Séminaire Lotharingien de Combinatoire XX (2020)
Article #YY, 12 pp.

Proceedings of the 32nd *Conference on Formal Power Series and Algebraic Combinatorics (Ramat Gan)*

Iterated-sums signature, quasi-symmetric functions and time series analysis

Joscha Diehl*[1] Kurusch Ebrahimi-Fard†[2] and Nikolas Tapia‡[3,4]

[1] *Universität Greifswald, Institut für Mathematik und Informatik, Walther-Rathenau-Str. 47, 17489 Greifswald, Germany.*
[2] *Department of Mathematical Sciences, NTNU, 7491 Trondheim, Norway.*
[3] *Weierstraß-Institut Berlin, Mohrenstr. 39, 10117 Berlin, Germany.*
[4] *Technische Universität Berlin, Str. des 17. Juni 136, 10623 Berlin, Germany.*

Abstract. We survey and extend results on a recently defined character on the quasi-shuffle algebra. This character, termed iterated-sums signature, appears in the context of time series analysis and originates from a problem in dynamic time warping. Algebraically, it relates to (multidimensional) quasi-symmetric functions as well as (deformed) quasi-shuffle algebras.

Keywords: Shuffle and quasi-shuffle Hopf algebras; half-shuffles; iterated integral- and sum-signatures; quasi-symmetric functions; Hoffman's exponential.

1 Introduction

In his seminal 1954 paper [3], K.-T. Chen introduced the iterated-integral signature of a smooth path taking values in a finite-dimensional smooth manifold. We recall the definition for the special case of curves on d-dimensional Euclidean space using the shuffle product introduced by Ree [13]. Let $A = \{1, \ldots, d\}$ be a finite alphabet, and let A^* denote the free monoid, which consists of all words with letters from A; the unit element, the *empty word* is denoted by e. A noncommutative product is *concatenation* of words, denoted by juxtaposition. The linear space H spanned[1] by A^* has an algebra structure given by the *shuffle product* $⧢ \colon H \otimes H \to H$, recursively defined by $e ⧢ u := u =: u ⧢ e$ for all $u \in H$, and

$$ua ⧢ vb := (u ⧢ vb)a + (ua ⧢ v)b$$

for $u, v \in H$ and $a, b \in A$. It is a standard result that $(H, ⧢, e)$ is a commutative algebra [14].

*joscha.diehl@gmail.com
†kurusch.ebrahimi-fard@ntnu.no
‡tapia@wias-berlin.de. Nikolas Tapia was supported by the BMS MATH+ EF1-5 project "On robustness of deep neural networks".

[1] In the following all algebraic structures are defined over a base field $\mathbb{F}$ of characteristic zero.

Given a smooth path $x\colon [0,1] \to \mathbb{R}^d$, $t \mapsto x(t) = (x^1(t),\dots,x^d(t))$ and a word $w = i_1 \cdots i_n \in A^*$ define the iterated integral

$$S(x)^w := \int\cdots\int_{0<s_1<\cdots<s_n<1} \dot{x}^{i_1}(s_1)\cdots\dot{x}^{i_n}(s_n)\,\mathrm{d}s_1\cdots\mathrm{d}s_n.$$

This can be linearly extended to all of H in a unique way. In the literature either the collection $(S(x)^w : w \in A^*)$ or its linear extension $S(x)\colon w \mapsto S(x)^w \in H^*$ are known as the *iterated-integral signature* of x.

Chen showed that the signature satisfies the so-called *shuffle relations*, generalizing integration-by-parts to iterated integrals: for any $v, w \in H$,

$$S(x)^{v \sqcup\!\sqcup w} = S(x)^v S(x)^w. \tag{1.1}$$

In other words, if we regard the map $S(x)$ as a formal word series $S(x) = \sum_{w\in A^*} S(x)^w w$, then it is a group-like element. One could also say that $S(x)$ is a character over the shuffle algebra. In any case, its *logarithm* $\Lambda(x) := \log S(x)$ is well defined as an element of the free Lie algebra on d generators.

In control theory, the coefficients of the signature actually provide a universal description of solutions to affinely controlled ODEs. For a fixed path $x\colon [0,1] \mapsto \mathbb{R}^d$ as before, and smooth vector fields $f_1,\dots,f_d$ on $\mathbb{R}^q$, consider the initial value problem

$$\dot{y}(t) = \sum_{i=1}^{d} f_i(y(t))\dot{x}^i(t), \quad y(0) = \xi \in \mathbb{R}^q. \tag{1.2}$$

Using Picard iteration, the final value $y(1)$ can be expressed as a series:

$$y(1) = \sum_{w\in A^*} f_w(\xi) S(x)^w \tag{1.3}$$

where the functions f_w are defined recursively via the relation $f_e(y) = y$ and $f_{iw}(y) = Df_w(y)f_i(y)$. T. Lyons' insight was, that this expansion generalizes to control systems with irregular (in particular: non-differentiable) drivers, in what is now known as *rough paths theory* [10]. The main philosophy of the theory is that Equation (1.2) should be interpreted as an integral equation, and the iterated integrals appearing in Equation (1.3), which may not exist, should be replaced by an object –a geometric rough path– satisfying properties similar to those of the signature, that has to be supplied as an input to the problem.

Numerical schemes for Equation (1.2) are obtained by integrating the equation over small interval of size $h > 0$, so that for a fixed $t \in (0,1)$

$$y(t+h) - y(t) = \sum_{i=1}^{d} \int_t^{t+h} f_i(y(u))\dot{x}^i(u)\,\mathrm{d}u = \sum_{i=1}^{d} f_i(y(t))(x^i(t+h) - x^i(t)) + \mathrm{o}(h).$$

Setting $y_k := y(kh)$ and $x_k := x(kh)$ for $k = 0, \dots, N$ we are led to consider the associated finite difference equation

$$y_{k+1} - y_k = \sum_{i=1}^{d} f_i(y_k)(x^i_{k+1} - x^i_k).$$

Similar to the continuous case, it can be shown that y_N can be expressed as a series:

$$y_N = \sum_{w \in A^*} f_w(\xi)\mathsf{DS}(x)^w$$

where now the coefficient $\mathsf{DS}(x)^w$ is defined using *iterated sums* instead of integrals, that is, if $w = i_1 \cdots i_n$ then

$$\mathsf{DS}(x)^w := \sum_{0<k_1<\cdots<k_n<N} \cdots \sum \delta x^{i_1}_{k_1} \cdots \delta x^{i_n}_{k_n} \tag{1.4}$$

and we have defined increments $\delta x_j := x_{j+1} - x_j$ for convenience. As before, we extend this definition linearly to H.

We observe that $\mathsf{DS}(x)$ *does not* satisfy the shuffle relations (1.1). For example,

$$\mathsf{DS}(x)^{i \sqcup\!\sqcup j} = \mathsf{DS}(x)^{ij} + \mathsf{DS}(x)^{ji} = \mathsf{DS}(x)^i \mathsf{DS}(x)^j - \sum_{0<k<N} \delta x^i_k \delta x^j_k.$$

The last term on the right-hand side cannot be expressed as a linear combination of the coefficients in Equation (1.4). The correct way to describe the product rule satisfied by the iterated sums requires another product on words, generalizing the shuffle product, $\sqcup\!\sqcup$, over a larger alphabet. This product is known as *quasi-shuffle*. See, e.g.-[6, 7].

To describe the quasi-shuffle product, we first need to extend the alphabet A to a commutative semigroup A. The internal law on A, which is associative and commutative, will be denoted by using square brackets. By construction, every element of A can be written uniquely (up to commutativity) as an iteration of brackets

$$[i_1[i_2[\cdots i_n]]] := [i_1 i_2 \cdots i_n], \quad i_1, \dots, i_n \in A$$

where the definition on the right is consistent by associativity. We now denote by A^* the free monoid, with empty word e. The linear space H spanned by A^* has an algebra structure through the quasi-shuffle product $\star\colon \mathsf{H} \otimes \mathsf{H} \to \mathsf{H}$, recursively defined by

$$e \star u = u \star e \quad \text{and} \quad ua \star vb := (u \star vb)a + (ua \star v)b + (u \star v)[ab]$$

for $u, v \in \mathsf{A}^*$ and $a, b \in \mathsf{A}$.

We now extend the definition of DS in (1.4) to include letters from A by setting

$$\delta x_j^{[i_1 \cdots i_n]} := \prod_{k=1}^{n} \delta x_j^{i_k}.$$

With this notation, the previous example rewrites

$$\mathsf{DS}(x)^{i \star j} = \mathsf{DS}(x)^{ij+ji+[ij]} = \mathsf{DS}(x)^{i}\mathsf{DS}(x)^{j}.$$

Definition 1 ([4]). *The* iterated-sums signature *of the discrete sequence $x = (x_0, \ldots, x_N)$ is the collection $\mathsf{DS}(x) := (\mathsf{DS}(x)^w : w \in \mathsf{A})$ defined by Equation (1.4).*

As before, we will not distinguish between the collection $\mathsf{DS}(x)$ and its unique linear extension to H.

Theorem 2 ([4]). *The iterated-sums signature satisfies the* quasi-shuffle relations, $\mathsf{DS}(x)^{v \star w} = \mathsf{DS}(x)^{v}\mathsf{DS}(x)^{w}$*, for all $v, w \in \mathsf{H}$.*

Using strict inequalities in (1.4) seems to be arbitrary. In fact, another signature is defined as follows

$$\mathsf{DS}_{-1}(x)^w := \sum_{0<k_1\le k_2 \le \cdots \le k_{n-1} \le k_n \le N} \delta x_{k_1}^{i_1} \cdots \delta x_{k_n}^{i_n}.$$

This is a character on another algebra. To formulate this we immediately introduce the general notation. On H define for $\theta \in \mathbb{R}$ the *θ-weight quasi-shuffle* $\star_\theta$ recursively as

$$e \star_\theta w = w = w \star_\theta e \quad \text{and} \quad wa \star_\theta vb = (w \star_\theta vb)a + (wa \star_\theta v)b + \theta(w \star_\theta v)[ab].$$

For $\theta = 0$ this is the (classical) shuffle, for $\theta = 1$ this is the quasi-shuffle $\star = \star_1$ defined earlier, and we shall keep using the former symbol when convenient. Now: $\mathsf{DS}_{-1}(x)$ is a character on $(\mathsf{H}, \star_{-1})$. In the next Section, we will see how to translate between $\mathsf{DS}_{-1}(x)$, $\mathsf{DS}(x)$ and more general "signatures".

Regarding Theorem 2, we mention here that for the case $d = 1$ there is an immediate interpretation in terms of quasi-symmetric functions [11], [9]. In the multidimensional case, Novelli and Thibon's quasi-symmetric functions of level d [12] are the right object.

Lastly, we briefly mention the connection to time series analysis: in [4], we set out to find polynomial functions of a time series $x = (x_0, x_1, \ldots, x_N) \in (\mathbb{R}^d)^{N+1}$ that are invariant to *time warping*. We skip the precise definition, but morally these are functions that do not change when the time series is run at a different speed. It turns out for $d = 1$ these are *exactly* the quasi-symmetric functions in the variables $x_0, x_1, \ldots, x_N, x_{N+1}, \ldots$ where we extend the time series constantly as $x_n = x_N$ for $n > N$. For $d \ge 2$ these invariants should correspond to "d-dimensional quasi-symmetric functions". These are Novelli–Thibon's quasi-symmetric function of level d [12]. In all cases $d \ge 1$ the iterated-sums signature, introduced in [4], stores these quasi-symmetric functions, evaluated on some time series, as the character on the quasi-shuffle Hopf algebra. This signature can be seen as the (polynomial) *feature map* corresponding to the *dynamic time warping (DTW)* distance [1], a heavily used distance in the realm of time series analysis.

In the next section we will look at the algebras $(\mathsf{H}, \star_\theta)$ and maps between them. In the last section we present some observations and open questions.

2 (Quasi)-shuffle morphisms

The rather elegant algebraic description by Hoffman and Ihara of quasi-shuffle homomorphisms [7, 8] will be used now. We first recall their notation. For a power series f in t, with zero constant coefficient, $f(t) = \sum_{n=1}^{\infty} c_n t^n$, define the linear map $\Psi_f : H \to H$

$$w \mapsto \Psi_f(w) = \sum_{I=(i_1,\ldots,i_m)\in C(\ell(w))} c_{i_1} \cdots c_{i_m} I[w].$$

Here $C(n)$ is the set of all compositions of the integer p, i.e., tuples $(i_1, \ldots, i_p)$ of positive integers such that $i_1 + \cdots + i_p = n$. Given $I = (i_1, \ldots, i_p) \in C(n)$ and a word $w = w_1 \cdots w_n \in \mathsf{A}^*$ of length $\ell(w) = n > 0$, we define a new word $I[w] \in \mathsf{A}^*$ by

$$I[w] := [w_1 \cdots w_{i_1}][w_{i_1+1} \cdots w_{i_1+i_2}] \cdots [w_{i_1+\cdots+i_{p-1}+1} \cdots w_n].$$

Here (as well as later) we are using the suitable convention that $[a] := a$ for all $a \in \mathsf{A}$.

In [8] an isomorphism from $(H, \star_{+1}) \to (H, \star_{-1})$ is given. We generalize this and let for $\theta \in \mathbb{R}$

$$f_\theta(t) := \frac{1}{\theta}(e^{\theta t} - 1) = \sum_{n\geq 1} \frac{\theta^{n-1}}{n!} t^n,$$

where the last line makes also sense for $\theta = 0$. Define

$$f_\theta^{-1}(t) = \frac{1}{\theta} \log(1 + \theta t) = \sum_{n\geq 1} \frac{\theta^{n-1}}{n} t^n,$$

which, again, makes also sense for $\theta = 0$.

Lemma 3. $\exp_\theta : (H, \star_0) \to (H, \star_\theta)$, $\exp_\theta := \Psi_{f_\theta}$, *is a Hopf algebra isomorphism, with inverse* $\log_\theta := \Psi_{f_\theta^{-1}}$.

Corollary 4. *For* $\theta, \theta' \in \mathbb{R}$, *the map* $E_{\theta\to\theta'} := \exp_{\theta'} \circ \log_\theta : (H, \star_\theta) \to (H, \star_{\theta'})$, *is a Hopf isomorphism and* $E_{\theta\to\theta'} = \Psi_{e_{\theta\to\theta'}}$, *where*

$$\begin{aligned} e_{\theta\to\theta'}(t) &= \frac{1}{\theta'}(e^{\frac{\theta'}{\theta}\log(1+\theta t)} - 1) \\ &= t - \frac{\theta - \theta'}{2!} t^2 + \frac{(\theta - \theta')(2\theta - \theta')}{3!} t^3 - \frac{(\theta - \theta')(2\theta - \theta')(3\theta - \theta')}{4!} t^4 \\ &\quad + \frac{(\theta - \theta')(2\theta - \theta')(3\theta - \theta')(4\theta - \theta')}{5!} t^5 - \ldots \end{aligned}$$

Remark 5. *Using this map E and starting from the character* $\mathsf{DS}(x)$ *on* $(H, \star_{+1})$ *we can construct characters* $\mathsf{DS}_\theta(x)$ *on* $(H, \star_\theta)$ *by defining*

$$\Big\langle w, \mathsf{DS}_\theta(x) \Big\rangle := \Big\langle E_{\theta\to 1} w, \mathsf{DS}(x) \Big\rangle.$$

We note that $\mathsf{DS}_{+1}(x) = \mathsf{DS}(x)$ *and one can show that*

$$\Big\langle [a_1]\cdots[a_p], \mathsf{DS}_{-1}(x) \Big\rangle = \sum_{0\le i_1\le\cdots\le i_p} \delta x_{i_1}^{[a_1]}\cdots\delta x_{i_p}^{[a_p]}.$$

In other words, $\mathsf{DS}_{-1}(x)$ *is defined like* $\mathsf{DS}(x)$ *but with all strict inequalities (in the sum over timepoints) replaced by weak inequalities.*

We also note that $\mathsf{DS}_0(x)$ *is the iterated-integrals signature of the piecewise linear interpolation of the (infinite dimensional) time series* X^a *indexed by* $a = [1^{k_1}\cdots d^{k_d}] \in \mathsf{A}$ *and given by*

$$n \mapsto X_n^a = \sum_{j=1}^n \delta x_j^a = \sum_{j=1}^n (\delta x_j^{[1]})^{k_1}\cdots(\delta x_j^{[d]})^{k_d}.$$

For $\theta \notin \{-1, 0, +1\}$ *we currently do not have a satisfying alternative characterization of* $\mathsf{DS}_\theta(x)$*.*

Example 6.

$$e_{1/2\to 1}(t) = (e^{2\log(1+t/2)} - 1) = ((1+t/2)^2 - 1) = t + \frac{1}{4}t^2.$$

We get for example $\langle [1], \mathsf{DS}_{1/2}(x)\rangle = \langle [1], \mathsf{DS}(x)\rangle$ *and*

$$\Big\langle [1][1], \mathsf{DS}_{1/2}(x) \Big\rangle = \Big\langle [1][1] + \frac{1}{4}[1^2], \mathsf{DS}(x) \Big\rangle \quad \Big\langle [1^2], \mathsf{DS}_{1/2}(x) \Big\rangle = \Big\langle [1^2], \mathsf{DS}(x) \Big\rangle,$$

and hence

$$\begin{aligned} \Big\langle [1], \mathsf{DS}_{1/2}(x) \Big\rangle^2 &= \Big\langle [1], \mathsf{DS}(x) \Big\rangle^2 = \Big\langle [1] \star_1 [1], \mathsf{DS}(x) \Big\rangle = \Big\langle 2[1][1] + [1^2], \mathsf{DS}(x) \Big\rangle \\ &= \Big\langle 2[1][1] + 1/2[1^2], \mathsf{DS}_{1/2}(x) \Big\rangle = \Big\langle [1] \star_{1/2} [1], \mathsf{DS}_{1/2}(x) \Big\rangle, \end{aligned}$$

as expected.

We finally note that these different concepts of summation / integration appear naturally in the field of *stochastic analysis*. Stochastic integration theory starts from Riemann-type sums over stochastic processes[2] X, Y, namely

$$I(X,Y) = \int_0^1 X\,\mathrm{d}Y \approx \sum_{i=0}^n X_{t_i}(Y_{t+i} - Y_{t_i}).$$

The approximation here is in a probabilistic sense, meaning that the limiting procedure should also take into account the stochastic nature of the setting. Due to the particular analytic properties of these processes, the choice of the evaluation point t_i in the above

[2] We recall that a stochastic process is a collection of random variables $(X_t : t \in [0,1])$ [15].

discrete approximation is a subtle matter. In a nutshell, different choices of this value lead to very different notions of stochastic integrals. The choices X_{t_i}, $X_{t_{i+1}}$ and $\frac{1}{2}(X_{t_i} + X_{t_{i+1}})$ corresponds to the main three stochastic integrals, i.e. Itō, backward Itō and Stratonovich, respectively. Each of these integrals has its own integration by parts rule, and they are tightly related analytically, as well as algebraically. Quasi-shuffles enter the picture when one considers these integrals at the level of the discrete approximations, since the multiplication of iterated sums follows summation-by-parts. For example, for Itō integration,

$$
\begin{aligned}
&(X_1 - X_0)(Y_1 - Y_0) = \big(\sum_i (X_{t_{i+1}} - X_{t_i})\big)\big(\sum_j (Y_{t_{j+1}} - Y_{t_j})\big) \\
&= \sum_{i<j} (X_{t_i} - X_0)(Y_{t_{j+1}} - Y_{t_j}) + \sum_{j<i} (Y_{t_j} - Y_0)(X_{t_{i+1}} - X_{t_i}) + \sum_i (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}) \\
&\approx \int_0^1 (X_t - X_0)\,\mathrm{d}Y_t + \int_0^1 (Y_t - Y_0)\,\mathrm{d}X_t + \langle X, Y\rangle_t \qquad \text{(Itō)} \\
&\approx \int_0^1 (X_t - X_0) \circ \mathrm{d}Y_t + \int_0^1 (Y_t - Y_0) \circ \mathrm{d}X_t \qquad \text{(Stratonovich)} \\
&\approx \int_0^1 (X_t - X_0)\,\hat{\mathrm{d}}Y_t + \int_0^1 (Y_t - Y_0)\,\hat{\mathrm{d}}X_t - \langle X, Y\rangle_t \qquad \text{(backward Itō)}
\end{aligned}
$$

The term $\langle X, Y\rangle$, known as stochastic *bracket*, is an artifact of the diagonal term that appears when multiplying iterated sums.

3 Observations

3.1 Expectation values of discrete signatures

We have seen that $\mathsf{DS}(x)$ is a character over the quasi-shuffle Hopf algebra $(H, \star)$. That means it is an element in the group of algebra morphisms, $\tilde{G} \subset G$, where G is the larger group of invertible linear maps sending the empty word to one. Both groups have corresponding Lie algebras, $\tilde{\mathfrak{g}} \subset \mathfrak{g}$, with $\tilde{\mathfrak{g}}$ containing the infinitesimal characters, whereas $\mathfrak{g}$ consists of linear maps that send the empty word to zero. It is clear that for any $\Phi \in \tilde{G}$ we can find a unique $\alpha \in \tilde{\mathfrak{g}}$, such that $\Phi = \exp^{\circledast}(\alpha)$. Analogously, the group G is in bijection with $\mathfrak{g}$. Recall that the convolution product of linear maps

$$\alpha \circledast \beta := m_{\mathbb{F}}(\alpha \otimes \beta)\Delta$$

is defined in terms of the deconcatenation coproduct on words

$$\Delta(w) = w \otimes e + e \otimes w + \sum_{i=1}^{n-1} w_1 \cdots w_i \otimes w_{i+1} \cdots w_n.$$

Let x be a random sequence, i.e. it has random variables as entries and hence $\mathsf{DS}(x)$ is itself a random variable. The expectation of the iterated-sums signature is defined as the liner map given by

$$\mu_x^w := \mathbb{E}[\mathsf{DS}(x)^w] = \mathbb{E}\Bigg[\sum_{0<k_1<\cdots<k_n<N}\!\!\cdots\sum \delta x_{k_1}^{i_1}\cdots\delta x_{k_n}^{i_n}\Bigg] \in \mathbb{R}.$$

Note that $\mu_x \in G$. Hence

$$\kappa_x := \log^{\circledast} \mu_x$$

is a well-defined linear map in the Lie algebra $\mathfrak{g}$, sending the empty word to zero. Defining $\mu_x' := \mu_x - \epsilon$, where ϵ is the counit in $(H, \star)$, we have

$$\begin{aligned}\kappa_x &= -\sum_{n>0} \frac{(-1)^n}{n}(\mu_x')^{\circledast n} \\ &= \mu_x' - \frac{1}{2}(\mu_x')^{\circledast 2} + \frac{1}{3}(\mu_x')^{\circledast 3} - \frac{1}{4}(\mu_x')^{\circledast 4} + \cdots .\end{aligned}$$

It is then easy to see that for any word

$$\mu_x^w = \sum_{m>0}^{|w|} \frac{1}{m!} \sum_{v_1\cdots v_m = w} \kappa_x^{v_1}\cdots\kappa_x^{v_m}$$

and

$$\kappa_x^w = -\sum_{m>0}^{|w|} \frac{(-1)^m}{m} \sum_{\substack{v_1\cdots v_m = w\\ v_i \neq \mathbf{1}}} \mu_x^{v_1}\cdots\mu_x^{v_m}.$$

Following the terminology for random variables in probability theory and call κ_x the *cumulant map* and μ_x the *moment map*. Motivated by [2], which compute so-called signature moments and cumulants, we address the more interesting problem of calculating expressions like $\kappa_x^{[1][2]\star[3]}$ without expanding the quasi-shuffle product.

$$\begin{aligned}\kappa_x^{[1][2]\star[3]} &= -\sum_{n>0}\frac{(-1)^n}{n}(\mu_x')^{\circledast n}([1][2]\star[3]) \\ &= \mu_x^{[1][2]\star[3]} - \mu_x^{[1][2]}\mu_x^{[3]} - \frac{1}{2}\mu_x^{[1]\star[3]}\mu_x^{[2]} - \frac{1}{2}\mu_x^{[1]}\mu_x^{[2]\star[3]} \\ &\quad + \frac{1}{3}(\mu_x' \otimes \mu_x' \otimes \mu_x')(\Delta\otimes \mathrm{id})\Delta([1][2]\star[3]) \\ &= \mathbb{E}\Big[\mathsf{DS}(x)^{[1][2]}\mathsf{DS}(x)^{[3]}\Big] - \mathbb{E}\Big[\mathsf{DS}(x)^{[1][2]}\Big]\mathbb{E}\Big[\mathsf{DS}(x)^{[3]}\Big] \\ &\quad - \frac{1}{2}\mathbb{E}\Big[\mathsf{DS}(x)^{[1]}\mathsf{DS}(x)^{[3]}\Big]\mathbb{E}\Big[\mathsf{DS}(x)^{[2]}\Big] - \frac{1}{2}\mathbb{E}\Big[\mathsf{DS}(x)^{[1]}\Big]\mathbb{E}\Big[\mathsf{DS}(x)^{[2]}\mathsf{DS}(x)^{[3]}\Big]\end{aligned}$$

$$+ \mathbb{E}\Big[\mathsf{DS}(x)^{[1]}\Big]\mathbb{E}\Big[\mathsf{DS}(x)^{[2]}\Big]\mathbb{E}\Big[\mathsf{DS}(x)^{[3]}\Big].$$

Here we used the fact that $\mathsf{DS}(x)$ is an algebra morphism for the quasi-shuffle product $\star$. Using that $\kappa_x^{\mathbf{1}} = 0$, we can invert this expansion and express μ_x in terms of κ_x

$$\mu_x^{[1][2]\star[3]} = \kappa_x^{[1][2]\star[3]} + \kappa_x^{[1][2]}\kappa_x^{[3]} + \frac{1}{2}\kappa_x^{[1]\star[3]}\kappa_x^{[2]} + \frac{1}{2}\kappa_x^{[1]}\kappa_x^{[2]\star[3]} + \frac{1}{2}\kappa_x^{[1]}\kappa_x^{[2]}\kappa_x^{[3]}.$$

The general formula for expressing cumulants and moments, κ_x resp. μ_x, in terms of each other is obtained by recalling the Hopf algebra isomorphism F from the Hopf subalgebra of rooted ladder trees with decorations in the Butcher–Connes–Kreimer Hopf algebra of rooted trees to the quasi-shuffle Hopf algebra $(H, \star)$. Quasi-shuffle products of words are identified with forests of decorated ladder trees. Hence, iterated (reduced) deconcatenation coproducts on $w_1 \star \cdots \star w_n$ can be computed in terms of iterated (reduced) coproducts on forests of ladders. For the above example we obtain

$$F([1][2] \star [3]) = F([1][2])F([3]) = \begin{smallmatrix}\bullet_2\\ \vert\\ \bullet_1\end{smallmatrix}\bullet_3$$

and compute the reduced Connes–Kreimer coproduct Δ'_{CK} on the forest of two ladders

$$\Delta'_{\mathrm{CK}}\begin{smallmatrix}\bullet_2\\ \vert\\ \bullet_1\end{smallmatrix}\bullet_3 = \begin{smallmatrix}\bullet_2\\ \vert\\ \bullet_1\end{smallmatrix} \otimes \bullet_3 + \bullet_3 \otimes \begin{smallmatrix}\bullet_2\\ \vert\\ \bullet_1\end{smallmatrix} + \bullet_2 \otimes \bullet_1\bullet_3 + \bullet_2\bullet_3 \otimes \bullet_1.$$

Here the root-part is on the right. The first iterated reduced coproduct

$$(\Delta'_{\mathrm{CK}} \otimes \mathrm{id})\Delta'_{\mathrm{CK}}\begin{smallmatrix}\bullet_2\\ \vert\\ \bullet_1\end{smallmatrix}\bullet_3 = \bullet_2 \otimes \bullet_1 \otimes \bullet_3 + \bullet_2 \otimes \bullet_3 \otimes \bullet_1 + \bullet_3 \otimes \bullet_2 \otimes \bullet_1.$$

This should be compared with the expression for $\kappa_x^{[1][2]\star[3]}$. Hence, a product $w_1 \star \cdots \star w_n$ corresponds to a forest $F(w_1 \star \cdots \star w_n) = t_1 \cdots t_n$. We denote the degree of a forest, i.e., the total number of its vertices, by $|t_1 \cdots t_n| := \sum_{i=1}^n |t_i|$, where $|t_i|$ corresponds to the number of letters of the word $w_i = F^{-1}(t_i)$. Defining $\tilde{\kappa}_x := \kappa_x \circ F^{-1}$ and $\tilde{\mu}_x := \mu_x \circ F^{-1}$, we find for

$$\begin{aligned}
&\kappa_x^{w_1\star\cdots\star w_n} = \tilde{\kappa}_x(t_1 \cdots t_n) \\
&= -\sum_{m=1}^{|t_1\cdots t_n|} \frac{(-1)^m}{m} \tilde{\mu}_x^{\otimes m} \Delta_{\mathrm{CK}}^{[m-1]}(t_1) \cdots \Delta_{\mathrm{CK}}^{[m-1]}(t_n) \\
&= -\sum_{m=1}^{|t_1\cdots t_n|} \frac{(-1)^m}{m} \sum_{(t_1,\ldots,t_n)}' \tilde{\mu}_x(t_1^{(1)} \cdots t_n^{(1)}) \cdots \tilde{\mu}_x(t_1^{(m)} \cdots t_n^{(m)}) \\
&= \sum_{m=1}^{N} \frac{(-1)^{m-1}}{m} \sum_{(w_1,\ldots,w_n)}' \mathbb{E}\Big[\mathsf{DS}(x)\big(w_1^{(1)}) \cdots \mathsf{DS}(x)(w_n^{(1)})\Big] \cdots \mathbb{E}\Big[\mathsf{DS}(x)(w_1^{(m)}) \cdots \mathsf{DS}(x)(w_n^{(m)})\Big],
\end{aligned}$$

where $N := |t_1 \cdots t_n| = |w_1 \star \cdots \star w_n|$ and the primed sums refer to the constraint that none of the forests $t_1^{(i)} \cdots t_n^{(i)}$ (words $w_1^{(i)} \cdots w_n^{(i)}$), for $i = 1, \ldots, m$, can be the empty word, i.e., $F^{-1}(t_1^{(i)} \cdots t_n^{(i)}) \neq \mathbf{1}$. This sum can be expressed in terms of linearly ordered partitions constructed as follows.

$$\kappa_x^{w_1 \star \cdots \star w_n} = \tilde{\kappa}_x(t_1 \cdots t_n) = -\sum_{m=1}^{|t_1 \cdots t_n|} \frac{(-1)^m}{m} \sum_{\pi \in OP_m} \tilde{\mu}'_x(t_{\pi_1}) \cdots \tilde{\mu}'_x(t_{\pi_m}).$$

The second sum on the right-hand side of the second equality runs over ordered partition with m blocks. The computation of the order m, $\pi := \{\pi_1, \cdots, \pi_m\}$ and its blocks π_i is summarized in the following algorithm. The first step consists in partitioning $I \cup J = [n]$ into subsets, where $I \neq \emptyset$. Then consider the corresponding subsets of trees, $t_I = t_{i_1} \cdots t_{i_p}$ and $t_J = t_{j_{p+1}} \cdots t_{j_n}$. Apply to each tree in t_I a single non-empty cut. This may include the full cut (below the root). This produces a tensor product of forests

$$t'_I \otimes t''_I,$$

where $t'_I \neq \emptyset$. Next, define the set $\pi_1 := \{t'_I\}$. Then, define the forest $t''_I t_J$ and repeat the procedure to define successively the block π_2, π_3, up to π_m for $1 \leq m \leq |t_1 \cdots t_n|$.

Remark 7. *The computations of iterated coproducts of forests can be represented by matrices. Each forest $t_{\lambda_1} \cdots t_{\lambda_n}$ determines a partition of order n of the degree $N = \lambda_1 + \cdots + \lambda_n$, where $\lambda_i = |t_{\lambda_i}| > 0$. We denote it as the $1 \times n$ matrix $(\lambda_1, \ldots, \lambda_n)$. Now, to compute the lth order convolution product, we may construct certain matrices of size $l \times n$, where in column q we put a weak composition of length l, $(c_{1q}, \ldots, c_{lq})$, of $\lambda_q = c_{1q} + \cdots + c_{lq}$, $1 \leq q \leq n$, $c_{iq} \geq 0$, for $1 \leq i \leq l$*

$$\Lambda_{ln} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{l1} & \cdots & c_{ln} \end{pmatrix}.$$

These matrices are constraint as follows:

1. *The sum of the entries in each row must be bigger than zero, that is, for all $1 \leq k \leq l$ we must have $\sum_{p=1}^n c_{kp} > 0$.*
2. *For $1 \leq p \leq n$, we have $\sum_{r=1}^n c_{rp} = \lambda_p$.*

Item (1) reflects the fact that $(\tilde{\mu} - \epsilon)(\mathbf{1}) = 0$. The l order convolution product is of the form

$$(\tilde{\mu} - \epsilon) \prod_{i=1}^n t_{c_{1i}} \otimes (\tilde{\mu} - \epsilon) \prod_{i=1}^n t_{c_{2i}} \otimes \cdots \otimes (\tilde{\mu} - \epsilon) \prod_{i=1}^n t_{c_{li}}.$$

Here, $t_{c_{ij}}$ is the ladder tree of size $|t_{c_{ij}}| = c_{ij}$. Hence, the composition $(c_{1q}, \ldots, c_{lq})$ amounts to cutting the tree t_{λ_q} into l (possibly empty) trees $t_{c_{iq}}$ of size c_{iq}, $1 \leq i \leq l$. Translated to words, we find for $[1][2] \star [3]$

$$\Lambda_{12}^{(1)} = \begin{pmatrix} [1][2] & [3] \end{pmatrix} \quad \Lambda_{22}^{(1)} = \begin{pmatrix} [1][2] & 0 \\ 0 & [3] \end{pmatrix} \quad \Lambda_{22}^{(2)} = \begin{pmatrix} 0 & [3] \\ [1][2] & 0 \end{pmatrix}$$

$$\Lambda_{22}^{(3)} = \begin{pmatrix} [1] & [3] \\ [2] & 0 \end{pmatrix} \quad \Lambda_{22}^{(4)} = \begin{pmatrix} [1] & 0 \\ [2] & [3] \end{pmatrix}$$

$$\Lambda_{32}^{(1)} = \begin{pmatrix} [1] & 0 \\ [2] & 0 \\ 0 & [3] \end{pmatrix} \quad \Lambda_{32}^{(2)} = \begin{pmatrix} [1] & 0 \\ 0 & [3] \\ [2] & 0 \end{pmatrix} \quad \Lambda_{32}^{(3)} = \begin{pmatrix} 0 & [3] \\ [1] & 0 \\ [2] & 0 \end{pmatrix}.$$

Correspondingly, we have the iterated reduced coproducts:

$$[1][2] \star [3] + [1][2] \otimes [3] + [3] \otimes [1][2] + [1] \star [3] \otimes [2] + [1] \otimes [2] \star [3] + 3[1] \otimes [2] \otimes [3]$$

3.2 Chow's theorem

Recall the "classical" Chow theorem for the iterated-integrals signature

Theorem 8 ([5, Theorem 7.28])**.** *Every (finite dimensional projection of) a grouplike element (of the unshuffle coalgebra) can be realized as (the finite dimensional projection of) the iterated-integral signature of some piecewise smooth path X.*

Heuristically: "iterated-integral signatures fill the entire group" or respectively, "the logarithms of iterated-integral signatures fill the entire Lie algebra". It turns out that something analogous is *not* true for the iterated-sums signature. Indeed, let $x = (x_0, x_, \ldots, x_N) \in \mathbb{R}^{N+1}$ then one can calculate

$$\langle [1^2], \log \mathsf{DS}(x) \rangle = \sum_j (\delta x_j)^2 \geq 0.$$

Therefore, the image of the logarithm of iterated-sums signatures only reaches a certain subset of the Lie algebra. This raises several questions

- Does the problem persist if $x \in \mathbb{C}^{N+1}$? (The above problem evaporates in this setting, since $\sum_j (\delta x_j)^2$ can then reach any complex number.)
- In the real case: how to describe the subset of the Lie algebra that can be reached?
- Are there any implications, if any, to time series analysis?

References

[1] D. J. Berndt and J. Clifford. "Using Dynamic Time Warping to Find Patterns in Time Series". In: *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining*. AAAIWS'94. Seattle, WA: AAAI Press, 1994, pp. 359–370.

[2] P. Bonnier and H. Oberhauser. "Signature Cumulants, Ordered Partitions, and Independence of Stochastic Processes". In: *arXiv:1908.06496 [math.PR]* (2019).

[3] K.-T. Chen. "Iterated Integrals and Exponential Homomorphisms". In: *Proc. Lond. Math. Soc.* s3-4.1 (1954), pp. 502–512. DOI: 10.1112/plms/s3-4.1.502.

[4] J. Diehl, K. Ebrahimi-Fard, and N. Tapia. "Time-warping invariants of multidimensional time series". In: (2019). arXiv: 1906.05823 [math.RA].

[5] P. K. Friz and N. B. Victoir. *Multidimensional Stochastic Processes as Rough Paths*. Vol. 120. Cambridge studies in advanced mathematics. Cambridge University Press, 2010.

[6] J. G. Gaines. "The algebra of iterated stochastic integrals". In: *Stochastics and Stochastic Reports* 49.3-4 (1994), pp. 169–179.

[7] M. E. Hoffman. "Quasi-shuffle products". In: *J. Alg. Combin.* 11.1 (2000), pp. 49–68. ISSN: 0925-9899. DOI: 10.1023/a:1008791603281.

[8] M. E. Hoffman and K. Ihara. "Quasi-shuffle products revisited." English. In: *J. Algebra* 481 (2017), pp. 293–326. ISSN: 0021-8693.

[9] K. Luoto, S. Mykytiuk, and S. van Willigenburg. *An Introduction to Quasisymmetric Schur Functions*. SpringerBriefs in Mathematics. Springer New York, 2013.

[10] T. Lyons. "Differential equations driven by rough signals". In: *Rev. Matem. Iberoam.* (1998), pp. 215–310. DOI: 10.4171/rmi/240.

[11] C. Malvenuto and C. Reutenauer. "Duality between quasi-symmetric functions and the Solomon descent algebra". In: *J. Algebra* 177.3 (Nov. 1995), pp. 967–982.

[12] J.-C. Novelli and J.-Y. Thibon. "Free quasi-symmetric functions and descent algebras for wreath products, and noncommutative multi-symmetric functions". In: *Discrete Math.* 310.24 (2010), pp. 3584–3606.

[13] R. Ree. "Lie elements and an algebra associated with shuffles". English. In: *Ann. Math. (2)* 68.2 (1958), pp. 210–220. ISSN: 0003-486X; 1939-8980/e. DOI: 10.2307/1970243.

[14] C. Reutenauer. *Free Lie algebra*. New Series 7. Oxford Science Publications, Clarendon Press, 1993.

[15] D. Revuz and M. Yor. "Continuous Martingales and Brownian Motion". In: *Grundlehren der mathematischen Wissenschaften* (1999). ISSN: 0072-7830. DOI: 10.1007/978-3-662-06400-9.