=1}^l p_j(x_{i_j+1} - x_{i_j}).$$

We argue that this new tensor space is a very interesting one from the point of view of statistics and numerical analysis. We will truncate this space and consider only the

E-mail address: bellinge@math.tu-berlin.de, raul.penaguiao@mis.mpg.de. *Date*: March 24, 2023.

²⁰¹⁰ Mathematics Subject Classification. 17B45,14Q15,60H99.

Key words and phrases. Lie algebras, signatures, varieties, Lyndon words.

span of the words with **total degree** bounded by $h, T^{\leq h}((\mathfrak{A}_d))$. For instance, at h = 2 we have

$$\mathfrak{S}^{\leq 2} = \varepsilon + \mathfrak{S}^{(1)} + \mathfrak{S}^{(1,1)} + \mathfrak{S}^{(2)}$$

where $S^{(1)}$ are all the signatures of the form S_t , $S^{(1,1)}$ are all the signatures of the form $S_{t\circ u}$, and $S^{(2)}$ are all the signatures of the form S_{tu} , for t and u linear monomials.

On this truncated space, we will study the different images of the signature map, the variety $\mathcal{V}_{d,h,N}$. This variety is an irreducible algebraic variety that arises from a very dificult implicitization problem. Part of this paper is studying the limit of this image when N is very large, which will give us the **universal variety**. This mimics what was done for Chen's signature in [AFS19]. It uses a basis indexed by Lyndon words. However, these words are skewed with a **height** function, generalising the original concept from [CFL58]. These combinatorially skewed Lyndon words arise in several places of the Hopf algebra landscape, for instance in quasi-symmetric functions (see [Haz01]) as well as permutations (see [Var14, BP20]) and marked permutations (see [Pen22]). Our study uses results from Lie algebras put forth in [BFPP22].

The signature of a time-series satisfies the so called **quasi-shuffle relations**, but the question stands: does any truncated tensor series satisfying the quasi-shuffle relations arise from the truncated signature of a time-series? A strategy for establishing precisely that will be outlined, and as a result we conjecture the dimension of the universal variety. Specifically, in Lemma 5.4 we relate this with the following equation, which we will study from an algebraic point of view in Sections 4 and 5.2:

Question 1.1. Fix a time-series length N. For which $w \in K[X_1, \ldots, X_d]$ can we find a time-series x of length N such that

(2)
$$\exp(\mathsf{w}) = \Phi_H^*(\mathcal{S}(x)),$$

where Φ_H is the Hofmann isomorphism, a linear map introduced below. Such timeseries x are called **primary elements**, due to their role in spanning the entire space of discrete signatures, *vide* Lemma 5.3.

In Section 5.2 we explicitly state how these relations look like for the case where h = 3.

This paper also studies the intermediate varieties $\mathcal{V}_{d,h,N}$. An avatar of these varieties is presented, where we answer dimension and degree questions for some small cases of d and h. Noticing that the iterated sums in Eq. (1) are polynomial functions on the time-series, we leverage computational algebra tools to present *Gröbner basis* of the vanishing ideals of the image $\mathcal{V}_{d,h,N}$, drawing inspiration from algebraic statistics (see [DSS08]). These vanishing ideals contain the quasi-shuffle relations, but are usually much larger.

Finally, we discuss the **degree problem** from the point of view of path recovery. This traces back to the study in [PSS19], where paths are recovered only knowing its signature of order three. We establish bounds for the degree of low order signature varieties. This means that for a set of linear constraints, there is a bounded number of time-series that have a specific signature.

We now state what we will present in this paper. We start with some preliminaries, introducing the K-algebras and Lie algebras of interest, as well as the main properties of

the shuffle relations. We will also present algebraic equations that define the signature space for small cases in Section 4. Then, in Section 5, we discuss the application of Chow theorem in the discrete signature varieties. Finally, in Section 5.4, we present an enumerative result for the dimension of the universal variety.

2. Preliminaries

Combinatorics. For an integer $n \ge 1$, write $[n] \coloneqq \{1, \ldots, n\}$. Given a countable alphabet I, a word $w = (i_1, \ldots, i_n)$ is an n-tuple of elements in I, for $n \ge 0$ integer. We may write $w = i_1 \bullet \cdots \bullet i_n$ for simplicity, and |w| = n is called the length of w. We denote by $\mathcal{W}(I)$ the set of words in I, including the empty word, which we denote ε .

Two words w and v may be concatenated, which we represent by $w \bullet v$. Furthermore, we abuse notation and write $w \bullet i$ to convey the concatenation of w with the length one word (i). An alphabet I may be equipped with a degree map deg : $I \to \mathbb{Z}_{>0}$. We call the sum $\sum_{j=1}^{n} \deg(i_j)$ the **height** $||w||_{ht}$ of a word $w = i_1 \bullet \cdots \bullet i_n$. One has that $|w| \leq ||w||_{\text{ht}}$ on any non-empty word w.

A composition α of an integer k is an n-tuple of positive integers $\alpha = (\alpha_1, \ldots, \alpha_n)$ such that $\sum_{i=1}^{n} \alpha_i = k$. We denote the set of all composition of k by C(k), and write $\ell(\alpha) = n$ for the **length** of the composition. We use the following statistics on compositions $\alpha! \coloneqq \prod_{i=1}^{n} \alpha_i!$ and $\Pi \alpha \coloneqq \prod_{i=1}^{n} \alpha_i$.

Assume now that I is an alphabet with an associative and commutative operation, which we denote by juxtaposition. If $\alpha \in C(k)$ and $w = i_1 \bullet \cdots \bullet i_k$ is a word such that |w| = k, then we define the **contracted word** $(w)_{\alpha}$ by multiplying the letters in w according to α , i.e.

$$(w)_{\alpha} = \tau_1 \bullet \cdots \bullet \tau_{\ell(\alpha)}, \quad \text{with} \quad \tau_j \coloneqq i_{s_j+1} \cdots i_{s_j+\alpha_j},$$

where $s_j = \sum_{i=1}^{j-1} \alpha_i$ for $j = 1, \dots, \ell(\alpha)$. Given a countable set A, a **multiset** $S \subset A$ is a collection of elements in A, allowing for repetitions. We denote the family of multisets of elements in A by MS_A^0 . This includes the empty multiset. We denote $MS_A := MS_A^0 \setminus \{\emptyset\}$. When A = [d], we denote MS_A^0 , MS_A by MS_d^0 , MS_d , respectively.

The alphabet MS_d as a multiset has an associative and commutative operation \forall , the union. Note that in the context of multisets, the union counts multiplicity of elements.

Tensor algebras. Let V be a vector space over a characteristic zero field K. We define T(V), the **tensor algebra**, and T(V), the **tensor series** over V as follows:

(3)
$$T(V) \coloneqq \bigoplus_{k=0}^{+\infty} V^{\otimes k} \qquad T((V)) \coloneqq \prod_{k=0}^{+\infty} V^{\otimes k} ,$$

with $V^{\otimes 0} := K$. By fixing a basis $\mathcal{B} = \{ e_i : i \in I \}$ of V we can write elements of T(V) and T(V) as finite linear combinations and formal series of elements in $\mathcal{W}(I)$, respectively, via the identification

$$i_1 \bullet \cdots \bullet i_k = \mathsf{e}_{i_1} \otimes \cdots \otimes \mathsf{e}_{i_k}$$
.

Define the \bullet operation $(w, v) \mapsto w \otimes v$. It can be seen that this map is well defined on both T(V) and T((V)), because it is locally finite. Both T(V) and T((V)) form an algebra under •. The algebras $(T(V), \bullet)$ and $(T((V)), \bullet)$ are called **algebra of tensors** and **algebra of tensor series**, respectively. We define a bilinear and non-singular bracket $\langle -, - \rangle : T((V)) \times T(V) \to K$ as

(4)
$$\left\langle \sum_{w \in \mathcal{W}(I)} \alpha_w w, v \right\rangle = \alpha_v \,.$$

In this way, we can identify T((V)) with the algebraic dual of T(V).

Equip the indexing set I of the basis of V with a grading, such that there are finitely many elements of I with a given degree. Then V becomes a graded vector space, and we can write $V = \bigoplus_{h\geq 0} V^h$ for the corresponding grading. The vector space T(V) also inherits a grading, by setting deg $(w) = ||w||_{\text{ht}}$. Define

$$T^{h}(V) \coloneqq \operatorname{span}\{w \in \mathcal{W}(I) \colon ||w||_{\operatorname{ht}} = h\}, \quad T(V) = \bigoplus_{h=0}^{+\infty} T^{h}(V)$$
$$T^{\leq h}(V) = \operatorname{span}\{w \in \mathcal{W}(I) \colon ||w||_{\operatorname{ht}} \leq h\}, \quad T^{>h}(V) = \operatorname{span}\{w \in \mathcal{W}(I) \colon ||w||_{\operatorname{ht}} > h\}$$

The vector space $T^{\leq h}(V) = T(V)/_{T^{>h}(V)}$ is called the space of **truncated tensors**. Since $T^{>h}(V)$ is a \bullet ideal, the corresponding truncated tensor product is well defined on the quotient. We write \bullet_h for the product on the quotient. We always have the isomorphism

$$T^{\leq h}(V) \cong T^{\leq h}\left(V^{\leq h}\right) \,$$

where $V^{\leq h} = \bigoplus_{k=0}^{h} V^k$. These vector spaces are all finite dimensional.

The same truncation procedure applies to T((V)) and we obtain a vector space which is isomorphic to $T^{\leq h}(V)$. For sake of simplicity, we will denote this vector space with the same notation $T^{\leq h}(V)$.

Shuffle and quasi-shuffle Hopf algebras. Fix $d \ge 1$ integer, and consider henceforth $V = \mathfrak{A}_d = \mathbb{K}[X_1, \ldots, X_d]/_{\{\text{constant polynomials}\}}$ as our vector space of interest, on which we study its tensor algebra and tensor series.

The vector space \mathfrak{A}_d has a basis of non-constant monomials. These are identified with MS_d . We abuse notation and refer to a multiset $I \in \mathrm{MS}_d$ as a monomial in the variables X_1, \ldots, X_d . For example, we identify the monomial $X_1^2 X_2$ with the multiset 112. This abuse of notation extends to the evaluation of a monomial, thus writing $112(y_1, y_2, y_3) = y_1^2 y_2$. To distinguish elements of the alphabet MS_d and scalars in \mathbb{Z} , we use typewritter typeset for multisets. The alphabet MS_d graded, with degree given by the polynomial degree of the associated monic polynomial, or equivalently by the set size. Given any $h \geq 1$ we use the shorthand notation $\mathfrak{A}_{d,h} := (\mathfrak{A}_d)^{\leq h}$.

Two products can be defined on $T(\mathfrak{A}_d)$: the **shuffle** \blacksquare and the **quasi-shuffle** \blacksquare products. We define them here recursively. For any $i, j \in MS_d$ and $w, v \in \mathcal{W}(MS_d)$,

(5)

$$w = \varepsilon \overline{\amalg} w = w \overline{\amalg} \varepsilon = \varepsilon \amalg w = w \amalg \varepsilon$$

$$w \bullet i \amalg v \bullet j = (w \amalg vj) \bullet i + (w \bullet i \amalg v) \bullet j$$

$$w \bullet i \overline{\amalg} v \bullet j = (w \overline{\amalg} vj) \bullet i + (w \bullet i \overline{\amalg} v) \bullet j + (w \overline{\amalg} v) \bullet ij$$

These relations define two commutative algebras on $T(\mathfrak{A}_d)$ which are compatible with the grading of $T(\mathfrak{A}_d)$ given above (see [Hof00, Theorem 2.1] for a proof of this fact).

The tensor algebra $T(\mathfrak{A}_d)$ can be further equipped with two structures of **Hopf** algebras by introducing the deconcatenation coproduct $\delta: T(\mathfrak{A}_d) \to T(\mathfrak{A}_d) \otimes T(\mathfrak{A}_d)$ and a matching counit $\eta^*: T(\mathfrak{A}_d) \to K$. For a word $w = i_1 \bullet \cdots \bullet i_k$, let

$$\delta(w) = \varepsilon \otimes w + w \otimes \varepsilon + \sum_{l=1}^{k-1} i_1 \bullet \cdots \bullet i_l \otimes i_{l+1} \bullet \cdots \bullet i_k, \quad \eta^*(w) := \begin{cases} 1 & \text{if } w = \varepsilon, \\ 0 & \text{otherwise.} \end{cases}$$

We define as well the reduced coproduct map $\tilde{\delta} = \delta - \mathrm{id} \otimes 1 - 1 \otimes \mathrm{id}$. This endows $(T(\mathfrak{A}_d), \mathbf{\underline{u}}, \delta)$ and $(T(\mathfrak{A}_d), \mathbf{\overline{u}}, \delta)$ with graded Hopf algebra structures. We expect there is no confusion between elements in T(V) and $T(V) \otimes T(V)$.

These Hopf algebras were shown to be isomorphic. Explicit algebra morphisms $\Phi_H, \Psi_H : T(\mathfrak{A}_d) \to T(\mathfrak{A}_d)$ were constructed in [Hof00], which are inverses of each other. Specifically, the maps Φ_H and Ψ_H are the linear maps that act on words as follows:

(6)
$$\Phi_{H}(w) := \sum_{\alpha \in C(|w|)} \frac{1}{\alpha!} (w)_{\alpha}, \qquad \Psi_{H}(w) := \sum_{\alpha \in C(|w|)} \frac{(-1)^{|w| - \ell(\alpha)}}{\Pi \alpha} (w)_{\alpha}.$$

For instance, we have the following identities

$$\begin{split} \Phi_H(1 \bullet 2) =& 1 \bullet 2 + \frac{1}{2} 12 , \qquad \Psi_H(1 \bullet 2) = 1 \bullet 2 - \frac{1}{2} 12 , \\ \Phi_H(1 \bullet 2 \bullet 3) =& 1 \bullet 2 \bullet 3 + \frac{1}{2} 12 \bullet 3 + \frac{1}{2} 1 \bullet 23 + \frac{1}{6} 123 , \\ \Psi_H(1 \bullet 2 \bullet 3) =& 1 \bullet 2 \bullet 3 - \frac{1}{2} 12 \bullet 3 - \frac{1}{2} 1 \bullet 23 + \frac{1}{3} 123 . \end{split}$$

The map Φ_H is a graded isomorphism from $(T(\mathfrak{A}_d), \mathbf{\underline{u}}, \delta)$ to $(T(\mathfrak{A}_d), \mathbf{\underline{u}}, \delta)$.

The adjoints of Φ_H , Ψ_H with respect to $\langle -, - \rangle$ are also explicitly described in [Hof00, Section 4.2]. Specifically, $\Phi_H^*, \Psi_H^* : T((\mathfrak{A}_d)) \to T((\mathfrak{A}_d))$ are maps satisfying the following identities

(7)

$$\begin{aligned}
\Phi_{H}^{*}(i_{1} \bullet \cdots \bullet i_{k}) &\coloneqq \Phi_{H}^{*}(i_{1}) \bullet \cdots \bullet \Phi_{H}^{*}(i_{k}), \\
\Psi_{H}^{*}(i_{1} \bullet \cdots \bullet i_{k}) &\coloneqq \Psi_{H}^{*}(i_{1}) \bullet \cdots \bullet \Psi_{H}^{*}(i_{k}), \\
\Phi_{H}^{*}(i) &\coloneqq \sum_{n \geq 1} \frac{1}{n!} \sum_{i_{1} \cdots i_{n} = i} i_{1} \bullet \cdots \bullet i_{n}, \quad \Psi_{H}^{*}(i) &\coloneqq \sum_{n \geq 1} \frac{(-1)^{n-1}}{n} \sum_{i_{1} \cdots i_{n} = i} i_{1} \bullet \cdots \bullet i_{n}.
\end{aligned}$$

for any $i_1, \ldots, i_k, i \in MS_d$. For instance, one has

$$\begin{split} \Phi_{H}^{*}(12) =& 12 + \frac{1}{2}\mathbf{1} \bullet 2 + \frac{1}{2}\mathbf{2} \bullet 1 \,, \qquad \Psi_{H}^{*}(12) = 12 - \frac{1}{2}\mathbf{1} \bullet 2 - \frac{1}{2}\mathbf{2} \bullet 1 \,, \\ \Phi_{H}^{*}(123) =& 123 + \frac{1}{2}\mathbf{12} \bullet 3 + \frac{1}{2}\mathbf{13} \bullet 2 + \frac{1}{2}\mathbf{23} \bullet 1 + \frac{1}{2}\mathbf{1} \bullet 23 \\ &\quad + \frac{1}{2}\mathbf{2} \bullet \mathbf{13} + \frac{1}{2}\mathbf{3} \bullet \mathbf{12} + \frac{1}{6}\sigma(\mathbf{1} \bullet \mathbf{2} \bullet 3) \\ \Psi_{H}^{*}(\mathbf{123}) =& \mathbf{123} - \frac{1}{2}\mathbf{12} \bullet 3 - \frac{1}{2}\mathbf{13} \bullet 2 - \frac{1}{2}\mathbf{23} \bullet \mathbf{1} - \frac{1}{2}\mathbf{1} \bullet \mathbf{23} \\ &\quad - \frac{1}{2}\mathbf{2} \bullet \mathbf{13} - \frac{1}{2}\mathbf{3} \bullet \mathbf{12} + \frac{1}{3}\sigma(\mathbf{1} \bullet \mathbf{2} \bullet \mathbf{3}) \,, \end{split}$$

where $\sigma(1 \bullet 2 \bullet 3) = 1 \bullet 2 \bullet 3 + 2 \bullet 1 \bullet 3 + 3 \bullet 2 \bullet 1 + 1 \bullet 3 \bullet 2 + 2 \bullet 3 \bullet 1 + 3 \bullet 1 \bullet 2$.

The Hopf algebras dual to $(T(\mathfrak{A}_d), \underline{\mathbf{u}}, \delta)$ and $(T(\mathfrak{A}_d), \overline{\mathbf{u}}, \delta)$ have the \bullet product, and have coproducts in T(V) that we denote by $\delta_{\underline{\mathbf{u}}}$, $\delta_{\overline{\mathbf{u}}}$, respectively. These were given explicitly in [Reu93]. In this way, Φ_H^*, Ψ_H^* are **graded Hopf algebra isomorphism** between $(T((\mathfrak{A}_d)), \bullet, \delta_{\overline{\mathbf{u}}})$ and $(T((\mathfrak{A}_d)), \bullet, \delta_{\underline{\mathbf{u}}})$, see *e.g.* [BFPP22, Proposition 3.5].

The maps Φ, Ψ, Φ^* and Ψ^* are graded. Therefore, for any $h \ge 1$ these maps restrict to isomorphisms. For instance, Φ^* is an isomorphism between $(T^{\le h}(\mathfrak{A}_d), \bullet, \delta_{\mathfrak{m}})$ and $(T^{\le h}(\mathfrak{A}_d), \bullet, \delta_{\overline{\mathfrak{m}}})$. We display these maps in Fig. 1, in the context of two important subspaces of $T((\mathfrak{A}_d))$, that we introduce in the next section.

3. LIE POLYNOMIALS AND GROUP-LIKE ELEMENTS

In this section, we introduce two fundamental subsets of $T^{\leq h}(\mathfrak{A}_d)$: the space of Lie polynomials and the variety of group-like elements. These will play a role in understanding discrete signatures and, as the name suggests, these will form a pair of Lie algebra $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ and Lie group $\mathcal{G}^{\leq h}(\mathfrak{A}_d)$.

We define the **Lie bracket** [-,-]: $T^{\leq h}(\mathfrak{A}_d) \otimes T^{\leq h}(\mathfrak{A}_d) \to T^{\leq h}(\mathfrak{A}_d)$ by setting

$$[v,w] \coloneqq w \bullet_h v - w \bullet_h v,$$

and extending it linearly. We define $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ as the smallest Lie subalgebra of $T^{\leq h}(\mathfrak{A}_d)$ containing \mathfrak{A}_d . Equivalently, $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ is the space of iterated Lie brackets starting from the finite dimensional vector space $\mathfrak{A}_{d,h}$ or \mathfrak{A}_d . We refer to this Lie algebra as the set of **height-h Lie polynomials**.

For a word $w = \mathbf{e}_1 \bullet \cdots \bullet \mathbf{e}_k$, write $u = \mathbf{e}_2 \bullet \cdots \bullet \mathbf{e}_k$ and define inductively the following Lie polynomials:

(8)
$$\mathfrak{l}_w = [\mathbf{e}_1, \mathfrak{l}_u], \qquad \mathfrak{l}_I = I, \qquad \mathfrak{l}_\varepsilon = 0$$

Notice how $\mathcal{L}^{\leq h}(\mathfrak{A}_d) \cap \mathfrak{A}_d = \mathfrak{A}_{d,h}$, so the principal elements of $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ are $\mathfrak{A}_{d,h}$, and $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ is finitely generated. We will now show that any height-*h* Lie polynomial is generated by the elements \mathfrak{l}_w .

We introduce an intermediary Lie algebra \mathfrak{L}^h that sits between $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ and T((V)). Consider the Lie bracket without truncation on T((V)):

$$[[v,w]] \coloneqq w \bullet v - w \bullet v \,.$$

The following is a Lie algebra

$$\mathfrak{L}^{h} \coloneqq \mathfrak{A}_{d,h} \oplus \left[\left[\mathfrak{A}_{d,h}, \mathfrak{A}_{d,h} \right] \right] \oplus \cdots \oplus \underbrace{\left[\left[\mathfrak{A}_{d,h}, \left[\left[\mathfrak{A}_{d,h}, \ldots, \left[\left[\mathfrak{A}_{d,h}, \mathfrak{A}_{d,h} \right] \right] \right]_{h-1 \text{ times}} \right] \right]_{h-1 \text{ times}},$$

where we set that any iterated sequence of h brackets vanishes. This follows [FV10, Definition 7.25] and is called the **free** h-step nilpotent Lie algebra. This intermediary algebra allows us to write $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ as a quotient:

Proposition 3.1. The space $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ is a Lie algebra. It arises as a quotient of a Lie algebra with a Lie ideal:

(9)
$$\mathcal{L}^{\leq h}(\mathfrak{A}_d) = \mathfrak{L}^h / \left(T^{>h}(\mathfrak{A}_d) \cap \mathfrak{L}^h \right) ,$$

where $\mathfrak{L}^{\leq h}(\mathfrak{A}_d)$ is defined above. Furthermore, one has the following identification

(10)
$$\mathcal{L}^{\leq h}(\mathfrak{A}_d) = \{ \mathbf{v} \in T_0^{\leq h}(\mathfrak{A}_d) : \langle \mathbf{v}, u \, \mathbf{u} \, k \rangle = 0 \quad \text{for all } ||u||_{\text{ht}} + ||k||_{\text{ht}} \leq h \}.$$

As a consequence, by projecting onto $T^{\leq h}(\mathfrak{A}_d)$ any basis of \mathfrak{L}^h one has a generating set for $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$. This gives us that $\{\mathfrak{l}_w : ||w||_{\mathrm{ht}} \leq h\}$ is a generating set for $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$.

Sketch of proof. We can use the fundamental property of \mathfrak{L}^h as a free *h*-step nilpotent Lie algebra (see [FV10, Remark 7.26]). This gives us an explicit Lie algebra morphism $\mathfrak{f}: \mathfrak{L}^h \to \mathcal{L}^{\leq h}(\mathfrak{A}_d).$

Degree considerations give us that this map \mathfrak{f} is surjective, and one can see that its kernel is $T^{>h}(\mathfrak{A}_d) \cap \mathfrak{L}^h$, which concludes the quotient claim. The equation (10) follows from [Reu93, Theorem 3.1].

For $k \in K$, let $T_k^{\leq h}(\mathfrak{A}_d) \coloneqq \{ \mathsf{v} \in T^{\leq h}(\mathfrak{A}_d) | \langle \mathsf{v}, \varepsilon \rangle = k \}$. We introduce the polynomial maps

$$\exp_{\bullet_h} \colon T_0^{\leq h}(\mathfrak{A}_d) \to T_1^{\leq h}(\mathfrak{A}_d) \quad \text{and} \quad \log_{\bullet_h} \colon T_1^{\leq h}(\mathfrak{A}_d) \to T_0^{\leq h}(\mathfrak{A}_d)$$

defined by

(11)
$$\exp_{\bullet_h}(\mathsf{v}) \coloneqq \sum_{n \ge 0} \frac{1}{n!} \mathsf{v}^{\bullet_h n}, \quad \log_{\bullet_h}(v) \coloneqq \sum_{n \ge 0} \frac{(-1)^{n-1}}{n} (\mathsf{v} - \varepsilon)^{\bullet_h n},$$

where $\mathbf{v}^{\bullet_h n} = \mathbf{v} \bullet_h \cdots \bullet_h \mathbf{v}$ stands for the *n*-th bullet product. The following was shown in [Reu93, Chapter 3]:

Proposition 3.2. For each w, the elements $\exp_{\bullet_h}(w)$ and $\log_{\bullet_h}(w)$ are finite sums. Moreover, \exp_{\bullet_h} is surjective and $\log_{\bullet_h} = \exp_{\bullet_h}^{-1}$.

We define the height-*h* free nilpotent Lie group $\mathcal{G}^{\leq h}(\mathfrak{A}_d)$:

$$\mathcal{G}^{\leq h}(\mathfrak{A}_d) \coloneqq \exp_{\bullet_h}(\mathcal{L}^{\leq h}(\mathfrak{A}_d)) \subset T_1^{\leq h}(\mathfrak{A}_d),$$

which is a Lie group when endowed it with the operation \bullet_h . Let $\pi^h : T((\mathfrak{A}_d)) \to T^h(\mathfrak{A}_d)$ be the canonical projection. To describe varieties of discrete signature we also introduce the following set

$$\mathcal{G}^h(\mathfrak{A}_d) \coloneqq \pi^h(\mathcal{G}^{\leq h}(\mathfrak{A}_d)).$$

The following is a consequence of Proposition 3.1, as well as some classical properties of free Nilpoltent Lie algebras. See e.g. [Reu93, Theorem 3.2] for details on this.

Proposition 3.3. The elements of $\mathcal{G}^{\leq h}(\mathfrak{A}_d)$ are characterized by:

(12)
$$\mathcal{G}^{\leq h}(\mathfrak{A}_d) = \left\{ \mathbf{v} \in T_1^{\leq h}(\mathfrak{A}_d) : \langle \mathbf{v}, w \, \mathbf{i} \, \mathbf{u} \rangle = \langle \mathbf{v}, w \rangle \langle \mathbf{v}, u \rangle \text{ for } ||w||_{\mathrm{ht}} + ||u||_{\mathrm{ht}} \leq h \right\}.$$

Remark that these equations are all polynomial (quadratic) equations on the entries of v.

We now introduce Chow theorem, which expresses elements of $\mathcal{G}^{\leq h}(\mathfrak{A}_d)$ as concatenation of simpler element.

Theorem 3.4 ([FV10, Theorem 7.28]). For any height $h \ge 1$ and $g \in \mathcal{G}^{\le h}(\mathfrak{A}_d)$ there exists an integer m and $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathfrak{A}_{d,h}$ such that

$$g = \exp_{\bullet_h}(\mathsf{v}_1) \bullet_h \cdots \bullet_h \exp_{\bullet_h}(\mathsf{v}_m)$$



FIGURE 1. The height h free Lie algebras and the height h Lie groups are connected via Φ^* and Ψ^* .

We now turn our attention to the quasi-shuffle relations, and define an analogous space to the one presented in Proposition 3.3, in the $\overline{\mathbf{u}}$ context. Note that this is the variety of interest for this paper, as the signature of a time-series is an element of this space, according to Theorem 4.4.

Definition 3.5. For any integer $h \ge 1$ we define the height-*h* free quasi-shuffle Lie group $\hat{\mathcal{G}}^{\le h}(\mathfrak{A}_d)$ and height-*h* quasi-shuffle Lie polynomials $\hat{\mathcal{L}}^{\le h}(\mathfrak{A}_d)$ as

$$\hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d) \coloneqq \left\{ \mathbf{v} \in T_1^{\leq h}(\mathfrak{A}_d) \colon \langle \mathbf{v}, w \overline{\mathbf{u}} u \rangle = \langle \mathbf{v}, w \rangle \langle \mathbf{v}, u \rangle \text{ for all } ||w||_{\mathrm{ht}} + ||u||_{\mathrm{ht}} \leq h \right\} ,$$
$$\hat{\mathcal{L}}^{\leq h}(\mathfrak{A}_d) \coloneqq \log_{\bullet_h}(\hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)) \subset T_0^{\leq h}(\mathfrak{A}_d) .$$

Similarly as before, we introduce the set

$$\hat{\mathcal{G}}^h(\mathfrak{A}_d) \coloneqq \pi^h(\hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d))$$
.

Recall that Φ_H and Ψ_H , defined in (6), and their adjoints, are Hopf algebra isomorphisms. In [Hof00, Theorem 4.2] and [BFPP22] it is shown that:

Proposition 3.6. The function Φ_H^* maps isomorphically $\hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)$ to $\mathcal{G}^{\leq h}(\mathfrak{A}_d)$ and $\hat{\mathcal{L}}^{\leq h}(\mathfrak{A}_d)$ to $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$. The maps Φ_H^*, Ψ_H^* commute with $\exp_{\bullet_h}, \log_{\bullet_h}$ on these domains.

Summing up the relation in a commutative diagram we can describe the properties of Φ_H^* and Ψ_H^* in Figure 1 above.

We conclude the section by defining an explicit adjoint of \log_{\bullet_h} . These are the **truncated shuffle eulerian map** and **truncated quasi-shuffle eulerian map**, the maps $e_1^{\underline{\mathsf{u}}}, e_1^{\underline{\mathsf{u}}} : T^{\leq h}(\mathfrak{A}_d) \to T^{\leq h}(\mathfrak{A}_d)$ defined as

$$e_1^{\mathbf{u}} = \sum_{n \ge 1} \frac{(-1)^{n-1}}{n} \, \mathbf{u}^{\circ(n-1)} \circ \tilde{\delta}^{\circ(n-1)} \,, \quad e_1^{\mathbf{u}} = \sum_{n \ge 1} \frac{(-1)^{n-1}}{n} \overline{\mathbf{u}}^{\circ(n-1)} \circ \tilde{\delta}^{\circ(n-1)} \,.$$

where we use the convention that $\overline{\mathbf{m}}^{\circ(0)} \circ \widetilde{\delta}^{\circ(0)} = \mathbf{m}^{\circ(0)} \circ \widetilde{\delta}^{\circ(0)}$ is the projection from $T^{\leq h}(\mathfrak{A}_d)$ to εK . For instance, one has

$$\begin{split} e_1^{\mathbf{u}}(\varepsilon) &= e_1^{\overline{\mathbf{u}}}(\varepsilon) = 0\,, \qquad e_1^{\mathbf{u}}(\mathbf{i}) = e_1^{\overline{\mathbf{u}}}(\mathbf{i}) = \mathbf{i}\,, \\ e_1^{\mathbf{u}}(\mathbf{i} \bullet \mathbf{j}) &= \frac{1}{2}(\mathbf{i} \bullet \mathbf{j} - \mathbf{j} \bullet \mathbf{i})\,, \quad e_1^{\overline{\mathbf{u}}}(\mathbf{i} \bullet \mathbf{j}) = \frac{1}{2}(\mathbf{i} \bullet \mathbf{j} - \mathbf{j} \bullet \mathbf{i}) + \frac{1}{2}\mathbf{i}\mathbf{j} \end{split}$$

These two maps allow to compute the adjoint of \log_{\bullet_h} on group-like elements.

Proposition 3.7. Let $v \in \mathcal{G}^{\leq h}(\mathfrak{A}_d)$ and $w \in \hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)$. For any word w with $||w||_{ht} \leq h$ one has:

(13)
$$\langle \log_{\bullet_h}(\mathbf{v}), w \rangle = \langle \mathbf{v}, e_1^{\mathbf{u}}(w) \rangle , \\ \langle \log_{\bullet_h}(\mathbf{w}), w \rangle = \langle \mathbf{w}, e_1^{\mathbf{u}}(w) \rangle .$$

We use the following fact, called the **duality between product and coproduct** associated to \bullet and δ , without proof:

$$\langle (\mathbf{v} - \varepsilon)^{\bullet_h n}, w \rangle = \langle \mathbf{v}^{\otimes n}, \tilde{\delta}^{\circ (n-1)} w \rangle_{T(\mathfrak{A}_d)^{\otimes n}},$$

Proof of Proposition 3.7. Using the duality above, and applying Proposition 3.3 we have:

$$\begin{split} \langle \log_{\bullet_{h}}(\mathbf{v}), w \rangle &= \sum_{n \ge 1} \frac{(-1)^{n-1}}{n} \langle (\mathbf{v} - \varepsilon)^{\bullet_{h}n}, w \rangle = \sum_{n \ge 1} \frac{(-1)^{n-1}}{n} \langle \mathbf{v}^{\otimes n}, \tilde{\delta}^{\circ(n-1)} w \rangle_{T(\mathfrak{A}_{d})^{\otimes n}} \\ &= \sum_{n \ge 1} \frac{(-1)^{n-1}}{n} \langle \mathbf{v}, \mathbf{u}^{\circ(n-1)} \tilde{\delta}^{\circ(n-1)} w \rangle \\ &= \langle \mathbf{v}, \sum_{n \ge 1} \frac{(-1)^{n-1}}{n} \mathbf{u}^{\circ(n-1)} \tilde{\delta}^{\circ(n-1)} w \rangle = \langle \mathbf{v}, e_{1}^{\mathbf{u}}(w) \rangle \,, \end{split}$$

which concludes one equality. The remaining follows via the same computations. \Box

4. VARIETIES OF DISCRETE SIGNATURES

In what follows we consider K to be an algebraically closed field. Fix integers $d, N \ge 1$ and we consider time-series $x = (x_1, \ldots, x_N)$ of elements in \mathbb{K}^d . Denote $\Delta x_i = x_{i+1} - x_i \in \mathbb{K}^d$. The discrete signature tensor in $T((\mathfrak{A}_d))$ is given by

(14)
$$\langle \mathfrak{S}(x), w \rangle \coloneqq \sum_{1 \le i_1 < \cdots < i_k < N} \mathsf{e}_1(\Delta x_{i_1}) \cdots \mathsf{e}_k(\Delta x_{i_k}),$$

for all $w = \mathbf{e}_1 \bullet \cdots \bullet \mathbf{e}_k \in \mathcal{W}(\mathrm{MS}_d)$. A first observation is that the discrete signature of a time-series x is invariant up to translations. We will therefore reparametrize (14) and only consider the signature of $y = \Delta x$, the difference time-series. We abuse notation and set, for any word $w = \mathbf{e}_1 \bullet \cdots \bullet \mathbf{e}_k$, that:

(15)
$$\langle \mathfrak{S}(y), w \rangle \coloneqq \sum_{1 \le i_1 < \cdots < i_k \le N} \mathsf{e}_1(y_{i_1}) \cdots \mathsf{e}_k(y_{i_k}) \, .$$

This reparametrisation allows for a more tractable computer assisted calculation, as it reduces de input space of S. Crucially, the image of this map is still the same. We will use the same notation for this reparametrisation. Giving $h \ge 1$ we denote the projection of the signature S(x) onto $T^{\le h}(\mathfrak{A}_d)$ by $S^{\le h}(x)$, and denote the projection of S(x) onto $T^h(\mathfrak{A}_d)$ by $S^h(x)$.

We identify a time-series in $(\mathbf{K}^d)^N$ with a $d \times N$ matrix in K. If $x = (x_1, \ldots, x_N)$, we identify x with a matrix whose *i*-th column is x_i . So we have for example

$$S^{\leq 2} \left(\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \end{bmatrix} \right) = \varepsilon + 6\mathbf{1} + 7\mathbf{2} + 14\mathbf{11} + 14\mathbf{12} + 17\mathbf{22} + 11\mathbf{1} \bullet \mathbf{1} + 9\mathbf{1} \bullet \mathbf{2} + 19\mathbf{2} \bullet \mathbf{1} + 16\mathbf{2} \bullet \mathbf{2}.$$

Definition 4.1 (The signature varieties and the universal variety). Fix d, h, N integers. We define the **discrete signature variety** $\mathcal{V}_{d,h,N}$ to be the Zariski closure of the image of S^h in $T^h(\mathfrak{A}_d)$.

Remark 4.2. Recall that we are considering K a algebraically closed field. Furthermore, $\mathcal{V}_{d,h,N}$ is a homogeneous, so its projectivisation is Zariski closed (see [SR94, Section 5.2]). This allows us to simply word with the image of the map from now on.

We observe that when using the new parametrisation, adding zero-columns to a matrix does not change $S^{\leq 2}$. For instance we have:

$$S^{\leq 2} \left(\begin{bmatrix} 1 & 0 & 2 & 3 & 0 \\ 2 & 0 & 3 & 2 & 0 \end{bmatrix} \right) = \varepsilon + 6\mathbf{1} + 7\mathbf{2} + 14\mathbf{11} + 14\mathbf{12} + 17\mathbf{22} + 11\mathbf{1} \cdot \mathbf{1} + 9\mathbf{1} \cdot \mathbf{2} + 19\mathbf{2} \cdot \mathbf{1} + 16\mathbf{2} \cdot \mathbf{2}.$$

As a consequence, we have the following ascending chain of varieties

(16)
$$\mathcal{V}_{d,h,0} \subseteq \mathcal{V}_{d,h,1} \subseteq \mathcal{V}_{d,h,2} \subseteq \cdots$$

we call $\mathcal{V}_{d,h}$ to the union of these varieties. These are varieties given by parametrisations, therefore they are irreducible. According to the variety-ideal dictionary (see for instance [CLO06]), these correspond to a descending chain of prime ideals.

The following fact is a consequence of **Krull's principal ideal theorem**.

Proposition 4.3. A descending chain of prime ideals in a polynomial ring on finite variables must stabilize.

Therefore, $\mathcal{V}_{d,h} = \mathcal{V}_{d,h,N_{d,h}}$ for some finite integer $N_{d,h}$. These are called the **universal** varieties of the discrete signature. We explore universal varieties in Section 5.

Theorem 4.4 (Quasi-shuffle relations). The signature of a time-series satisfies the quasi-shuffle relations. That is, for any two words w, v in MS_d we have

$$\langle \mathfrak{S}(x), w \overline{\mathfrak{u}} v \rangle = \langle \mathfrak{S}(x), w \rangle \langle \mathfrak{S}(x), v \rangle$$

This means that $\mathcal{S}^h(x) \in \hat{\mathcal{G}}^h(\mathfrak{A}_d)$. In particular, $\mathcal{V}_{d,h} \subseteq \hat{\mathcal{G}}^h(\mathfrak{A}_d)$. This was shown in [DET20, Theorem 3.4]. In Conjecture 5.1 we conjecture that this inclusion is tight.

4.1. The height one varieties. For completeness sake we include this case here. For h = 1, the map S covers the height one component $T^1(\mathfrak{A}_d)$. In this case there are no relations, and for $N \ge 1$, we have that $\mathcal{V}_{d,1,N} = \mathcal{V}_{d,1,1} \cong \mathbb{K}^d$.

4.2. The height two varieties. The first non-trivial quasi-shuffle relation arises at height two. Specifically the rank of the following matrix is at most one:

$$\begin{bmatrix} \mathbb{S}_{\mathbf{e}}(x)^2 & \mathbb{S}_{\mathbf{e}}(x) \, \mathbb{S}_{\mathbf{f}}(x) \\ \mathbb{S}_{\mathbf{e}}(x) \, \mathbb{S}_{\mathbf{f}}(x) & \mathbb{S}_{\mathbf{f}}(x)^2 \end{bmatrix} \, .$$

This gives us the following equation on height two discrete signatures

(17) $(2 \,\mathfrak{S}_{\mathsf{e}\bullet\mathsf{e}} + \mathfrak{S}_{\mathsf{e}\mathsf{e}}) \,(2 \,\mathfrak{S}_{\mathsf{f}\bullet\mathsf{f}} + \mathfrak{S}_{\mathsf{f}\mathsf{f}}) = (\mathfrak{S}_{\mathsf{e}\mathsf{f}} + \mathfrak{S}_{\mathsf{f}\bullet\mathsf{e}} + \mathfrak{S}_{\mathsf{e}\bullet\mathsf{f}})^2 \;.$

This equation determines $S_{f \bullet e}$ given S_{ef} and $S_{e \bullet f}$, for $e \neq f$. In Section 5 we show that, $\hat{\mathcal{G}}^2(\mathfrak{A}_d)$ is precisely the universal variety. By dimension counting, together with Corollary 5.7 we get the following generator result:

Theorem 4.5. The quadratic equations given in (17) generate the variety $\hat{\mathcal{G}}^h(\mathfrak{A}_d)$. Furthermore, because the equations in (17) are all transversal, the degree of this variety is $2^{\binom{d}{2}}$.

4.3. Paths on the line. We now look at the case where d = 1. Here, \mathfrak{A}_d has a basis element for each degree, and $T^h(\mathfrak{A}_d)$ has a basis element indexed by compositions of h. Therefore, it has dimension 2^{h-1} for $h \ge 1$. This is the ambient space of $\mathcal{V}_{1,h,N}$. However, one can see that $\langle S_{d,N}(x), w \rangle$ is an evaluation of a quasi-symmetric function:

Example 4.6. First, we see $w \in \mathcal{W}(MS_1)$ as compositions of $||w||_{ht}$. For instance, if $w = 11 \bullet 1 \bullet 11$, we identify w with the composition (2, 1, 2) of 5.

Fix h, N integers, and let $w \in \mathcal{W}(MS_1)$ be seen as a composition. Write M_w for the monomial quasi-symmetric function indexed by the composition corresponding to w, that is

$$M_{(\alpha_1,\ldots,\alpha_k)} = \sum_{1 \le i_1 < \cdots < i_k} x_{i_1}^{\alpha_1} \cdots x_{i_k}^{\alpha_k} ,$$

Then $\langle S_{d,N}(x), w \rangle$ is the evaluation of M_w on x appended with zeroes.

For instance, if we consider $w = 1 \bullet 11$, this corresponds to the composition $\alpha = (1, 2)$. If x is the time-series (1, 4, 2, 3) in \mathbb{R}^1 , then

$$\langle S(x), w \rangle = M_{(1,2)}(1, 4, 2, 3, 0, 0, 0, \dots) = 1 \cdot 4^2 + 1 \cdot 2^2 + 1 \cdot 3^2 + 4 \cdot 2^2 + 4 \cdot 3^2 + 2 \cdot 3^2.$$

We recall that [Haz01] has shown that QSym, the algebra of quasi-symmetric functions, is freely generated over the integers, and indeed gives a generating set for this algebra. We add that this result is independent of the characteristic of K. A simple consequence of this is that the dimension of our varieties of interest are the ones predicted by Conjecture 5.1.

Proposition 4.7. The variety $\mathcal{V}_{1,h,N}$ has the dimension expected in Conjecture 5.1.

We analyse here further particular cases with d = 1. For h = 3 and N = 3, the signature variety is embedded in K^4 and has dimension three: it is the image of the map $K^3 \rightarrow K^4$ explicitly given by

 $(a_1, a_2, a_3) \mapsto (a_1 a_2 a_3, a_1 a_1 a_2 + a_1 a_1 a_3 + a_2 a_2 a_3, a_1 a_2 a_2 + a_1 a_3 a_3 + a_2 a_3 a_3, a_1 a_2 a_3).$

Let us denote by $(s_{00}, s_{01}, s_{10}, s_{11})$ the coordinates of K⁴. The variety $\mathcal{V}_{1,3,3}$ is generated by one equation of degree nine, here is an excerpt of this equation:

$$81s_{00}^{9} + 162s_{00}^{8}s_{01} + 351s_{00}^{7}s_{01}^{2} + 333s_{00}^{6}s_{01}^{3} + 72s_{00}^{5}s_{01}^{4} - 63s_{00}^{4}s_{01}^{5} - 30s_{00}^{3}s_{01}^{6} + 6s_{00}^{2}s_{01}^{7} + 6s_{00}s_{01}^{8} + s_{01}^{9} + 162s_{00}^{8}s_{10} - 30s_{00}^{2}s_{01}s_{10}^{6} + \dots + 4s_{00}^{3}s_{10}^{2}s_{11}^{4} + 4s_{00}^{2}s_{01}s_{10}^{2}s_{11}^{4} + s_{00}s_{01}^{2}s_{10}^{2}s_{11}^{4}.$$

For h = 4 and N = 3, the variety $\mathcal{V}_{1,4,3}$ is embedded in \mathbb{K}^8 and is generated by 20 polynomials, whose degree counts are the following:

The degree one polynomial is s_{000} , the degree two polynomial is:

$$s_{001}^{2} + 2s_{001}s_{010} + s_{010}^{2} + 2s_{001}s_{011} + 2s_{010}s_{011} + s_{011}^{2} + 2s_{001}s_{100} + 2s_{010}s_{100} + 2s_{011}s_{100} + s_{100}^{2} - 4s_{001}s_{101} - 4s_{010}s_{101} - 4s_{100}s_{101} - 2s_{101}^{2} + 2s_{001}s_{110} + 2s_{010}s_{110} + 2s_{011}s_{110} + 2s_{010}s_{110} + 2s_{010}s_{110} + 2s_{010}s_{110} - 2s_{100}s_{111} - 2s_{100}s_{111} - 2s_{100}s_{111} - s_{101}s_{111}.$$

4.4. Three steps. Let us now focus on N = 3, d = 2 and h = 3. To write out the parametrization of $\mathcal{V}_{2,3,3}$, we rename some coordinates of $\mathfrak{A}_{2,3}$ as follows:

- For $w = \mathbf{e}_1 \bullet \mathbf{e}_2 \bullet \mathbf{e}_3$, we write $s_{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3}$ for $\langle \mathfrak{S}(x), w \rangle$.
- For $w = \mathbf{e}_1 \mathbf{e}_2 \bullet \mathbf{e}_3$, we write $t_{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3}$ for $\langle S(x), w \rangle$.
- For $w = \mathbf{e}_1 \bullet \mathbf{e}_2 \mathbf{e}_3$, we write $u_{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3}$ for $\langle \mathfrak{S}(x), w \rangle$.
- For $w = \mathbf{e}_1 \mathbf{e}_2 \mathbf{e}_3$, we write $v_{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3}$ for $\langle S(x), w \rangle$.

If
$$x = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix}$$
, then

$$\begin{aligned} s_{i,j,k} = a_i b_j c_k \\ t_{i,j,k} = a_i a_j b_k + a_i a_j c_k + b_i b_j c_k \\ u_{i,j,k} = a_i b_j b_k + a_i c_j c_k + b_i c_j c_k \\ v_{i,j,k} = a_i a_j a_k + b_i b_j b_k + c_i c_j c_k \end{aligned}$$

Proposition 4.8. There are 226 minimal generators of $\mathcal{I}(\mathcal{V}_{2,3,3})$ of degree at most four. These break down into 58 quadrics, 74 cubic and 134 quartic generators.

We present code for this fact in [BP23]. Here is one of the cubics generating $\mathcal{I}(\mathcal{V}_{2,3,3})$.

 $s_{121}s_{222}v_{222} - s_{121}t_{222}u_{222} - s_{122}s_{222}v_{122} + s_{122}t_{222}u_{212}$ $- s_{221}s_{222}v_{122} + s_{221}t_{122}u_{222} + s_{222}^2v_{112} - s_{222}t_{122}u_{212}.$

5. Universal varieties

This section relates to a conjecture measuring the difference between the universal variety and $\hat{\mathcal{G}}^h(\mathfrak{A}_d)$, as well as some consequences of this. Specifically, this conjecture guarantees that \mathcal{S}^h : $(\mathbb{K}^d)^N \to \hat{\mathcal{G}}(\mathfrak{A}_d)$ maps surjectively to $\hat{\mathcal{G}}^h(\mathfrak{A}_d)$, when taking N sufficiently large.

In this section we outline a proof strategy for this conjecture. This strategy entails showing that many elements of \mathfrak{A}_d are so called **reachable**. We show that for length at most two, the reachability conditions are satisfied. We display the equations necessary to solve the reachability problem for length three.

In the rest of this section we present some important consequences of this conjecture. Specifically, we present some dimension results and an enumeration result.

Conjecture 5.1. The universal variety $\mathcal{V}_{d,h}$ is precisely $\hat{\mathcal{G}}^h(\mathfrak{A}_d)$.

5.1. The dimension conjecture. The following definition and lemma outline a strategy for showing Conjecture 5.1. Recall from Eq. (8) that for a word $w = \mathbf{e}_1 \bullet \cdots \bullet \mathbf{e}_k$, we write the corresponding Lie polynomial as \mathfrak{l}_w .

Definition 5.2. Fix d, h, N and an element $\mathbf{v} \in \mathfrak{A}_d$ of degree at most h. We say that \mathbf{v} is (d, h, N)-reachable if there is a time-series $x \in (\mathbf{K}^d)^N$ that satisfies the following equations for all words w such that $||w||_{\text{ht}} \leq h$ and $|w| \geq 2$, and monomials $I \in \text{MS}_d$ of degree at most h:

$$\begin{split} \langle \mathbb{S}(x),I\rangle &= \langle \mathsf{v},I\rangle \\ \langle \mathbb{S}(x),\Phi_{H}e_{1}^{\mathsf{u}}(\mathfrak{l}_{w})\rangle &= 0\,. \end{split}$$

Lemma 5.3. A given $v \in \mathfrak{A}_d$ is (d, h, N)-reachable if and only if there is a time-series $x \in (\mathbf{K}^d)^N$ such that

$$\log_{\bullet_h} \Phi_H^* \, \mathcal{S}^{\leq h}(x) = \mathsf{v} \, .$$

Proof. We first observe that $\log_{\bullet_h} \Phi_H^* S^{\leq h}(x) = \mathsf{v}$ if and only if $\langle \Phi_H^* \log_{\bullet} S(x), \mathsf{b} \rangle = \langle \mathsf{v}, \mathsf{b} \rangle$ for all b running over a generating set of $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$. We use the set described in Proposition 3.1. By using the fact that $S(x) \in \hat{\mathcal{G}}(\mathfrak{A}_d)$, Proposition 3.6 tells us that $\Phi_H^* S(x) \in \mathcal{G}(\mathfrak{A}_d)$ so Proposition 3.7 gives us

$$\langle \log_{\bullet} \Phi_{H}^{*} \, \mathfrak{S}(x), \mathfrak{l}_{w} \rangle = \langle \Phi_{H}^{*} \, \mathfrak{S}(x), e_{1}^{\mathsf{u}}(\mathfrak{l}_{w}) \rangle = \langle \mathfrak{S}(x), \Phi_{H} e_{1}^{\mathsf{u}}(\mathfrak{l}_{w}) \rangle .$$

We obtain the desired equations once we note that whenever |w| = 1, $e_1^{\tt m}\mathfrak{l}_w = w$. Furthermore, the map $\Phi_H e_1^{\tt m}$ is graded, so $\langle \mathsf{v}, \Phi_H e_1^{\tt m}(w) \rangle = 0$ for |w| > 1.

We call the equations in Definition 5.2 the **reachability** equations. These are split into **levels** according to the length of w. In this way, the reachability equations of level k correspond to all equations where w words have length k. For instance, the reachability equations of lenght 1 are the non-homogeneous equations. Our strategy is to show that all v of degree at most h are reachable: the following lemma establishes that this is enough to infer Conjecture 5.1.

Lemma 5.4. Assume that there exists some N such that any $\mathbf{v} \in \mathfrak{A}_d$ of degree at most h is (d, h, N)-reachable. Consider $S^{\leq h}$ as a map $S^{\leq h} : (\mathbf{K}^d)^m \to \hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)$. Then, for some integer m, we have im $S^{\leq h} = \hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)$.

If $a \in (\mathbb{K}^d)^N$, $b \in (\mathbb{K}^d)^M$ are two time-series, we denote its concatenation by a|b. Specifically, it denotes the times-series in \mathbb{K}^d with M + N vectors resulting from appending b to a. For the proof of this lemma we use the following fact without proof:

Lemma 5.5. We have:

$$S(a|b) = S(a) \bullet S(b) ,$$

where we are using the parametrisation described in Eq. (15).

Proof of Lemma 5.4. Assume that for each $\mathbf{v} = \sum_{I \in \mathrm{MS}_d} \mathbf{v}_I I$ of degre at most h there exists a time-series $x = x(\mathbf{v})$ in $(\mathbf{K}^d)^N$ such that $\log_{\bullet_h} \Phi_H^* S^{\leq h}(x) = \mathbf{v}$. Theorem 3.4 guarantees that there exists some integer m such that for any $\mathbf{w} \in \mathcal{G}^{\leq h}(\mathfrak{A}_d)$ we can find $\mathbf{v}^1, \ldots, \mathbf{v}^m \in \mathfrak{A}_d$ satisfying

$$\mathsf{w} = \exp_{\bullet_h}(\mathsf{v}^1) \bullet_h \cdots \bullet_h \exp_{\bullet_h}(\mathsf{v}^m).$$

Consider the time-series $x = x(\mathbf{v}^1) | \cdots | x(\mathbf{v}^m)$ in $(\mathbf{K}^d)^{mN}$. Then Lemma 5.5, together with the fact that Φ_H^* is an algebra homomorphism (see Eq. (7)) gives us:

 $\Phi_H^* \, \mathbb{S}^{\leq h}(x) = \Phi_H^* \, \mathbb{S}^{\leq h}(x(\mathbf{v}^1)) \bullet_h \cdots \bullet_h \Phi_H^* \, \mathbb{S}^{\leq h}(x(\mathbf{v}^m)) = \mathbf{w} \, .$

This shows that the map $\Phi_H^* \circ S^{\leq h}$ is surjective. Because Φ_H^* is an isomorphism, we have that $S^{\leq h}$ is surjective, as desired.

Lemma 5.6. Assume that K is algebraically closed, and fix d, h integers. There exists an N such any $\mathbf{v} \in \mathfrak{A}_d$ of degree at most h is (d, h, N)-reachable. That is, for any \mathbf{v} we can find a time-series that satisfies the reachability equations in Definition 5.2 of length at most two.

Corollary 5.7. Conjecture 5.1 holds for h = 2.

Proof of Lemma 5.6. We argue as follows: First we construct a time-series $x = x(\mathbf{v})$ that satisfies the equations of length one. That is, for any element $\mathbf{v} \in \mathfrak{A}_d$ of degree at most h, the time-series satisfies

 $\langle S(x), I \rangle = \langle v, I \rangle$ for all I monomial.

This concludes the level one. From these solutions we construct solutions to the level two reachability equations, which are amenable to algebraic manipulations on the level one. This allows us to construct the desired solution.

Let us find the time-series $x(\mathbf{v})$ for each \mathbf{v} by a dimension argument. Let \mathcal{U} be a basis of $\mathfrak{A}_{d,h}$. Let Y be the variety in $K^{\mathcal{U}}$ given by the image of the following map from K^d :

$$\vec{x} \mapsto (p(\vec{x}))_{p \in \mathcal{U}}$$

This is an irreducible variety. Say it has dimension e. Fix $N = \lceil |\mathcal{U}|/_e \rceil$. The linear relations $\sum_{j=1}^{N} x_j^i = \mathbf{v}_i$ for $i \in \mathcal{U}$ trace out an affine space L of codimention $|\mathcal{U}|$ in K^N . Given that Y^N has dimension $eN \geq |\mathcal{U}|$, and since Y^N is not contained in any hyperplane, the intersection $L \cap Y^N$ is non-empty. The time series x in this intersection space is the desired $x(\mathbf{v})$.

To establish the equations on level two, for a time series a let \overline{a} be the reversed-time time-series of a. Then, we show that for any time-series a, the time-series $g(a) = a | \overline{a} |$ satisfies the equations of height two, while having an amenable level one result, that is:

$$\langle \mathfrak{S}(g(a)), I \rangle = 2^{\deg I} \langle \mathfrak{S}(a), I \rangle$$
 for all I monomial.
 $\langle \mathfrak{S}(g(a)), \Phi_H^* e_1^{\mathsf{u}}(\mathfrak{l}_w) \rangle = 0$ for all w of length two.

It follows that $g(2^{-1}v)$ satisfies the desired equations, and the lemma is proven.

5.2. Length three equations. We remark that the map $e_1^{\mathfrak{u}}$ is a projection in $\mathcal{L}(\mathfrak{A}_d)$ for any element of $\mathcal{L}^{\leq h}(\mathfrak{A}_d)$ of length one, two and three. This is why the equations of level one and two, as well as the equations of level three, as we will see below, are tractable. However, it is not true that $e_1^{\mathfrak{u}}$ is the identity in $\mathcal{L}(\mathfrak{A}_d)$, and counter examples arise for level four and above.

Example 5.8 (The level three reachability equations). Here we present Definition 5.2 for $w = 1 \bullet 2 \bullet 3$ and $w = 12 \bullet 3 \bullet 4$, which is a generic element of length three. That is, we compute $\Phi_H^* e_1^{\mathsf{u}}(\mathfrak{l}_w)$. From the same token as in Lemma 5.6, finding a time-series that has

$$\langle \mathfrak{S}(x), \Phi_H^* e_1^{\mathbf{u}}(\mathfrak{l}_w) \rangle = 0,$$

implies Conjecture 5.1 for h = 3.

From the remark above we have that $e_1^{\mathfrak{u}}(\mathfrak{l}_w) = \mathfrak{l}_w$. Therefore, by using the examples given after Eq. (7) we get:

For
$$w = 1 \cdot 2 \cdot 3$$

$$\Phi_{H}^{*} e_{1}^{\mathsf{m}}(\mathfrak{l}_{w}) = \Phi_{H}^{*}(\mathfrak{l}_{w})$$

$$= \Phi_{H}^{*}(1 \cdot 2 \cdot 3 - 1 \cdot 3 \cdot 2 - 2 \cdot 3 \cdot 1 + 3 \cdot 2 \cdot 1)$$

$$= 1 \cdot 2 \cdot 3 - 1 \cdot 3 \cdot 2 - 2 \cdot 3 \cdot 1 + 3 \cdot 2 \cdot 1 = \mathfrak{l}_{w}.$$
For $w = 12 \cdot 3 \cdot 4$

$$\Phi_{H}^{*} e_{1}^{\mathsf{m}}(\mathfrak{l}_{w}) = \Phi_{H}^{*}(\mathfrak{l}_{w})$$

$$= \Phi_{H}^{*}(12 \cdot 3 \cdot 4 - 12 \cdot 4 \cdot 3 - 3 \cdot 4 \cdot 12 + 4 \cdot 3 \cdot 12)$$

$$= (12 + \frac{1}{2} \cdot 1 \cdot 2 + \frac{1}{2} \cdot 2 \cdot 1) \cdot 3 \cdot 4 - (12 + \frac{1}{2} \cdot 1 \cdot 2 + \frac{1}{2} \cdot 2 \cdot 1) \cdot 4 \cdot 3$$

$$- \cdot 3 \cdot 4 \cdot (12 + \frac{1}{2} \cdot 1 \cdot 2 + \frac{1}{2} \cdot 2 \cdot 1) + \cdot 4 \cdot 3 \cdot (12 + \frac{1}{2} \cdot 1 \cdot 2 + \frac{1}{2} \cdot 2 \cdot 1)$$

$$= \mathfrak{l}_{w} + \frac{1}{2} (\mathfrak{l}_{12 \cdot 3 \cdot 4} + \mathfrak{l}_{12 \cdot 3 \cdot 4}).$$

A general equation was, however, not found.

5.3. **Dimension results.** The following is a direct corollary from Conjecture 5.1.

Corollary 5.9. For any $d, h \ge 1$, we have dim $\mathcal{V}_{d,h} = \sum_{l \le h} \lambda_{d,l}$.

Proof. We follow the strategy laid out in [AFS19, Section 6]. Specifically, we show that the projection on $T^h(\mathfrak{A}_d)$, denoted by π^h is a generically *h*-to-1 map on $\hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)$, showing that

(18)
$$\dim \pi^h(\hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)) = \dim \hat{\mathcal{G}}^h(\mathfrak{A}_d).$$

To show that π^h is generically *h*-to-1, let $\mathbf{v} \in \hat{\mathcal{G}}^h(\mathfrak{A}_d)$ such that $\langle \mathbf{v}, \mathbf{1}^{\underline{\mathbf{u}}h} \rangle \neq 0$, and write

$$\mathbf{v} = \sum_{\substack{w \in \mathcal{W}(\mathrm{MS}_d) \\ ||w||_{\mathrm{ht}} = h}} \alpha_w w \, .$$

If $\mathbf{v} = \pi^h(\mathbf{w})$ for $\mathbf{w} \in \hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d)$, then

$$\langle \mathsf{w}, \mathsf{1}
angle^h = \langle \mathsf{w}, \mathsf{1}^{\overline{\mathsf{u}}h}
angle = \langle \mathsf{v}, \mathsf{1}^{\overline{\mathsf{u}}h}
angle$$

because all elements in $\mathbf{1}^{\overline{\mathbf{u}}h}$ have heigh h. This equation determines a non-zero value for $\langle \mathbf{w}, \mathbf{1} \rangle$ up to an h-root of unity of 1.

Now for any w of height l < h, note that

$$\langle \mathsf{v}, w \square \mathbf{1}^{\square h-l} \rangle = \langle \mathsf{w}, w \square \mathbf{1}^{\square h-l} \rangle = \langle \mathsf{w}, w \rangle \langle \mathsf{w}, \mathbf{1} \rangle^{h-l}$$

This determines

$$\langle \mathsf{w}, w \rangle = \langle \mathsf{v}, w \overline{\mathbf{u}} \mathbf{1}^{\overline{\mathbf{u}}h-l} \rangle /_{\langle \mathsf{w}, \mathbf{1} \rangle^{h-l}}$$

We conclude that for \mathbf{v} in an open set of $\hat{\mathcal{G}}^h(\mathfrak{A}_d)$, there are h many values of \mathbf{w} that map to \mathbf{v} . This concludes the proof of (18).

Now in [MR89] it was shown that for any vector space V, there is a basis of $\mathcal{L}^{\leq h}(V)$ given by Lyndon words, of height at most h, on the basis of V. Furthermore, it can be seen from the definition in Eq. (11) that \exp_{\bullet_h} is locally a diffeomorphism, and Proposition 3.6 gives us that Ψ_H^* is an isomorphism of vector spaces. Thus, dimension is preserved, $\dim \hat{\mathcal{G}}^{\leq h}(\mathfrak{A}_d) = \dim \mathcal{L}^{\leq h}(\mathfrak{A}_d) = \sum_{l \leq h} \lambda_{d,l}$. This, together with (18), concludes the proof.

5.4. Enumerative considerations. The dimension of the universal variety dim $\mathcal{V}_{d,h}$ arises as the number of Lyndon words with a specific height. The tilting of the usual grading on word algebras with the introduction of a degree function is uncommon in the study of Lyndon words. For instance, Lyndon words arise in the study of words on the alphabet $\{1, \ldots, d\}$, taken with the degree function constant equal to one. There, words of length h correspond to words of height h. Therefore, the number of Lyndon words of height h in such alphabets is given (see [Reu93, Section 0]) by

$$\mu_{d,h} = \frac{1}{h} \sum_{k|h} \mu\left(\frac{h}{k}\right) d^k \,,$$

where μ is the Möbius function on integers. This reflects the fact that the height and the length play the same role for this alphabet. In the context of discrete signatures, we do not have the luxury of having the same height and length on most words. **Theorem 5.10.** The number of Lyndon words in $\mathcal{W}(MS_d)$ of height h is

$$\lambda_{d,h} = \sum_{k|h} \frac{k}{h} \mu\left(\frac{h}{k}\right) \sum_{\alpha \models k} \frac{1}{\ell(\alpha)} \prod_{i} \binom{\alpha_i + d - 1}{d - 1}.$$

This formula deserves some coments. It is somewhat surprising that this always yields integer values. However, it is a corollary of the proof below that the intermediary terms

$$k\sum_{\alpha\models k}\frac{1}{\ell(\alpha)}\prod_{i}\binom{\alpha_{i}+d-1}{d-1}$$

are also integers.

Proof. First we note that the height grading on T(V) gives us the following power series

$$H(x) = \sum_{w \in \mathcal{W}(\mathrm{MS}_d)} x^{\mathrm{ht}(w)} = \frac{1}{1 - \sum_{I \in \mathrm{MS}_d} x^{\mathrm{deg}(I)}}.$$

Furthermore, for $h \ge 1$ there are $\binom{h+d-1}{d-1}$ multisets on $\{1, \ldots, d\}$ size h, so

$$\sum_{I \in MS_d} x^{\deg(I)} = \sum_{k \ge 1} \binom{k+d-1}{d-1} x^k = (1-x)^{-d} - 1,$$

so we have $H(X) = [1 - ((1 - x)^{-d} - 1)]^{-1}$.

On the other hand, the Lyndon unique factorization theorem (see [CFL58]) guarantees that each word $w \in \mathcal{W}(MS_d)$ can be written uniquely as

$$w=\tau_1\bullet\cdots\bullet\tau_j\,,$$

where τ_i are Lyndon words with $\tau_1 \geq_{lex} \cdots \geq_{lex} \tau_k$. Therefore

$$H(x) = \sum_{w \in \mathcal{W}(\mathrm{MS}_d)} x^{\mathrm{ht}(w)} = \prod_{\substack{\tau \text{ Lyndon word} \\ \tau \in \mathcal{W}(\mathrm{MS}_d)}} \left(1 + x^{\mathrm{ht}(\tau)} + x^{2\mathrm{ht}(\tau)} + x^{3\mathrm{ht}(\tau)} + \cdots\right)$$
$$= \prod_{\substack{\tau \text{ Lyndon word} \\ \tau \in \mathcal{W}(\mathrm{MS}_d)}} (1 - x^{\mathrm{ht}(\tau)})^{-1} = \prod_{k \ge 1} (1 - x^k)^{-\lambda_{d,k}},$$

where we recall that $\lambda_{d,k}$ is the number of Lyndon words $\tau \in \mathcal{W}(MS_d)$ of length k.

Putting it together, applying log on both sides and using

$$-\log(1 - f(x)) = \sum_{n \ge 1} \frac{1}{n} f(x)^n,$$

we get

$$-\log(H(x)) = \sum_{k \ge 1} -\lambda_{d,k} \log(1 - x^k) = \sum_{j,k \ge 1} \frac{1}{j} x^{jk} \lambda_{d,k} =$$
$$= \sum_{n \ge 1} x^n \sum_{k|n} \frac{1}{n/k} \lambda_{d,k} = \sum_{n \ge 1} \frac{1}{n} x^n \sum_{k|n} k \lambda_{d,k}$$

$$-\log(H(x)) = -\log[1 - ((1-x)^{-d} - 1)] =$$
$$= \sum_{k \ge 1} \frac{1}{k} \left(\sum_{t \ge 1} {\binom{t+d-1}{d-1}} x^t \right)^k$$
$$= \sum_{n \ge 0} x^n \sum_{\alpha \models n} \frac{1}{\ell(\alpha)} \prod_i {\binom{\alpha_i+d-1}{d-1}}$$

Equating both sides it follows that for each n we have

$$\sum_{k|n} k\lambda_{d,k} = n \sum_{\alpha \models n} \frac{1}{\ell(\alpha)} \prod_{i} \binom{\alpha_i + d - 1}{d - 1}.$$

Summing both sides for all n divisors of h and multiplying by $\mu\left(\frac{h}{n}\right)$ we get throught Möbius inversion that

$$h\lambda_{d,h} = \sum_{n|h} n \ \mu\left(\frac{h}{n}\right) \sum_{\alpha \models n} \frac{1}{\ell(\alpha)} \prod_{i} \binom{\alpha_i + d - 1}{d - 1},$$

from which the theorem follows.

This allows us to create the following values of $\lambda_{d,h}$. Code created to generate this table can be found in [BP23].

h	1	2	3	4	5	6	7	8	9
d = 1	1	1	2	3	6	9	18	30	56
d = 2	2	4	12	31	92	256	772	2291	7000
d = 3	3	9	36	132	534	2140	8982	38031	164150
d = 4	4	16	80	380	1960	10228	55352	304223	1700712
d = 5	5	25	150	875	5500	35335	234530	1584845	10885640
d = 6	6	36	252	1743	12936	98686	776412	6226008	50732712
d = 7	7	49	392	3136	26852	237160	2158156	20028764	188856934

6. Further work

- Given an element $S \in \hat{\mathcal{G}}(\mathfrak{A}_d)$, can we compute all possible time-series of a fixed size that have S(x) = S? This is related with the degree of the variety $\hat{\mathcal{G}}(\mathfrak{A}_d)$, which is not easy to compute in full generality.
- Does the dimension change when the field characteristic is non-zero?

18

Aknowledgments. The first author is supported in part by the DFG Research Unit FOR2402. The second author is supported by the Max Planck institute for the mathematics in the sciences. Both authors would like to thank fruitful conversations with Sylvie Paycha and Bernd Sturmfels. We would also like to thank Ángel Ríos Ortiz and Pierpaola Santarsiero for all the suggestions and coments on tensor algebras and algebraic varieties.

References

- [AFS19] Carlos Améndola, Peter Friz, and Bernd Sturmfels. Varieties of signature tensors. In *Forum of Mathematics, Sigma*, volume 7, page e10. Cambridge University Press, 2019.
- [BBSK⁺17] Afonso S Bandeira, Ben Blum-Smith, Joe Kileel, Amelia Perry, Jonathan Weed, and Alexander S Wein. Estimation under group actions: recovering orbits from invariants. *arXiv preprint arXiv:1712.10163*, 2017.
- [BFPP22] Carlo Bellingeri, Peter K. Friz, Sylvie Paycha, and Rosa Preiß. Smooth rough paths, their geometry and algebraic renormalization. *Vietnam J. Math.*, 50(3):719–761, 2022.
- [BP20] Jacopo Borga and Raul Penaguiao. The feasible region for consecutive patterns of permutations is a cycle polytope. *Algebraic Combinatorics*, 3(6):1259–1281, 2020.
- [BP23] Carlo Bellingeri and Raul Penaguiao. Varieties of signatures. https://mathrepo.mis.mpg.de, 2023.
- [CFL58] K.-T. Chen, R. H. Fox, and R. C. Lyndon. Free differential calculus. IV. The quotient groups of the lower central series. *Ann. of Math.* (2), 68:81–95, 1958.
- [Che57] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized bakerhausdorff formula. Annals of Mathematics, pages 163–178, 1957.
- [CLO06] David A Cox, John Little, and Donal O'shea. Using algebraic geometry, volume 185. Springer Science & Business Media, 2006.
- [DET20] Joscha Diehl, Kurusch Ebrahimi-Fard, and Nikolas Tapia. Time-warping invariants of multidimensional time series. Acta Appl. Math., 170:265–290, 2020.
- [DSS08] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. Lectures on algebraic statistics, volume 39. Springer Science & Business Media, 2008.
- [FH20] Peter K. Friz and Martin Hairer. A Course on Rough Paths, with introduction to Regularity Structures (2nd extended edition). Springer International Publishing, 2020.
- [FV10] Peter K. Friz and Nicolas Victoir. Multidimensional stochastic processes as rough paths, volume 120 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010. Theory and applications.
- [Haz01] Michiel Hazewinkel. The algebra of quasi-symmetric functions is free over the integers. Advances in Mathematics, 164(2):283–300, 2001.
- [Hof00] Michael E. Hoffman. Quasi-shuffle products. J. Algebraic Combin., 11(1):49–68, 2000.
- [LCL07] Terry J Lyons, Michael Caruana, and Thierry Lévy. Differential equations driven by rough paths. Springer, 2007.
- [MR89] Guy Melançon and Christophe Reutenauer. Lyndon words, free algebras and shuffles. Canadian Journal of Mathematics. Journal Canadian de Mathématiques, 41(4):577–591, 1989.
- [Pen22] Raul Penaguiao. Pattern hopf algebras. Annals of Combinatorics, 26(2):405–451, 2022.
- [PSS19] Max Pfeffer, Anna Seigal, and Bernd Sturmfels. Learning paths from signature tensors. SIAM J. Matrix Anal. Appl., 40(2):394–416, 2019.
- [Reu93] Christophe Reutenauer. Free Lie Algebras. LMS monographs. Clarendon Press, 1993.
- [SR94] Igor Rostislavovich Shafarevich and Miles Reid. *Basic algebraic geometry*, volume 2. Springer, 1994.
- [Var14] Yannic Vargas. Hopf algebra of permutation pattern functions. Discrete Mathematics & Theoretical Computer Science, 2014.