Weierstraß-Institut
für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Preprint ISSN 2198-5855

Generalized iterated-sums signatures

Joscha Diehl[1], Kurusch Ebrahimi-Fard[2], Nikolas Tapia[3]

submitted: December 10, 2020

[1] Universität Greifswald
Institut für Mathematik und Informatik
Walther-Rathenau-Str. 47
17489 Greifswald
Germany
E-Mail: joscha.diehl@gmail.com

[2] Department of Mathematical Sciences
NTNU
7491 Trondheim
Norway
E-Mail: kurusch.ebrahimi-fard@ntnu.no

[3] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
and
Institut für Mathematik
Technische Universität Berlin
Fakultät II
Str. des 17. Juni 136
10587 Berlin
Germany
E-Mail: nikolasesteban.tapiamunoz@wias-berlin.de

No. 2795
Berlin 2020

2010 *Mathematics Subject Classification.* 60L10, 60L70, 16Y60, 93C55.

Key words and phrases. Time series analysis, time warping, signature, quasi-shuffle product, Hoffman's exponential, Hopf algebra.

The second author is supported by the Research Council of Norway through project 302831 "Computational Dynamics and Stochastics on Manifolds" (CODYSMA). The third author is supported the DFG MATH+ Excellence Cluster.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Leibniz-Institut im Forschungsverbund Berlin e. V.
Mohrenstraße 39
10117 Berlin
Germany

Fax: +49 30 20372-303
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Generalized iterated-sums signatures

Joscha Diehl, Kurusch Ebrahimi-Fard, Nikolas Tapia

ABSTRACT. We explore the algebraic properties of a generalized version of the iterated-sums signature, inspired by previous work of F. Király and H. Oberhauser. In particular, we show how to recover the character property of the associated linear map over the tensor algebra by considering a deformed quasi-shuffle product of words on the latter. We introduce three non-linear transformations on iterated-sums signatures, close in spirit to Machine Learning applications, and show some of their properties.

1. INTRODUCTION

Recently, a series of papers [3, 4, 11, 13] have highlighted the importance of *signature-like objects* (following T. Lyons' nomenclature) for capturing features of sequentially ordered data. Part of the reason for their success is that these transformations posses a universality property, meaning that they are able to approximate arbitrary (bounded) nonlinear mappings on sequence space by linear functionals on feature space. They can also be efficiently computed thanks to an inherent recursive structure.

Both properties can be succinctly described by using Hopf-algebraic language, which has by now become standard in the field. For the iterated-integrals signature, the underlying Hopf algebra is the space of words together with the commutative *shuffle product* [16, 17] and the noncocommutative deconcatenation coproduct, whereas for the iterated-sums signature it turns out to be the commutative *quasi-shuffle algebra* over words [1, 8, 9, 14] with the same coproduct. Shuffle and quasi-shuffle products algebraically encode integration respectively summation by parts for iterated integrals respectively sums. In both cases the properties mentioned in the preceding paragraph amount to saying that the corresponding signature-like maps are characters, i.e., algebra morphisms, and that they satisfy Chen's relation.

Recently, F. Király and H. Oberhauser [11] introduced a higher order version of the iterated-sums signature as a way of approximating the iterated-integrals signature. The main disadvantage of this generalization is that the character quality is lost and consequently the universality property ceases to hold. In the paper at hand we unfold the algebraic underpinning of Király's and Oberhauser's definition of the higher order iterated-sums signature, which permits us to further generalize it to arbitrary nonlinearities, as opposed to the more restricted class of exponential-type nonlinearities underlying previous approaches. We show that this generalization enjoys a character property but with respect to a different Hopf algebra defined on words in terms of a modified quasi-shuffle product and the deconcatenation coproduct; in fact, we show that the algebraic structure actually depends on the selected nonlinear transformation.

Thanks to the more general approach, we are able to introduce three new transformations of the iterated-sums signature. The first transformation is obtained by applying a tensorized nonlinear transformation to each time slice, the second one is constructed by applying a polynomial map to increments, whereas the third is obtained by first transforming the data and then considering its increments. We show that these transformations can be expressed in terms of the un-transformed iterated-sums signature. In the third case, this rewriting procedure generalizes earlier work by L. Colmenarejo and R. Preiß on iterated-integrals signatures defined with respect to paths transformed by polynomial maps [2].

The rest of the paper is organized as follows. In Section 2 we review the algebraic underpinning of our construction, i.e., the notion of quasi-shuffle Hopf algebra. In Section 3 we introduce the generalized iterated-sums signature map and provide a complete description of its most important properties using the developments of the previous section.

DOI 10.20347/WIAS.PREPRINT.2795 Berlin 2020

2. Quasi-shuffle Hopf algebra

In this section we briefly recall the notion of commutative quasi-shuffle product, the algebraic framework present in [3]. However, we shall emphasise the refined viewpoint based on the notion of half-shuffle product. Readers interested in the details are directed to further references [5–7, 9, 10].

Following Foissy and Patras [6] we define the notion of commutative quasi-shuffle algebra over the base field $\mathbb{R}$.

Definition 2.1. A commutative quasi-shuffle algebra $(A, \succ, \bullet)$ consists of a nonunital commutative $\mathbb{R}$-algebra $(A, \bullet)$ equipped with a linear *half-shuffle product* $\succ\colon A \otimes A \to A$ satisfying

$$x \succ (y \succ z) = (x * y) \succ z \tag{1}$$

$$(x \succ y) \bullet z = x \succ (y \bullet z), \tag{2}$$

where the quasi-shuffle product is defined as

$$x * y := x \succ y + y \succ x + x \bullet y. \tag{3}$$

One sees that (1) and (2) imply that the quasi-shuffle product (3) is both commutative and associative. Observe that one may characterise a commutative quasi-shuffle algebra as a space with two commutative products related through a –symmetrized– half-shuffle product. In the following quasi-shuffle algebra means commutative quasi-shuffle algebra.

Definition 2.2. A quasi-shuffle morphism between two quasi-shuffle algebras $(A, \succ, \bullet)$ and $(\tilde{A}, \tilde{\succ}, \tilde{\bullet})$ is a linear map $\Lambda\colon A \to \tilde{A}$ such that

$$\Lambda(x \succ y) = \Lambda(x) \mathrel{\tilde{\succ}} \Lambda(y), \quad \Lambda(x \bullet y) = \Lambda(x) \mathbin{\tilde{\bullet}} \Lambda(y)$$

for all $x, y \in A$.

Any quasi-shuffle morphism is an algebra morphism between quasi-shuffle algebras, that is, $\Lambda(x * y) = \Lambda(x) \mathbin{\tilde{*}} \Lambda(y)$. A quasi-shuffle algebra $(A, \succ, \bullet)$ has a unital extension. Indeed, set $\bar{A} := \mathbb{R}1 \oplus A$ and define for $a \in A$: $1 \bullet x = x \bullet 1 := 0$, $1 \succ x := 0$ and $x \succ 1 := x$. Note that the product $1 \succ 1$ as well as $1 \bullet 1$ are excluded and that $1 * 1 := 1$. This turns $\bar{A}$ into a unital algebra.

It is noted that if the commutative algebra $(A, \bullet)$ in Definition 2.1 has a trivial product, then the notion of commutative quasi-shuffle algebra reduces to that of commutative shuffle algebra, which is defined solely in terms of relation (1). In this case the commutative and associative product (3) is called shuffle product.

Remark 2.3. We note that different terminology is used in the literature. Commutative shuffle and quasi-shuffle algebras are also known as Zinbiel and commutative tridendriform algebras, respectively. The noncommutative generalisations of both shuffle and quasi-shuffle algebra, are also known as dendriform and tridendriform algebras, respectively. In this work we follow [6], where the preference for the terminology used here is explained.

Our main example provides also the paradigm of commutative quasi-shuffle algebra, i.e., the free commutative quasi-shuffle algebra. Let $A = \{1, \dots, d\}$ be a finite alphabet and consider the reduced symmetric algebra $S(A)$ over the vector space spanned by it. In other words, $S(A)$ is the space spanned by words in commuting letters from A; here *reduced* means that we do not consider $S(A)$ to have a unit, i.e. we consider only the augmentation ideal of the standard symmetric algebra over A. Keeping up with our previous convention, we denote the commutative product in $S(A)$ by square brackets. We also do not endow $S(A)$ with any additional algebraic structure other than its product. Finally we recall that $S(A)$ has a natural grading

$$S(A) = \bigoplus_{n=1}^{\infty} S^n A$$

where $S^n A$ is spanned by products of the form $[\mathtt{i}_1 \cdots \mathtt{i}_n]$ with $\mathtt{i}_1, \ldots, \mathtt{i}_n \in A$. We denote this basis by $\mathfrak{A}_n$. It is well known that $\dim S^n A = \binom{d+n-1}{n}$. Furthermore, the set

$$\mathfrak{A} = \bigcup_{n=1}^{\infty} \mathfrak{A}_n$$

constitutes a basis for $S(A)$.

Now, we let $H := \overline{T}(S(A))$ be the tensor algebra over $S(A)$. As a vector space, it is the direct sum

$$H = \bigoplus_{n=0}^{\infty} S(A)^{\otimes n} = \bigoplus_{n=0}^{\infty} H_n$$

where $H_0 = \mathbb{R}\mathbf{e}$ and

$$H_n = \bigoplus_{k=1}^{n} \left(\bigoplus_{i_1+\cdots+i_k=n} S^{i_1} A \otimes \cdots \otimes S^{i_k} A \right).$$

In the following we will use the word notation for elements in H. Concatenation, written by juxtaposition of symbols, is the standard product on H. In particular, H inherits a grading from $S(A)$ in this way, which we call the *weight*, and denote it by $|\cdot|$. The length of a word $w = s_1 \cdots s_k \in H$ is defined to be $\ell(w) = k$. In [3] we show that

$$\sum_{n=0}^{\infty} t^n \dim H_n = \frac{(1-t)^d}{2(1-t)^d - 1} = 1 + dt + \frac{d(3d+1)}{2} t^2 + \frac{d(13d^2+9d+2)}{6} t^3 + \cdots .$$

The standard basis of H is the set of words over $\mathfrak{A}$, here denoted by $\mathfrak{A}^*$. We endow H with an inductively defined product obtained from the bracket product of $S(A)$: $\mathbf{e} \star u := u =: u \star \mathbf{e}$ and

$$ua \star vb = (u \star vb)a + (ua \star v)b + (u \star v)[ab] \tag{4}$$

for $u, v \in H$ and $a, b \in S(A)$. Hoffman [9] called (4) quasi-shuffle product and showed that it is commutative and associative as well as compatible with the deconcatenation coproduct Δ and the counit ε determined by the grading, so that $H_{\text{qsh}} := (H, \star, \Delta, \mathbf{e}, \varepsilon, \alpha)$ is a Hopf algebra. See also [14] and [8].

The quasi-shuffle algebra H_{qsh} carries a commutative quasi-shuffle structure [12], defined recursively by

$$u \succ va := (u \star v)a, \quad ua \bullet vb = (u \star v)[ab]. \tag{5}$$

Observe that any word $w = s_1 \cdots s_k \in H$ can be written as

$$w = (\cdots((s_1 \succ s_2) \succ s_3) \cdots) \succ s_k.$$

Remark 2.4. It is natural to consider the relation between the deconcatenation coproduct Δ and half-shuffle as well as the $\bullet$ products. It turns that they form together a quasi-shuffle bialgebra. See [6] for more details.

Finally, we recall the following important result due to Loday [12, Theorem 2.5].

Theorem 2.5. *The free commutative unital quasi-shuffle algebra over $\mathbb{R}^d$ is isomorphic to H_{qsh}.*

The dual space H^* can be identified with formal word series of the form

$$\mathsf{S} = \sum_{w \in \mathfrak{A}^*} \langle \mathsf{S}, w \rangle w$$

in the sense that there is an isomorphism between such formal series and elements of H^*. The convolution product of two maps $\mathsf{R}, \mathsf{S} \in H^*$ is defined by

$$\mathsf{RS} := \sum_w \langle \mathsf{R} \otimes \mathsf{S}, \Delta w \rangle w = \sum_w \left(\sum_{uv=w} \langle \mathsf{R}, u \rangle \langle \mathsf{S}, v \rangle \right) w.$$

Observe that this product is associative but not commutative.

2.1. **Deformed quasi-shuffle products.** We recall the following set-up from [9, 10]. A composition of an integer $n \geq 1$ is a sequence $I = (i_1, \ldots, i_k)$ of positive integers such that $i_1 + \cdots + i_k = n$. We write $C(n)$ for the set of compositions of n. For any word $w = s_1 \cdots s_n \in H$ and composition $I = (i_1, \ldots, i_k) \in C(n)$ we define

$$I[w] := [s_1 \cdots s_{i_1}][s_{i_1+1} \cdots s_{i_1+i_2}] \cdots [s_{i_1+\cdots+i_{k-1}+1} \cdots s_n].$$

Formal diffeomorphisms $f \in t\mathbb{R}[[t]]$ induce linear automorphisms of H in the following way: suppose that

$$f(t) = \sum_{n=1}^{\infty} c_n t^n,$$

and define

$$\Psi_f(w) := \sum_{I \in C(\ell(w))} c_{i_1} \cdots c_{i_k} I[w].$$

Formal diffeomorphisms are invertible with respect to composition of formal power series, and it can be shown that $\Psi_{f^{-1}} = \Psi_f^{-1}$ [10]. Finally, we observe that Ψ_f is always a coalgebra morphism [7]; this means that the identity

$$\Delta \circ \Psi_f = (\Psi_f \otimes \Psi_f) \circ \Delta$$

holds. Moreover, by definition the map Ψ_f is graded, that is, $\Psi_f(H_n) \subset H_n$ for all $n \geq 0$. These maps can be used to define a *deformed quasi-shuffle algebra* with deformed half-shuffle.

Proposition 2.6. *H equipped with the deformed products*

$$u \succ_f v := \Psi_f^{-1}(\Psi_f(u) \succ \Psi_f(v)) \qquad u \bullet_f v := \Psi_f^{-1}(\Psi_f(u) \bullet \Psi_f(v))$$

is a commutative quasi-shuffle algebra. The deformed quasi-shuffle product

$$u \star_f v := \Psi_f^{-1}(\Psi_f(u) \star \Psi_f(v)). \tag{6}$$

is associative and commutative.

Proof. Commutativity and associativity of $\bullet_f$ are evident, since the $\bullet$ is itself commutative and associative. The half-shuffle identities (1) and (2) are verified directly.

$$\begin{aligned} u \succ_f (v \succ_f w) &= \Psi_f^{-1}\Big(\Psi_f(u) \succ (\Psi_f(v) \succ \Psi_f(w))\Big) \\ &= \Psi_f^{-1}\Big((\Psi_f(u) \star \Psi_f(v)) \succ \Psi_f(w)\Big) \\ &= (u \star_f v) \succ_f w \end{aligned}$$

The second half-shuffle identity follows analogously

$$\begin{aligned}(u \succ_f v) \bullet_f w &= (\Psi_f(u) \succ \Psi_f(v)) \bullet \Psi_f(w)\\ &= \Psi_f(u) \succ (\Psi_f(v) \bullet \Psi_f(w))\\ &= u \succ_f (v \bullet_f w).\end{aligned}$$

Commutativity and associativity of $\star_f$ follow immediately. □

Even more, we have the following.

Theorem 2.7. *The space $H_f := (H, \star_f, \Delta, \mathbf{e}, \varepsilon)$ is a connected graded Hopf algebra.*

Proof. We have already seen that $(H, \star_f, \mathbf{e})$ is a commutative algebra. Since the coproduct is the (unchanged) deconcatenation coproduct of H_{qsh}, (H, Δ, ε) is a coalgebra. Clearly, the relation $\varepsilon \circ \Psi_f = \Psi_f \circ \varepsilon$ holds, and since $\Psi_f(\mathbf{e}) = \Psi_f^{-1}(\mathbf{e}) = \mathbf{e}$ we see that

$$\varepsilon(u \star_f v) = \varepsilon(\Psi_f(u))\varepsilon(\Psi_f(v)) = \varepsilon(u)\varepsilon(v).$$

Therefore, ε is an algebra morphism with respect to the deformed quasi-shuffle product $\star_f$.

Finally, we check that Δ is an algebra morphism as well. Using the notation $m_f(u \otimes v) := u \star_f v$, we have

$$\begin{aligned}\Delta \circ m_f &= \Delta \circ \Psi_f^{-1} \circ m \circ (\Psi_f \otimes \Psi_f)\\ &= (\Psi_f^{-1} \otimes \Psi_f^{-1}) \circ \Delta \circ m \circ (\Psi_f \otimes \Psi_f)\\ &= (\Psi_f^{-1} \otimes \Psi_f^{-1}) \circ \tilde{m} \circ (\Delta \otimes \Delta) \circ (\Psi_f \otimes \Psi_f)\\ &= (\Psi_f^{-1} \otimes \Psi_f^{-1}) \circ \tilde{m} \circ (\Psi_f \otimes \Psi_f \otimes \Psi_f \otimes \Psi_f) \circ (\Delta \otimes \Delta)\\ &= \tilde{m}_f \circ (\Delta \otimes \Delta).\end{aligned}$$

Here, we have defined $\tilde{m}\colon H^{\otimes 4} \to H \otimes H$ by $m := (m \otimes m) \circ (\mathrm{id} \otimes \tau \otimes \mathrm{id})$ and $\tau\colon H \otimes H \to H \otimes H$, $\tau(x \otimes y) := y \otimes x$ is the standard *braiding isomorphism*. □

Finally, we can show

Proposition 2.8. *The map $\Psi_f\colon H_f \to H_{\mathrm{qsh}}$ is a Hopf algebra isomorphism.*

Proof. In general, $\Psi_f\colon (H, \Delta) \to (H, \Delta)$ is a coalgebra morphism [7, 10]. It is an algebra morphism by definition. Indeed,

$$\Psi_f(u \star_f v) = \Psi_f\Big(\Psi_f^{-1}(\Psi_f(u) \star \Psi_f(v))\Big) = \Psi_f(u) \star \Psi_f(v).$$

Therefore, $\Psi_f\colon H_f \to H_{\mathrm{qsh}}$ is an isomorphism of bialgebras. It is a general fact that this implies that Ψ_f is already a Hopf algebra isomorphism. □

Remark 2.9. It is possible to show [7, Theorem 2.2] that in fact, all possible coalgebra automorphisms of (H, Δ) are of the form Ψ_f for some $f \in t\mathbb{R}[[t]]$. In particular, this means that any Hopf algebra preserving the deconcatenation coproduct has to be of the form H_f.

3. Iterated-sums signatures

We start this section by recalling the definition of the iterated-sums signature introduced in [3]. Fix integers $d \geq 1$ and $N > 0$. A *d-dimensional time series of length N* is a sequence of vectors

$x = (x_0, \ldots, x_{N-1}) \in (\mathbb{R}^d)^N$. The following notation for elements in the time series x is put in place:

$$x_j = (x_j^{(1)}, \ldots, x_j^{(d)}),$$

and it is extended to include brackets in $\mathfrak{A}$ by defining

$$x_j^{[\mathtt{i}_1 \cdots \mathtt{i}_n]} := x_j^{(i_1)} \cdots x_j^{(i_n)}. \tag{7}$$

Given a d-dimensional time series of length N, its increment series, denoted by δx, is also a d-dimensional time series of length $N-1$ with entries defined by $\delta x_k := x_{k+1} - x_k$.

Definition 3.1. Let x be a d-dimensional time series, and denote by δx its increment series. The *iterated-sums signature* of x is the two-parameter family $(\mathrm{ISS}(x)_{n,m} : 0 \le n \le m \le N)$ of linear maps in H^*_{qsh} such that $\mathrm{ISS}(x)_{n,n} = \varepsilon$, and defined recursively by $\langle \mathrm{ISS}(x)_{n,m}, \mathbf{e} \rangle := 1$, and

$$\langle \mathrm{ISS}(x)_{n,m}, a_1 \cdots a_p \rangle := \sum_{j=n}^{m-1} \langle \mathrm{ISS}(x)_{n,j}, a_1 \cdots a_{p-1} \rangle \delta x_j^{a_p}$$

for all words $a_1 \cdots a_p \in \mathfrak{A}^*$.

We recall that as a formal word series, the map $\mathrm{ISS}(x)_{n,m}$ can be expressed as the time-ordered product

$$\mathrm{ISS}(x)_{n,m} = \overrightarrow{\prod_{n \le j < m}} \left(\varepsilon + \sum_{a \in \mathfrak{A}} \delta x_j^a \, a \right). \tag{8}$$

In fact, eq. (8) can be seen to arise as the solution to a fixed-point equation in H_{qsh}. Indeed, from Definition 3.1 we see that for any word $a_1 \cdots a_p \in \mathfrak{A}^*$,

$$\begin{aligned} \delta \langle \mathrm{ISS}(x)_{n,\cdot}, a_1 \cdots a_p \rangle_m &= \langle \mathrm{ISS}(x)_{n,m+1}, a_1 \cdots a_p \rangle - \langle \mathrm{ISS}(x)_{n,m}, a_1 \cdots a_p \rangle \\ &= \langle \mathrm{ISS}(x)_{n,m}, a_1 \cdots a_{p-1} \rangle \delta x_m^{a_p}. \end{aligned}$$

Therefore, the equality between word series

$$\delta\, \mathrm{ISS}(x)_{n,m} = \mathrm{ISS}(x)_{n,m} \Phi(\delta x_m), \quad \mathrm{ISS}(x)_{n,n} = \varepsilon$$

holds, where the "polynomial extension" map $\Phi \colon \mathbb{R}^d \to S(\!(A)\!)$ (c.f. [15, eq. 60]) is defined by

$$\Phi(z) = \sum_{n=1}^{\infty} \left(\sum_{\mathtt{i} \in A} z^{(i)} [\mathtt{i}] \right)^n = \sum_{a \in \mathfrak{A}} z^a a.$$

More concisely, Φ amounts to a geometric series in the completed symmetric algebra $S(\!(A)\!)$ – recall from Section 2 that $[\cdot\cdot]$ denotes the symmetric tensor product in $S(A)$. This extension has also been considered by Toth, Bonnier and Oberhauser [18] in a Machine Learning context.

We now record the two most relevant properties of the iterated-sums signature of a d-dimensional time series, shown in [3, Theorem 3.4].

Theorem 3.2.

1 *For each $0 \le n \le m \le N$, the map $\mathrm{ISS}(x)_{n,m}$ is a quasi-shuffle character.*
2 *For any $0 \le n < n' < n'' \le N$ we have*

$$\mathrm{ISS}(x)_{n,n'}\, \mathrm{ISS}(x)_{n',n''} = \mathrm{ISS}(x)_{n,n''}.$$

In fact, in light of the commutative quasi-shuffle structure of H_{qsh}, one can be more precise about the nature of point (1) in Theorem 3.2. Let $\mathfrak{X}_N$ denote the space of real-valued time series with fixed time horizon $N \in \mathbb{N}$. It carries a commutative quasi-shuffle structure, given by

$$(x \succeq y)_k := \sum_{j=0}^{k-1}(x_j - x_0)\delta y_j, \quad (x \odot y)_k := \sum_{j=0}^{k-1} \delta x_j \delta y_j.$$

The corresponding associative product is $(x, y) \mapsto (x_\cdot - x_0)(y_\cdot - y_0)$ in $\mathfrak{X}_N$. For a given d-dimensional time series, we define a map $\sigma(x)\colon A \to \mathfrak{X}_N$ by

$$\langle\sigma(x), [\mathtt{i}]\rangle_k := x_k^{(i)} - x_0^{(i)}, \quad 0 \le k \le N.$$

By Theorem 2.5, it admits a unique extension to H_{qsh} as a commutative quasi-shuffle morphism.

Proposition 3.3. *The unique extension $\sigma\colon H_{\mathrm{qsh}} \to \mathfrak{X}_N$ is such that for all $0 \le k \le N$ and words $w \in \mathfrak{A}^*$ we have*

$$\langle\sigma(x), w\rangle_k = \langle\mathrm{ISS}(x)_{0,k}, w\rangle.$$

Proof. We first observe that since $[\mathtt{i}_1] \bullet \cdots \bullet [\mathtt{i}_n] = [\mathtt{i}_1 \cdots \mathtt{i}_n]$ for all $\mathtt{i}_1, \dots, \mathtt{i}_n \in A$, we have

$$\begin{aligned}\langle\sigma(x), [\mathtt{i}_1 \cdots \mathtt{i}_n]\rangle_k &= (\langle\sigma(x), \mathtt{i}_1\rangle \odot \cdots \odot \langle\sigma(x), \mathtt{i}_n\rangle)_k \\ &= \sum_{j=0}^{k-1} \delta x_j^{(i_1)} \cdots \delta x_j^{(i_n)} \\ &= \langle\mathrm{ISS}(x)_{0,k}, [\mathtt{i}_1 \cdots \mathtt{i}_n]\rangle\end{aligned}$$

by eq. (7). This shows the identity for all words of length 1. Now, suppose the equality is proven for all words up to length p. Any word w of length $p+1$ can be decomposed as $w = ua$ for some $u \in \mathfrak{A}^*$ with $\ell(u) = p$ and $a \in \mathfrak{A}$. Since, from eq. (5) (set $v = \mathbf{e}$), $u \succ a = ua$ for any $u \in \mathfrak{A}^*$ and $a \in \mathfrak{A}$, we see that

$$\begin{aligned}\langle\sigma(x), ua\rangle_k &= \langle\sigma(x), u \succ a\rangle_k \\ &= (\langle\sigma(x), u\rangle \succeq \langle\sigma(x), a\rangle)_k \\ &= \sum_{j=0}^{k-1}\langle\mathrm{ISS}(x)_{0,j}, u\rangle \delta x_j^a \\ &= \langle\mathrm{ISS}(x)_{0,k}, ua\rangle\end{aligned}$$

by Definition 3.1. □

To construct the maps $k \mapsto \mathrm{ISS}(x)_{n,k}$ for $0 \le n \le N$, one can follow a similar route, by first considering the map $x \mapsto \tilde{x} = (\tilde{x}_k : n \le k \le N)$ with $\tilde{x}_k = x_k - x_n$. The image of $\mathfrak{X}_N$ under this map will be denoted by $\tilde{\mathfrak{X}}_{n,N}$. The quasi-shuffle structure defined above can be transported to $\tilde{\mathfrak{X}}_{n,N}$ via this map, i.e., $\tilde{x} \succeq \tilde{y} := \widetilde{x \succeq y}$, $\tilde{x} \odot \tilde{y} := \widetilde{x \odot y}$. It is not difficult to see that then the same procedure applied now to $\tilde{\mathfrak{X}}_{n,N}$ gives rise to $\mathrm{ISS}(x)_{n,k}$ for $0 \le n \le k < N$. In particular we obtain the following

Proposition 3.4. *Let x be d-dimensional time series and fix $0 \le n \le m \le N$. The identities*

$$\langle\mathrm{ISS}(x)_{n,m}, u \succ v\rangle = \sum_{k=n}^{m-1}\langle\mathrm{ISS}(x)_{n,k}, u\rangle\delta\langle\mathrm{ISS}(x)_{n,\cdot}, v\rangle_k$$

$$\langle\mathrm{ISS}(x)_{n,m}, u \bullet v\rangle = \sum_{n \le k < m} \delta\langle\mathrm{ISS}(x)_{n,\cdot}, u\rangle_k \delta\langle\mathrm{ISS}(x)_{n,\cdot}, v\rangle_k$$

hold for all $u, v \in H_{\mathrm{qsh}}$.

3.1. **Generalized iterated-sums signatures.** Let $f \in t\mathbb{R}[[t]]$ be a formal diffeomorphism. Apart from the map Ψ_f described in the previous section, it induces a transformation on formal word series by

$$f(\mathrm{S}) = \sum_{n=1}^{\infty} c_n \mathrm{S}^n.$$

Definition 3.5. Let x be a d-dimensional time series and $f \in t\mathbb{R}[[t]]$ a formal diffeomorphism. The *generalized iterated-sums signature* is the family of linear maps $(\mathrm{ISS}^f(x)_{n,m} : 0 \le n \le m \le N)$ defined by

$$\mathrm{ISS}^f(x)_{n,m} := \prod_{n\le j<m}^{\rightarrow} \left(\varepsilon + f\left(\sum_{a\in\mathfrak{A}} \delta x_j^a a\right)\right).$$

We immediately have

Proposition 3.6. *The generalized iterated-sums signature satisfies Chen's property, that is, for any* $0 \le n \le n' \le n'' \le N$

$$\mathrm{ISS}^f(x)_{n,n'}\, \mathrm{ISS}^f(x)_{n',n''} = \mathrm{ISS}^f(x)_{n,n''}.$$

We observe that due to the nonlinear nature of the transformation f applied inside the product, expansion of this expression as a proper word series is, in principle, not straightforward. However we have the following result.

Proposition 3.7. *For every* $w \in H$*,*

$$\langle \mathrm{ISS}(x)^f_{n,m}, w\rangle = \langle \mathrm{ISS}(x)_{n,m}, \Psi_f(w)\rangle.$$

Proof. First we observe that, by Chen's property and the fact that Ψ_f is a coalgebra morphism, it suffices to show that the equality holds when $m = n+1$, i.e., we only need to show that

$$\left\langle f\left(\sum_{a\in\mathfrak{A}} \delta x_n^a a\right), w\right\rangle = \sum_{a\in\mathfrak{A}} \delta x_n^a \langle a, \Psi_f(w)\rangle.$$

Moreover, since the identity is linear in w, we can further restrict ourselves to the case $w \in \mathfrak{A}^*$.

Now, by definition,

$$\begin{aligned} f\left(\sum_{a\in\mathfrak{A}} \delta x_n^a a\right) &= \sum_{m=1}^{\infty} c_m \left(\sum_{a\in\mathfrak{A}} \delta x_n^a a\right)^m \\ &= \sum_{m=1}^{\infty} c_m \sum_{a_1,\dots,a_m\in\mathfrak{A}} \delta x_n^{a_1}\cdots\delta x_n^{a_m} a_1\cdots a_m. \end{aligned}$$

Since $\delta x_n^{a_1}\cdots\delta x_n^{a_m} = \delta x_n^{[a_1\cdots a_m]}$ we obtain that, if $w = a_1\cdots a_m \in \mathfrak{A}^*$,

$$\left\langle f\left(\sum_{a\in\mathfrak{A}} \delta x_n^a a\right), w\right\rangle = c_m \delta x_n^{[a_1\cdots a_m]}.$$

On the other hand, we have

$$\begin{aligned} \left\langle \sum_{a\in\mathfrak{A}} \delta x_n^a a, \Psi_f(w)\right\rangle &= \sum_{a\in\mathfrak{A}} \delta x_n^a \langle a, \Psi_f(w)\rangle \\ &= \sum_{a\in\mathfrak{A}} \delta x_n^a \sum_{J\in C(m)} c_{i_1}\cdots c_{i_k} \langle a, I[w]\rangle. \end{aligned}$$

However, in the last sum the only word of length 1 of the form $I[w]$ is $(m)[w] = [a_1 \cdots a_m]$. Therefore

$$\left\langle \sum_{a\in\mathfrak{A}} \delta x_n^a a, \Psi_f(w) \right\rangle = c_m \delta x_n^{[a_1 \dots a_m]}$$

and the equality is proven in this case. □

An application of Proposition 2.8 yields

Corollary 3.8. *The generalized iterated-sums signature is a character over the Hopf algebra H_f with deformed quasi-shuffle product.*

Remark 3.9. We observe that as a consequence of Remark 2.9, an even stronger statement is true. Suppose that $\Psi\colon H \longrightarrow H$ is a transformation, and define $\overline{\mathrm{ISS}}(x) = \mathrm{ISS}(x) \circ \Psi$. This map satisfies Chen's relation if (and only if) Ψ is a coalgebra morphism, thus it must be of the form $\Psi = \Psi_f$ for some $f \in t\mathbb{R}[[t]]$. Moreover, $f = \pi \circ \Psi$ where $\pi\colon H \longrightarrow S(A)$ is the canonical projection on words of length one.

Before concluding this part, we relate our results to the higher-order discrete signatures introduced by F. Király and H. Oberhauser [11, Definition B.4]. Given an integer $p \geq 1$, these authors define for a d-dimensional time series x the map $\mathrm{S}^+_{(p)}(x) \in T((\mathbb{R}^d))$ by

$$\mathrm{S}^+_{(p)}(x) = \prod_{0\le i<N-1}^{\rightarrow} \sum_{j=0}^{p} \frac{(\delta x_i)^{\otimes j}}{j!}. \tag{9}$$

Here $\delta x_i = \sum_{\mathtt{j}\in A} \delta x_i^{(j)}[\mathtt{j}]$ and $(\delta x_i)^{\otimes n} = \sum_{\mathtt{j}_1,\dots,\mathtt{j}_n\in A} \delta x_i^{(j_1)} \cdots \delta x_i^{(j_n)} [\mathtt{j}_1] \cdots [\mathtt{j}_n]$. We note that this map, considered as a linear map on $T(\mathbb{R}^d)$, is not an algebra morphism over *any* product defined on $T(\mathbb{R}^d)$ that is compatible with the grading. Indeed, suppose there is such a product and denote it by $\circledast$. Consider the map $\mathrm{S}^+_{(p)}(x)$ over a single time step with a non-zero increment. Fix moreover a single symbol, say $\mathtt{1} \in A$. Then

$$(\delta x_0^{(1)})^{p+1} = \langle \mathtt{1}^{\circledast(p+1)}, \mathrm{S}^+_{(p)}(x) \rangle = 0$$

which is a contradiction. This is resolved by considering the infinite-dimensional polynomial extension of the time series, including all powers of increments [18, Example B.2]. In essence, this is what the quasi-shuffle approach does – the extension is obtained by considering the bracket terms $[\mathtt{i}_1 \cdots \mathtt{i}_n] \in \mathfrak{A}$ and the corresponding extended increments $\delta x^{[\mathtt{i}_1 \cdots \mathtt{i}_n]}$.

However, even when considering the proper extension, the map so obtained does not yield a character over the quasi-shuffle Hopf algebra when $p > 1$ (the case $p = 1$ corresponds to $\mathrm{ISS}(x)$, see eq. (8)). Taking $p = 2$, a single time step and considering the product $\mathtt{1} \star \mathtt{1}$ constitutes a simple example.[1] In this case, $\mathrm{S}^+_{(p)}(x)$ equals $\mathrm{ISS}^{f_p}(x)$, with $f_p = t + \frac{1}{2}t^2 + \cdots + \frac{1}{p!}t^p$; Corollary 3.8 restores the character property of this map, with respect to a different product.

Finally, we mention that in the limit $p \longrightarrow \infty$, $\mathrm{S}^+_{(\infty)}(x)$ coincides with the iterated-integrals signature of the path X interpolating the values of x piecewise linearly with unit speed. In the same way, the extended version $\mathrm{ISS}^{f_\infty}(x)$ coincides with the iterated-integrals signature of an infinite-dimensional extension of X [3, Theorem 5.3]. Both statements are consistent with the fact that $f_\infty(x) = \exp(t) - 1$, so that Ψ_{f_∞} is the Hoffman exponential and $\star_{f_\infty}$ becomes the shuffle product (over A and $\mathfrak{A}$, respectively).

[1] The reader is invited to work out the details.

3.2. **Vectorized transformation.** In some applications, one might have only access to observables of a time series and not to the time series itself. In others, such as in Machine Learning, applying non-linearities to the data might be of use. We introduce now an analogue of the iterated-sums signature, acting on transformed data.

Let $f \in t\mathbb{R}[t]$ be a polynomial with $f(0) = 0$, and fix a d-dimensional time series x. We are interested in describing the algebraic properties of the iterated-sums signature of the transformed series $(f(\delta x_0), \dots, f(\delta x_{N-1}))$ where f acts in a vectorized fashion. That is, we wish to study the map

$$\langle \mathrm{ISS}^{(f)}(x)_{n,m}, a_1 \cdots a_p \rangle = \sum_{n \le i_1 < \cdots < i_p < m} f(\delta x_{i_1})^{a_1} \cdots f(\delta x_{i_p})^{a_p}.$$

It is immediate from the definition that, as a word series, $\mathrm{ISS}^{(f)}_{n,m}(x)$ admits the factorization

$$\mathrm{ISS}^{(f)}_{n,m}(x) = \overrightarrow{\prod_{n \le j < m}} \left(\varepsilon + \sum_{a \in \mathfrak{A}} f(\delta x_j)^a a \right).$$

In particular we have

Proposition 3.10. *The identity*

$$\mathrm{ISS}^{(f)}(x)_{n,n'}\, \mathrm{ISS}^{(f)}(x)_{n',n''} = \mathrm{ISS}^{(f)}(x)_{n,n''}$$

holds.

Since f vanishes at 0, the entries of $\mathrm{ISS}^{(f)}(x)$ are invariant to time-warping. Therefore, since the iterated-sums signature contains all such invariants, we are guaranteed to be able to express all said entries in terms of those in $\mathrm{ISS}(x)$. In order to describe this relation, we consider the map $f_\diamond \colon H \longrightarrow H$ induced by f in the following way: first we declare that

$$f_\diamond(\mathtt{i}) = \sum_{n=1}^{\deg f} c_n [\mathtt{i}^n]$$

for all $\mathtt{i} \in A$. This map extends uniquely to $S(A)$ as an algebra morphism $f_\diamond \colon S(A) \longrightarrow S(A)$. It further has a unique extension to all of H as a *concatenation* morphism, i.e. if $w = a_1 \cdots a_m \in \mathfrak{A}^*$ then $f_\diamond(w) = f_\diamond(a_1) \cdots f_\diamond(a_m)$.

Lemma 3.11. *The map $f_\diamond \colon H_{\mathrm{qsh}} \longrightarrow H_{\mathrm{qsh}}$ is an algebra morphism of the Hopf algebra H_{qsh} with quasi-shuffle product.*

Proof. We first show that $f_\diamond$ preserves the quasi-shuffle product, and we do so by induction. The base case holds by definition. The inductive definition of the quasi-shuffle product (4) then yields

$$\begin{aligned} f_\diamond(ua \star vb) &= f_\diamond\Big((u \star vb)a + (ua \star v)b + (u \star v)[ab]\Big) \\ &= f_\diamond(u \star vb) f_\diamond(a) + f_\diamond(ua \star v) f_\diamond(b) + f_\diamond(u \star v)[f_\diamond(a) f_\diamond(b)] \\ &= (f_\diamond(u) \star f_\diamond(vb)) f_\diamond(a) + (f_\diamond(ua) \star f_\diamond(v)) f_\diamond(b) + (f_\diamond(u) \star f_\diamond(v))[f_\diamond(a) f_\diamond(b)] \\ &= f_\diamond(u) f_\diamond(a) \star f_\diamond(v) f_\diamond(b) \\ &= f_\diamond(ua) \star f_\diamond(vb). \end{aligned}$$ □

Remark 3.12. The map $f_\diamond$ is not invertible. This is due to the fact that f^{-1} will in general be a formal power series and not just a polynomial. We can say that there is a loss of information, in terms of time-warping invariance, if we are only allowed to observe some polynomial transformation of the data instead of the data itself.

Theorem 3.13. *Let $f \in t\mathbb{R}[t]$. For all $w \in H$, the relation*

$$\langle \mathrm{ISS}^{(f)}(x)_{n,m}, w\rangle = \langle \mathrm{ISS}(x)_{n,m}, f_\diamond(w)\rangle$$

holds. In particular, $\mathrm{ISS}^{(f)}(x)$ *is a quasi-shuffle character.*

Proof. We first prove the identity on $S(A)$. For this, we first show it for $\mathtt{i} \in A$. Observe that

$$\langle \mathrm{ISS}(x)_{n,m}, f_\diamond(\mathtt{i})\rangle = \sum_{j=n}^{m-1} \sum_{p=1}^{\deg f} c_p(\delta x_j^{(i)})^p = \sum_{j=n}^{m-1} f(\delta x_j^{(i)}) = \langle \mathrm{ISS}^{(f)}(x)_{n,m}, \mathtt{i}\rangle.$$

If now $a = [\mathtt{i}_1 \cdots \mathtt{i}_p] \in \mathfrak{A}$ we see that

$$\begin{aligned}\langle \mathrm{ISS}(x)_{n,m}, f_\diamond(a)\rangle &= \langle \mathrm{ISS}(x)_{n,m}, [f_\diamond(\mathtt{i}_1)\cdots f_\diamond(\mathtt{i}_p)]\rangle \\ &= \sum_{j=n}^{m-1} f(\delta x_j^{(i_1)})\cdots f(\delta x_j^{(i_p)}) \\ &= \langle \mathrm{ISS}^{(f)}(x)_{n,m}, a\rangle.\end{aligned}$$

By linearity, the identity holds for all $a \in S(A)$.

Finally, by definition, if $w = a_1 \cdots a_p \in \mathfrak{A}^*$ then

$$\begin{aligned}\langle \mathrm{ISS}(x)_{n,m}, f_\diamond(w)\rangle &= \langle \mathrm{ISS}(x)_{n,m}, f_\diamond(a_1\cdots a_{p-1}) f_\diamond(a_p)\rangle \\ &= \sum_{j=n}^{m-1} \langle \mathrm{ISS}(x)_{n,m}, f_\diamond(a_1\cdots a_{p-1})\rangle f(\delta x_j)^{a_p} \\ &= \sum_{j=n}^{m-1} \langle \mathrm{ISS}^{(f)}(x)_{n,m}, a_1\cdots a_{p-1}\rangle f(\delta x_j)^{a_p} \\ &= \langle \mathrm{ISS}^{(f)}(x)_{n,m}, a_1\cdots a_p\rangle.\end{aligned}$$

The quasi-shuffle property follows immediately from Lemma 3.11. □

3.3. **Polynomial transformations.** We now consider polynomial transformations of the data. Let $P\colon \mathbb{R}^d \to \mathbb{R}^e$ be a polynomial map, for some $e \geq 1$. We write $P = (p_1, \ldots, p_e)$ where $p_k \in \mathbb{R}[x^{(1)}, \ldots, x^{(d)}]$ is a multivariate polynomial.

Recall from Section 2 that the quasi-shuffle algebra H_{qsh} carries a commutative quasi-shuffle structure. Moreover, by Theorem 2.5 it realizes the free commutative quasi-shuffle over $\mathbb{R}^d$; in other words, if H' is any other commutative quasi-shuffle algebra and $\Lambda\colon A \to H'$ is a map, there exists a unique extension $\Lambda\colon H_{\mathrm{qsh}} \to H'$ respecting the corresponding quasi-shuffle structures. In the following, we will work with different base alphabets, so we explicitly include the size of it in the notation. Hence from now on we write e.g. $H_{\mathrm{qsh}}(\mathbb{R}^d)$ to indicate this.

Given a time series x, we consider its transform $X = P(x) := (P(x_0), \ldots, P(x_N))$, which is an $\mathbb{R}^e$-valued time series. Interestingly enough, the iterated-sums signature of X can be computed just by knowing that of the untransformed data x. More precisely we have

Theorem 3.14. *Let $P\colon \mathbb{R}^d \to \mathbb{R}^e$, be a polynomial map without constant term, i.e., with $P(0) = 0$. Given a d-dimensional time series x with $x_0 = 0$, define the e-dimensional time series $X := P(x)$. Then, for all $0 \leq k \leq N$,*

$$\langle \mathrm{ISS}_{0,k}(X), w\rangle = \langle \mathrm{ISS}_{0,k}(x), \Lambda_P(w)\rangle, \tag{10}$$

DOI 10.20347/WIAS.PREPRINT.2795

where $\Lambda_P \colon H_{\mathrm{qsh}}(\mathbb{R}^e) \to H_{\mathrm{qsh}}(\mathbb{R}^d)$ *is the unique quasi-shuffle morphism (in the sense of Definition 2.2), determined by its action on* $[\mathtt{1}], \dots, [\mathtt{e}]$ *as*

$$\Lambda_P([\mathtt{i}]) \coloneqq \iota(p_i) \in H_{\mathrm{qsh}}(\mathbb{R}^d),$$

where $\iota \colon \mathbb{R}[x^{(1)}, \dots, x^{(d)}] \to H_{\mathrm{qsh}}(\mathbb{R}^d)$ *is the unique morphism of commutative algebras satisfying* $\iota(x^{(i)}) = [\mathtt{i}]$.

Example 3.15. Let $P \colon \mathbb{R}^2 \to \mathbb{R}^3$, $P = (p_1, p_2, p_3) = ((x^{(1)})^2, (x^{(2)})^3, x^{(1)}(x^{(2)})^2)$. Then

$$\begin{aligned}
\Lambda_P([\mathtt{1}]) &= [\mathtt{1}] \star [\mathtt{1}] \\
&= 2[\mathtt{1}][\mathtt{1}] + [\mathtt{1}^2] \\
\Lambda_P([\mathtt{2}]) &= [\mathtt{2}] \star [\mathtt{2}] \star [\mathtt{2}] \\
&= 6[\mathtt{2}][\mathtt{2}][\mathtt{2}] + 3[\mathtt{2}^2][\mathtt{2}] + 3[\mathtt{2}][\mathtt{2}^2] + [\mathtt{2}^3] \\
\Lambda_P([\mathtt{3}]) &= [\mathtt{1}] \star [\mathtt{2}] \star [\mathtt{2}] \\
&= 2[\mathtt{2}][\mathtt{2}][\mathtt{1}] + 2[\mathtt{1}][\mathtt{2}][\mathtt{2}] + 2[\mathtt{2}][\mathtt{1}][\mathtt{2}] + 2[\mathtt{2}][\mathtt{12}] + 2[\mathtt{12}][\mathtt{2}] + [\mathtt{1}][\mathtt{2}^2] + [\mathtt{2}^2][\mathtt{1}] + [\mathtt{12}^2].
\end{aligned}$$

Proof. Since by Proposition 3.4 both sides of eq. (10) are quasi-shuffle morphisms, it is enough to show that it holds for letters $[\mathtt{1}], \dots, [\mathtt{e}]$. Now, on one hand, by definition

$$\langle \mathrm{ISS}_{0,k}(X), [\mathtt{i}] \rangle = \sum_{j=0}^{k-1} \delta X_j^{(i)} = X_k^{(i)} = p_i(x_k)$$

On the other hand, by definition

$$\begin{aligned}
\langle \mathrm{ISS}_{0,k}(x), \Lambda_P([\mathtt{i}]) \rangle &= \langle \mathrm{ISS}_{0,k}(x), \iota(p_i) \rangle \\
&= p_i(\langle \mathrm{ISS}(x)_{0,k}, \mathtt{1} \rangle, \dots, \langle \mathrm{ISS}(x)_{0,k}, \mathtt{d} \rangle) \\
&= p_i(x_k)
\end{aligned}$$

and the proof is finished.

□

Corollary 3.16. *Let* $P \colon \mathbb{R}^d \to \mathbb{R}^e$ *be a polynomial map, and* x *a* d*-dimensional time series. Define the* e*-dimensional time series* $X \coloneqq P(x)$*, and set* $\tilde{P}_{x_0} \coloneqq P(\cdot + x_0) - P(x_0)$*. Then, for all* $0 \le k \le N$,

$$\mathrm{ISS}(X)_{0,k} = \mathrm{ISS}(x) \circ \Lambda_{\tilde{P}_{x_0}}.$$

Proof. The result follows from Theorem 3.14 and the fact that, if $\tilde{x} = x_\cdot - x_0$ then $\tilde{x}_0 = 0$ and $\mathrm{ISS}(\tilde{x})_{0,k} = \mathrm{ISS}(x)_{0,k}$ for all $0 \le k \le N$. □

References

[1] P. Cartier, *On the structure of free Baxter algebras*, Adv. Math. **9** (1972), 253–265 (English).

[2] L. Colmenarejo and R. Preiß, *Signatures of paths transformed by polynomial maps*, Beitr. Algebra Geom. **61** (2020), no. 4, 695–717.

[3] J. Diehl, K. Ebrahimi-Fard, and N. Tapia, *Time-warping invariants of multidimensional time series*, Acta Appl. Math. **170** (2020), no. 1, 265–290.

[4] ———, *Tropical time series, iterated-sums signature and quasisymmetric functions*, 2020, arXiv:2009.08443 [math.RA].

[5] K. Ebrahimi-Fard and F. Patras, *Quasi-shuffle algebras in non-commutative stochastic calculus*, 2020, arXiv:2004.06945 [math.PR], (to appear in proceedings volume "Geometry and Invariance in Stochastic Dynamics", Springer Proceedings in Mathematics & Statistics).

[6] L. Foissy and F. Patras, *Lie theory for quasi-shuffle bialgebras*, Periods in Quantum Field Theory and Arithmetic (J. I. Burgos Gil, K. Ebrahimi-Fard, and D. Manchon, eds.), Springer Proceedings in Mathematics & Statistics, vol. 314, Springer Berlin Heidelberg, 2020, pp. 537–615.

[7] L. Foissy, F. Patras, and J.-Y. Thibon, *Deformations of shuffles and quasi-shuffles*, Ann. Inst. Fourier (Grenoble) **66** (2016), no. 1, 209–237.
[8] J. G. Gaines, *The algebra of iterated stochastic integrals*, Stochastics and Stochastic Reports **49** (1994), no. 3-4, 169–179.
[9] M. E. Hoffman, *Quasi-shuffle products*, J. Alg. Combin. **11** (2000), no. 1, 49–68.
[10] M. E. Hoffman and K. Ihara, *Quasi-shuffle products revisited*, J. Algebra **481** (2017), 293–326.
[11] F. J. Király and H. Oberhauser, *Kernels for sequentially ordered data*, J. Mach. Learn. Res. **20** (2019), no. 31, 1–45.
[12] J.-L. Loday, *On the algebra of quasi-shuffles*, Manuscripta Math. **123** (2007), no. 1, 79–93. MR 2300061
[13] T. Lyons and H. Oberhauser, *Sketching the order of events*, 2017, `arXiv:1708.09708 [stat.ML]`.
[14] K. Newman and D. E. Radford, *The cofree irreducible Hopf algebra on an algebra*, Am. J. Math. **101** (1979), no. 5, 1025–1045 (English).
[15] J.-C. Novelli and J.-Y. Thibon, *Free quasi-symmetric functions and descent algebras for wreath products, and non-commutative multi-symmetric functions*, Discrete Math. **310** (2010), no. 24, 3584–3606.
[16] R. Ree, *Lie elements and an algebra associated with shuffles*, Ann. Math. (2) **68** (1958), no. 2, 210–220 (English).
[17] C. Reutenauer, *Free Lie algebras*, Handbook of Algebra, Oxford: Clarendon Press, 1993, pp. xvii + 269 (English).
[18] C. Toth, P. Bonnier, and H. Oberhauser, *Seq2tens: An efficient representation of sequences by low-rank tensor projections*, 2020, `arXiv:2006.07027 [cs.LG]`.