Structure theorems for streamed information

https://www.datasig.ac.uk

Cris Salvi, Joscha Diehl, Terry Lyons, Rosa Preiß, and Jeremy Reizenstein

DataSig

A rough path between
mathematics and data science

Framing assumptions for this talk

- on some very fine scale, most instances of streamed information can be conceptually modelled as (rough) path segments $\gamma_u, u \in [s, t]$
- we are interested in such streams because of their potential effects and applications, in addition to their value at any given time
- Unparameterized path segments are a real-world data type

the signature of a path transforms a path segment γ in the Banach space V into a sequence of tensors within the algebra $T((V))$. This non-commutative exponential is obtained by evaluating the solution S_t of the controlled differential equation $dS_u = S_u \otimes d\gamma_u,\ S_s = 1$.
the signature of a path characterises a path up to a generalised re-parameterization (treelike equivalence) [3], [1]. We call the set of unparameterized paths $\Omega(V)$. The signatures of unparameterized paths form a subgroup G of the grouplike elements in $T((V))$. Each path has a unique tree reduced and unparameterised representative.

The fundamental questions

The signature representation filters out a symmetry made more powerful because we can answer the basic questions

- Understanding the spaces of functions on G.
- Give structure to the space of “polynomial” functions on G.

However, the signature is big, and does not directly offer a way of describing the data that is scalable and localisable. Logsignature - there is a hall basis, but the coefficients are global -change the basis elsewhere and you change the coefficients here.

Explicit algebraic bases

- Evaluating a function on a particular stream has a cost. After all it involves testing the stream γ in some way.
- Having evaluated $f(\gamma)$ and $g(\gamma)$ the cost of additionally evaluating $f(\gamma)g(\gamma)$ is a scalar multiplication without reference to the stream γ.
- Identifying an algebraic basis splits the process of evaluating a "polynomial" function on G into two parts:
  - evaluating the basis sensors on the underlying data γ and then
  - evaluating a unique polynomial function in these expensive but informative precomputed values
- Moreover if the bases are defined hierarchically using Hall sets, they can be computed recursively in a localised way (to compute one, one must compute its ancestors but not others).

The results in this work

- give a direct proof of [2]: polynomials in pure areas generate the shuffle algebra
- prove that polynomials in Hall integrals uniquely generate the shuffle algebra
- prove that polynomials in Hall areas uniquely generate the shuffle algebra

These results are fundamental structure theorems for streamed information. Polynomial functions split uniquely into two components -

- a first that engages with the underlying stream and which is potentially evaluated via some "physical" integration process that responds to the underlying signal
  - is intrinsically nasty operator
  - controlled differential equations are not closable in the uniform topology on γ - see [4]
- and a second that is a robust and continuous numerical polynomial evaluation made without reference to the stream.

Words as functions on paths

If A is an alphabet of d letters, then the tensor algebra is the space spanned by words

$$T((A)) = \mathbb{R} \oplus V\langle A\rangle \oplus V\langle A\rangle^{\otimes 2} \oplus \dots$$

Definition

Let $I \subset \mathbb{R}_+$ be a compact time interval and let $\gamma \in C^1(I, V)$ be a continuous path of bounded variation. For any sub-interval $[a,b] \in I$ define the signature $Sig(\gamma)_{[a,b]}$ of the path γ over $[a,b]$ as the following element of $T(V)$

$$\begin{aligned} Sig(\gamma|_{[a,b]}) &= 1 + \sum_{k=1}^{\infty} \int_{a<t_1<\dots<t_n<b} \dots \int d\gamma_{t_1} \otimes \dots \otimes d\gamma_{t_n} \\ w(\gamma|_{[a,b]}) &= \int_{a<t_1<\dots<t_n<b} \dots \int \langle w_1, d\gamma_{t_1} \rangle \otimes \dots \otimes \langle w_n, d\gamma_{t_n} \rangle \end{aligned}$$

Products and shuffles

These words span an algebra of real functions on path segments (the shuffle algebra) that contains the constants and separates paths (that are distinct modulo paramterisation)

$$\langle f \sqcup\!\sqcup g, x\rangle = \langle f, x\rangle\langle g, x\rangle \tag{1}$$

Definition

Let u and v be words in W_A then the shuffle product $u \sqcup\!\sqcup v$ is defined on $T(A)$ as follows

- $u \sqcup\!\sqcup e = e \sqcup\!\sqcup u = u$
- $au \sqcup\!\sqcup bv = a(u \sqcup\!\sqcup bv) + b(au \sqcup\!\sqcup v)$.

Functions on paths as real valued paths

Let $I \subset \mathbb{R}_+$ be a compact time interval and let $\gamma \in C^1(I, V)$ then we can define the historical path of γ in path space

$$u \in I \to \gamma|_{[a,u]}$$

$$w(\gamma)_u = w\left(\gamma|_{[a,u]}\right)$$

Empty word goes to the path that is identically 1. For all other words the real valued path begins at zero. We can think of words as *sensors*.

Combining sensors

- Words are sensors that convert the path into a real number evolving in time.
- A letter w simply projects the path increment onto that channel

$$w(\gamma)_v := \int_{u=a}^{u=v} \langle w, d\gamma_u \rangle$$

- Suppose that σ and τ are two elements of the shuffle algebra then new sensors are

$$\begin{aligned} (\sigma \prec \tau)()_v &:= \int_{u=\gamma^-}^{u=v} \tau()_u \, d\sigma()_u \\ \mathrm{area}_A(\sigma, \tau) &:= \sigma \prec \tau - \tau \prec \sigma \end{aligned}$$

- Remarkably both of these operation can be expressed in pure algebra as operations in the shuffle algebra.

The algebraic interpretation of integration

Definition

The left half shuffle $\prec_A : Sh(A) \times Sh(A) \to Sh(A)$ is a bi-linear form initially defined on basis elements: If u is the empty word e then $u \prec_A v := 0$. If u is nonempty, then let the letters in u be given by $u = u_1 ... u_n$ and define

$$u \prec_A v := u_1((u_2 \dots u_n) ⧢ v) \tag{2}$$

In particular, if u is a nonempty word then $u \prec_A e = u$.

Other trivial identities

- (FTC) Let $e \in Sh(A)$ be the empty word, then for any $x \in Sh(A)$

$$x \prec_A e = x - \langle x, e \rangle_A e \tag{3}$$

This is an algebraic version of the statement: $\int_0^{\cdot} 1\, df = f(.) - f(0)$.

- (Product Rule) It is easy to check from the definition of half shuffle that for any $x, y \in Sh(A)$ the following relation is satisfied

$$(x \sqcup\!\sqcup y) \prec_A e = x \prec_A y + y \prec_A x \tag{4}$$

- (Integration by parts) Rewriting the left hand side in the product rule gives that for any $x, y \in Sh(A)$

$$x \sqcup\!\sqcup y - \langle x, e \rangle_A \langle y, e \rangle_A e = x \prec_A y + y \prec_A x \tag{5}$$

- (Chain rule) It follows directly from the definition of $\prec_A$ that for any $x, y, z \in Sh(A)$

$$x \prec_A (y \sqcup\!\sqcup z) = (x \prec_A y) \prec_A z \tag{6}$$

- This is an algebraic version of the statement: $\int_0^{\cdot} fgdh = \int_0^{\cdot} fd(\int_0^{\cdot} gdh)$.

The Salvi identities

Lemma (Shuffle-pullout identity)

For any triple $x, y, z \in Sh(A)$ *the following relation is satisfied*

$$3\,\mathrm{area}_A(z, x \sqcup\!\sqcup y) = x \sqcup\!\sqcup \mathrm{area}_A(z,y) + y \sqcup\!\sqcup \mathrm{area}_A(z,x) - x \sqcup\!\sqcup y \sqcup\!\sqcup z + \mathrm{area}_A(\mathrm{area}_A(z,y),x) + \mathrm{area}_A(\mathrm{area}_A(z,x),y)$$

Lemma (Area-Jacobi identity)

For any triple $x, y, z \in Sh(A)$ *the following relation is satisfied*

$$\mathrm{area}_A(\mathrm{area}_A(x,y),z) + \mathrm{area}_A(\mathrm{area}_A(y,z),x) + \mathrm{area}_A(\mathrm{area}_A(z,x),y) = x \sqcup\!\sqcup \mathrm{area}_A(y,z) + y \sqcup\!\sqcup \mathrm{area}_A(z,x) + z \sqcup\!\sqcup \mathrm{area}_A(x,y)$$

The computational magma

Definition

$M(A)$ denotes the magma over A. $M(A)$ is the minimal set satisfying $\forall a \in A, (a) \in M(A)$ and if $\tau', \tau'' \in M(A)$ then $(\tau', \tau'') \in M(A)$.

- [5] $M(A)$ can be equivalently identified to the set of binary, planar, rooted trees with leaves labelled in A. For a given element $t \in M(A)$ we will refer to the ordered collection of letters appearing in its leaves as its foliage.
- $M(A)$ is free. Any map from A to a space with a product extends uniquely to a map from $M(A)$. (for example the foliage map, degree, multidegree).

Pure Foliage, Pure Areas and Pure Integrals

- $\text{area}_A(z, x)$, $z \prec x$ are product operators on the shuffle algebra defined on letters by the identity map.
  - An element x of $Sh(A)$ is a *pure area* if it is in the image under area_A of the magma. That is to say, there exists a tree $\tau \in M(A)$ so that $x = \text{area}_A(\tau)$.
  - An element x of $Sh(A)$ is a *pure integral* if it is in the image under $\prec_A$ of the magma. That is to say, there exists a tree $\tau \in M(A)$ so that $x = \prec_A(\tau)$.

Theorem

[2, Corollary 5.6] Any element in $Sh(A)$ can be written as a shuffle polynomial in pure areas $\{\text{area}_A(\tau) \mid \tau \in M(A)\}$.

Areas of shuffles

Theorem

For any $n \geq 1$ and any n pure areas $A_1, \ldots, A_n$ and an additional pure area A, the following relation holds

$$\operatorname{area}_A(A, A_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup A_n) = \beta_n A \sqcup\!\sqcup A_1 \sqcup\!\sqcup \ldots \sqcup\!\sqcup A_n + Q$$

where $\beta_n = -(n-1)/(n+1)$, and Q is a shuffle polynomial in pure areas of shuffle-degree at most n.

Polynomials in areas span

Lets write a word as a polynomial in Hall words. If $|w| > 0$, w can be written as follows

$$w = av = a \prec_A v$$

where $v \in W_A$ is of word of length $|v| = n - 1$ and $a \in A \subset S(A)$ is a letter. Moreover for any elements of $S(A)$

$$\begin{aligned} a \prec_A v &= \frac{1}{2}(\mathrm{area}_A(a, v) + a \sqcup\!\sqcup v - < a, e >< v, e > e) \\ &= \frac{1}{2}(\mathrm{area}_A(a, v) + a \sqcup\!\sqcup v) \end{aligned}$$

since a is a letter. The length of the word v is equal to $n - 1$, so by induction it can be written as a polynomial in pure areas of shuffle-degree $n - 1$. Hence, the term $a \sqcup\!\sqcup v$ is a shuffle polynomial in pure areas of shuffle-degree n. By the area of shuffles, the term $\mathrm{area}_A(a, v)$ is also a polynomial in pure areas of shuffle-degree n, and so w a polynomial in pure areas of shuffle-degree n.

Ancestral orders and Hall sets

Definition

A total order $<$ on a sub-magma M is an ancestral order if for any tree $t = (t', t'') \in M(A)$ of degree ≥ 2 one has $t < t''$.

Definition

A sub-magma H of $M(A)$ is a Hall set if

- $<$ is an ancestral order on H.
- $A \subset H$.
- For any tree $h = (h_1, h_2) \in M(A)$ of degree ≥ 2 one has $h \in H$ if and only if:
  - $h_1, h_2 \in H$ and $h_1 < h_2$
  - either $h_1 \in A$ or $h_2 \leq h_1''$ where $h_1 = (h_1', h_1'')$.

Hall sets exist!

Hall sets exist

Lemma

Hall sets exist, any ancestral order on the full magma leads in a canonical way to to a unique Hall set, and every Hall set can be obtained in this way.

Lemma

[5, corollary 4.14] The number of Hall trees of degree n and the dimension of the space of homogeneous Lie polynomials of degree n are equal to

$$\mathcal{D}_H = \frac{1}{n}\sum_{d|n}\mu(d)q^{n/d}$$

where μ is the Mobius function.

Hall Words, Hall areas, Hall Integrals

- An element x of $Sh(A)$ is a Hall area if it is in the image under area_A of the Hall set. That is to say, there exists a tree $h \in H \subset M(A)$ so that $x = \text{area}_A(h)$.
- An element x of $Sh(A)$ is a Hall integral if it is in the image under area_A of the Hall set. That is to say, there exists a tree $h \in H \subset M(A)$ so that $x = \prec_A(h)$.
- Hall words are defined similarly using the foliage map f

Decreasing sequences of hall words

Lemma

[5, Corollary 4.7] Every word $w \in W_A$ can be written uniquely as a decreasing product of Hall words

$$w = f(h_1)^{k_1} \ldots f(h_n)^{k_n}, \qquad h_i \in H, h_1 > \ldots > h_n$$

Definition

If $h \in H$ then because [] is a binary operator one defines $[h] \in T(A)$
Consider the collection of all decreasing sequences $h_i \in H$, $h_1 > \ldots > h_n$ then $\{[h_1]^{k_1} \ldots [h_n]^{k_n}\}$ are the PBW basis for $T(A)$.

A shuffle basis from hall integrals

Theorem

Consider all decreasing sequences $h_i \in H$, $h_1 > \ldots > h_n$, and strictly positive integers $k_i > 0$; then the elements

$$\frac{A_{h_1}^{k_1} \ldots A_{h_n}^{k_n}}{k_1! \ldots k_n!} (\prec_A (h_1))^{ш k_1} ш \ldots ш (\prec_A (h_n))^{ш k_n}$$

are the dual basis in $Sh(A)$ to the PBW basis $\{[h_1]^{k_1} \ldots [h_n]^{k_n}\}$ for $T(A)$. Every element of $Sh(A)$ is uniquely expressible as a shuffle polynomial in Hall integrals. A_{h_1} is the accumulated Lazard depth of h.

Definition

If $h = (xh''^k)$ is the Lazard decomposition of $h \in H$, where $x = (x', x'')$, $h'' \in H$ and $x'' \neq h''$, then we define the Lazard depth α_h of h to be $1/k$. The accumulated Lazard depth of h is defined recursively: $A_h = 1$ if $h \in A$, otherwise $h = (h', h'')$ and $A_h = \alpha_h A_{h'} A_{h''}$.

Any sensor is a polynomial in Hall Areas

DataSig

Definition

A shuffle-polynomial in Hall areas is a linear combination of terms of the form

$$\mathrm{area}_A(h_1) ⧢ \ldots ⧢ \mathrm{area}_A(h_n), \qquad h_i \in H. \tag{7}$$

Theorem

Any element in $Sh(A)$ can be written uniquely as a shuffle-polynomial in Hall areas $\{\mathrm{area}_A(h) \mid h \in H\}$.

Proof ideas

Definition

Given X we define

- - $M(X)$ the free magma
  - $T((X))$ the tensor algebra of infinite tensor series
  - $L(X)$ the free Lie sub-algebra of $T(X)$
  - W_X the space of words in the alphabet X and a canonical basis for $T(X)$
  - $Sh(X)$ the dual space to $T((X))$
  - W_X^* the dual basis to W_X
  - $\prec_X$, area_X, $\langle\ ,\ \rangle_X$ the various products on these spaces

Now consider a particular choice of X based on a lower central series decomposition of $L(A)$.

Elimination trick

- Let c be the greatest element of A with respect to the ancestral ordering $<$. Define the subset of trees

$$X = \{(ac^n), a \in A \setminus \{c\}, n \geq 0\} \subset M(A) \tag{8}$$

- A treacherous path - implications in Shuffles are contravariant.
- $L(X)$ is a Lie ideal and sub-algebra of co-dimension one in $L(A)$

Lemma

[5, Theorem 0.6] The Lie algebra $L(A)$ is the semi-direct product of $L(X)$ and $\mathbb{R}c$

$$L(A) = L(X) \ltimes \mathbb{R}c \tag{9}$$

The Main Result

Theorem

For any magma $M(A)$, ancestral ordering $<$, Hall set H on $M(A)$, and any Hall tree $h \in H$ there exists a unique collection of Hall trees $h_1, \ldots, h_n \in H$ with the same multidegree as h and scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that

$$\prec_A (h) = \sum_{i=1}^{n} \alpha_i \mathrm{area}_A(h_i) + P \tag{10}$$

where P is a shuffle polynomial in areas of Hall trees $s \in H$. Moreover each monomial in this sum is a (shuffle) product of two or more Hall areas and has a net A-multidegree equal to the A-multidegree of h; in particular every Hall tree defining the Hall areas in P has A-degree strictly less than the A-degree of h.

A shuffle basis - more abstraction

Lemma

Any element of $Sh(A)$ is a polynomial on the vector space that is the free Lie algebra $L(A)$

Proof

Recall that $L(A) \subset T(A)$ and that we can take exponentials to get the grouplike elements or signatures. $S \in T^n(A)$ is a truncated signature if and only if for some $l \in L^n(A)$ one has $S = \exp l$. If x is a word we may consider the coordinate iterated integral $\langle x, S \rangle$ as a real valued function on the signatures. Choose a basis l_i to $L(A)$ that is homogeneous. Then

$$\left\langle x, \exp \sum \lambda_i l_i \right\rangle$$

is, by expanding the exponential and stopping at an appropriate degree, a finite polynomial in λ_i. □

Linear Polynomials

If V is a vector space and x_i are an ordered basis for V^* then every polynomial $P(v)$ on V is a unique linear combination of decreasing products of products of $x_i(v)$. So it is tempting to think that the $\langle \prec_A (h), \exp \sum \lambda_i l_i \rangle$ are linear functions on the λ and that they are idependent. But they are not linear!

Example

Any polynomial function on $\mathbb{R}^2$can be uniquely written as a polynomial in $x,\ y+x^2$

Problem

What are the linear polynomials on $L(A)$? Define $M \subset Sh(A)$ by

$$M := \left\{ x \mid \forall l, l' \in L(A), \ \ \langle x, \exp(l+l') - (\exp(l) + \exp(l')) \rangle = 0 \right\}$$

At least $A \subset M$. Can we identify M, find a basis for M that is treelike, local.

Thank You

Horatio Boedihardjo, Xi Geng, Terry Lyons, and Danyu Yang.
The signature of a rough path: uniqueness.
Advances in Mathematics **293**, 720–737 (2016).

Joscha Diehl, Terry Lyons, Rosa Preiß, and Jeremy Reizenstein.
Areas of areas generate the shuffle algebra.
arXiv preprint arXiv:2002.02338 (2020).

Ben Hambly and Terry Lyons.
Uniqueness for the signature of a path of bounded variation and the reduced path group.
Annals of Mathematics pages 109–167 (2010).

Terry J Lyons.
Differential equations driven by rough signals.
Revista Matemática Iberoamericana **14**(2), 215–310 (1998).

C Reutenauer.
"Free Lie Algebras".
London Mathematical Society Monographs. Oxford Science Publications, The Clarendon Press, Oxford University Press (1993).