Iterated sums over semiring

J. Diehl (Universität Greifswald)

joint with K. Ebrahimi-Fard (NTNU), N. Tapia (TU Berlin)

March 9th, CIRM

Overview

Semirings

Reason for iterated-sums

Algebraic setting

Outlook

Start with the ring $\mathbb{R}$, with operations

$$x + y \qquad x \cdot y.$$

For $h > 0$ define

$$\begin{aligned} x \oplus_h y &:= h \log(e^{\frac{x}{h}} + e^{\frac{y}{h}}) \\ x \odot_h y &:= h \log(e^{\frac{x}{h}} \cdot e^{\frac{y}{h}}) = x + y. \end{aligned}$$

For $h \to 0$ this converges to (*Maslov dequantization*)

$$\begin{aligned} x \oplus y &:= \max\{x, y\} \\ x \odot y &:= x + y. \end{aligned}$$

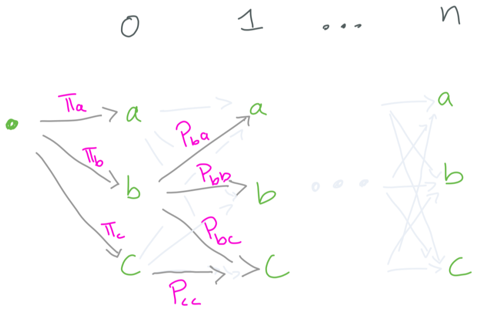
This does <u>not</u> have additive inverses anymore!

It is hence a semiring, the max-plus semiring.

A semiring (i.e. a ring without demand for additive inverse) pops up in many places.

Dynamic programming

Consider a time-homogeneous Markov chain $X_0, X_1, X_2, \ldots$ on states $\{a, b, c\}$.



A costly way to obtain the terminal distribution is

$$\mathbb{P}[X_n = a] = \sum_{w \in \{a,b,c\}^{n+1}, w_n = a} \pi_{w_0} p_{w_0 w_1} \ldots p_{w_{n-1} n} \quad \rightsquigarrow O(3^n)$$

Dynamic programming

There is, of course, a more economic way:

$$\mathbb{P}[X_n = a] = \mathbb{P}[X_{n-1} = a] \cdot p_{aa} + \mathbb{P}[X_{n-1} = b] \cdot p_{ca} \\ + \mathbb{P}[X_{n-1} = c] \cdot p_{ba}.$$

Iterating, one gets an $O(n)$ algorithm.

What if we are interested in the most probable path instead?

Dynamic programming

There is, of course, a more economic way:

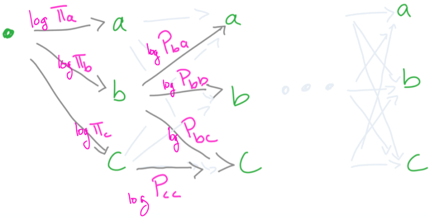
$$\mathbb{P}[X_n = a] = \mathbb{P}[X_{n-1} = a] \cdot p_{aa} + \mathbb{P}[X_{n-1} = b] \cdot p_{ca} \\ + \mathbb{P}[X_{n-1} = c] \cdot p_{ba}.$$

Iterating, one gets an $O(n)$ algorithm.

What if we are interested in the most probable path instead?

What if we are interested in the most probable path instead?

We just put the log-probabilities



and calculate in the max-plus semiring:

$$\max_{w \in \{a,b,c\}^{n+1}, w_n = a} \Big(\log \pi_{w_0} + \log p_{w_0 w_1} + \log p_{w_1 w_2} + \cdots \log p_{w_{n-1} w_n} \Big).$$

Dynamic programming still works!

Overview

Semirings

Reason for iterated-sums

Algebraic setting

Outlook

Convolutional Neural Networks

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 3 & 4 & 1 \\ 1 & 2 & 4 & \boxed{3} & 3 \\ 1 & 2 & 3 & 4 & 1 \\ 1 & 3 & 3 & 1 & 1 \\ 3 & 3 & 1 & 1 & 0 \end{pmatrix}$$

Why they work so well (probably ...)

1. Weight sharing.
2. Structure compatible with image data ("receptive field", approximate translation invariance).

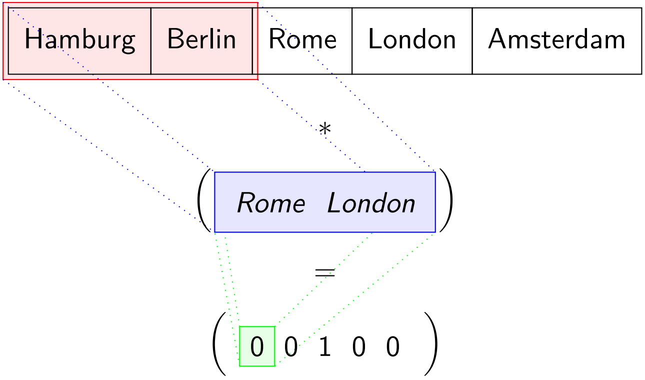
CNNs can, of course, be applied to sequential data.

$$\underset{I}{\begin{pmatrix} 0 & 1 & 1 & \boxed{1 \quad 0 \quad 3} & 0 \end{pmatrix}} * \underset{K}{\begin{pmatrix} \boxed{1 \quad 0 \quad 1} \end{pmatrix}} = \underset{I * K}{\begin{pmatrix} 1 & 2 & 1 & \boxed{4} & 0 \end{pmatrix}}$$

Does it make sense?

1. Weight sharing. ✓
2. Structure compatible with time-series data ?

Using a CNN to answer:
“Did a person visit Rome directly before visiting London?”



Using a CNN to answer:
“Did a person visit Rome directly before visiting London?”

Hamburg	Berlin	Rome	London	Amsterdam

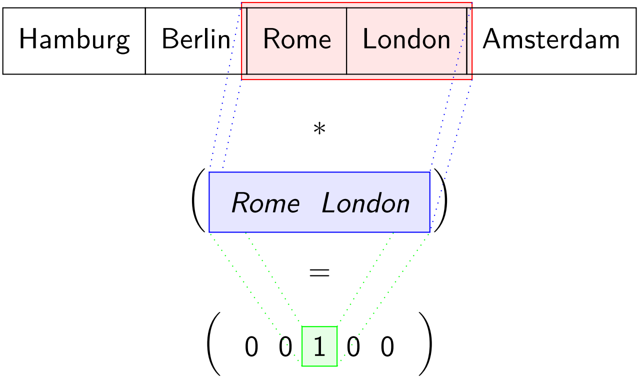
$$*$$

$$\begin{pmatrix} \boxed{\mathit{Rome} \quad \mathit{London}} \end{pmatrix}$$

$$=$$

$$\begin{pmatrix} 0 & \boxed{0} & 1 & 0 & 0 \end{pmatrix}$$

Using a CNN to answer:
“Did a person visit Rome directly before visiting London?”



But what if the person visits Rome some time before visiting London?

Hamburg	Rome	Berlin	Amsterdam	London

A (one-layer) CNN has difficulties detecting this (unless the kernel is large enough).

Chronological question:

“Did a person visit Rome some time before visiting London?”

Hamburg	Rome	Berlin	Amsterdam	London

$$*$$

$$\left(\boxed{\mathit{Rome} \quad \mathit{London}} \right)$$

$$=$$

$$\left(\boxed{0}\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0 \right)$$

Chronological question:

"Did a person visit Rome some time before visiting London?"

Hamburg	Rome	Berlin	Amsterdam	London

$$*$$

$$\left(\boxed{\mathit{Rome} \quad \mathit{London}}\right)$$

$$=$$

$$\begin{pmatrix} 0 & \boxed{0} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Chronological question:

"Did a person visit Rome some time before visiting London?"

Hamburg	Rome	Berlin	Amsterdam	London

$$*$$

$$\left(\boxed{\mathit{Rome} \quad \mathit{London}} \right)$$

$$=$$

$$\left(\begin{matrix} 0 & 0 & \boxed{0} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right)$$

Chronological question:

"Did a person visit Rome some time before visiting London?"

Hamburg	Rome	Berlin	Amsterdam	London

$$*$$

$$\left(\boxed{\mathit{Rome} \quad \mathit{London}} \right)$$

$$=$$

$$\left(\begin{matrix} 0 & 0 & 0 & \boxed{0} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \right)$$

Chronological question:

“Did a person visit Rome some time before visiting London?”

Hamburg	Rome	Berlin	Amsterdam	London

$$*$$

$$\left(\boxed{\textit{Rome} \quad \textit{London}} \right)$$

$$=$$

$$\begin{pmatrix} 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

More formal

Let

$$K : \text{Cities} \times \text{Cities} \to \{\texttt{true}, \texttt{false}\}$$

$$(\text{cityA}, \text{cityB}) \mapsto \left(\text{cityA} = \right.$$  $$\left.\right) \wedge \left(\text{cityB} = \right.$$  $$\left.\right)$$

$$\text{pool} : \{\texttt{true}, \texttt{false}\}^{\binom{n_{\text{in}}}{2}} \to \{\texttt{true}, \texttt{false}\}$$

$$z \mapsto z_1 \ \vee \ z_2 \ \vee \ \dots \ \vee \ z_{\binom{n_{\text{in}}}{2}}.$$

Then

$$\text{pool}\left(K(x_I) : I \in \binom{[n_{\text{in}}]}{2}\right) = \bigvee_{0 < i_1 < i_2 \leq n_{\text{in}}} \left(x_{i_1} = \right.$$  $$\left.\right) \wedge \left(x_{i_2} = \right.$$  $$\left.\right),$$

is $\texttt{true}$ <ins>if and only if</ins> Rome was visited some time before London.

More formal

Let

$$K : \text{Cities} \times \text{Cities} \to \{\texttt{true}, \texttt{false}\}$$

$(\text{cityA}, \text{cityB}) \mapsto (\text{cityA} =$  $) \wedge (\text{cityB} =$  $)$

$$\text{pool} : \{\texttt{true}, \texttt{false}\}^{\binom{n_{\text{in}}}{2}} \to \{\texttt{tru}$$

$$z \mapsto z_1 \vee$$

Chronological information

Then

$\text{pool}\left(K(x_I) : I \in \binom{[n_{\text{in}}]}{2}\right) = \bigvee_{0<i_1<i_2\leq n_{\text{in}}} (x_{i_1} =$  $) \wedge (x_{i_2} =$  $),$

is $\texttt{true}$ <u>if and only if</u> Rome was visited some time before London.

(There is nothing "learnable" here yet, we'll come to this later.)

First, we want to deal with a problem: $\binom{n_{\text{in}}}{2}$ gets large real quick !

(There is nothing "learnable" here yet, we'll come to this later.)

First, we want to deal with a problem: $\binom{n_{\text{in}}}{2}$ gets large real quick !

To clarify, let us do 3 cities whose ordered visit we want to detect:

$K(\dots) := (\text{cityA} =$  $) \wedge (\text{cityB} =$  $) \wedge (\text{cityC} =$  $)$

$$\operatorname{pool}\left(K(x_I) : I \in \binom{[n_{\text{in}}]}{3}\right) := \bigvee_{I \in \binom{[n_{\text{in}}]}{3}} K(x_I).$$

This needs $O(n_{\text{in}}^3)$ evaluations of K.

But! There is a better way.

$\displaystyle\bigvee_{I\in\binom{[n_{\text{in}}]}{3}} K(x_I) = \bigvee_{i_1<i_2<i_3} (x_{i_1} =$  $) \wedge (x_{i_2} =$  $) \wedge (x_{i_3} =$  $)$

$\displaystyle = \bigvee_{i_3} \Bigg(\bigvee_{i_1<i_2<i_3} (x_{i_1} =$  $) \wedge (x_{i_2} =$  $) \Bigg) \wedge (x_{i_3} =$  $)$

$\displaystyle =: \bigvee_{i_3} \mathrm{pool}'_{i_3} \wedge (x_{i_3} =$  $).$

Only n_{in} evaluations!

Further

$\mathrm{pool}'_{i_3} = \bigvee_{i_1<i_2<i_3} (x_{i_1} =$  $) \wedge (x_{i_2} =$  $)$

$= \bigvee_{i_2<i_3} \left(\bigvee_{i_1<i_2} (x_{i_1} =\right.$  $\left.)\right) \wedge (x_{i_2} =$  $)$

$=: \bigvee_{i_2<i_3} \mathrm{pool}''_{i_2} \wedge (x_{i_2} =$  $).$

Only n_{in} evaluations (to calculate all of $\mathrm{pool}'_\bullet$)!

Finally,

$\mathrm{pool}''_{i_2} = \bigvee_{i_1<i_2} (x_{i_1} =$  $)$

Only n_{in} evaluations (to calculate all of $\mathrm{pool}''_\bullet$)!

total amount of evaluations: $O(3n_{\mathrm{in}}) = O(n_{\mathrm{in}})$

What have we achieved?

We calulated

$$\operatorname{pool}\left(K(x_I) : I \in \binom{[n_{\text{in}}]}{3}\right)$$

$$= \bigvee_{I \in \binom{[n_{\text{in}}]}{3}} K(x_I)$$

$= \bigvee_{i_1<i_2<i_3} (x_{i_1} =$  $) \wedge (x_{i_2} =$  $) \wedge (x_{i_3} =$  $),$

which, on paper, costs $O(n_{\text{in}}^3)$, in only $O(n_{\text{in}})$ time !

What did we use?

- $\wedge$ distributes over $\vee$
- $\wedge$ and $\vee$ are associative

And that's it.

What have we achieved?

We calulated

$$\operatorname{pool}\left(K(x_I) : I \in \binom{[n_{\text{in}}]}{3}\right)$$

$$= \bigvee_{I \in \binom{[n_{\text{in}}]}{3}} K(x_I)$$

$= \bigvee_{i_1 < i_2 < i_3} (x_{i_1} =$  $) \wedge (x_{i_2} =$  $) \wedge (x_{i_3} =$  $),$

which, on paper, costs $O(n_{\text{in}}^3)$, in only $O(n_{\text{in}})$ time !

What did we use?

- $\wedge$ distributes over $\vee$
- $\wedge$ and $\vee$ are associative

And that's it.

Definition

The tuple $(\mathbb{S}, \oplus_{\mathbb{S}}, \odot_{\mathbb{S}}, \mathbf{0}_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}})$ is a commutative **semiring** if

- $(\mathbb{S}, \oplus_{\mathbb{S}}, \mathbf{0}_{\mathbb{S}})$ is a commutative monoid with unit $\mathbf{0}_{\mathbb{S}}$
- $(\mathbb{S}, \odot_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}})$ is a commutative monoid with unit $\mathbf{1}_{\mathbb{S}}$
- $\mathbf{0}_{\mathbb{S}} \odot_{\mathbb{S}} \mathbb{S} = \{\mathbf{0}_{\mathbb{S}}\}$
- multiplication distributes over addition, i.e.

$$a \odot_{\mathbb{S}} (b \oplus_{\mathbb{S}} c) = (a \odot_{\mathbb{S}} b) \oplus_{\mathbb{S}} (a \odot_{\mathbb{S}} c)$$

Examples of semirings

- any commutative ring
- boolean semiring $(\{\mathtt{false}, \mathtt{true}\}, \vee, \wedge, \mathtt{false}, \mathtt{true})$
- min-plus ("tropical") semiring $(\mathbb{R} \cup \{+\infty\}, \min, +, +\infty, 0)$
- possibilistic (or Viterbi or Bayesian) semiring $([0, 1], \max, \cdot, 0, 1)$

Definition

The tuple $(\mathbb{S}, \oplus_{\mathbb{S}}, \odot_{\mathbb{S}}, \mathbf{0}_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}})$ is a commutative **semiring** if

- $(\mathbb{S}, \oplus_{\mathbb{S}}, \mathbf{0}_{\mathbb{S}})$ is a commutative monoid with unit $\mathbf{0}_{\mathbb{S}}$
- $(\mathbb{S}, \odot_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}})$ is a commutative monoid with unit $\mathbf{1}_{\mathbb{S}}$
- $\mathbf{0}_{\mathbb{S}} \odot_{\mathbb{S}} \mathbb{S} = \{\mathbf{0}_{\mathbb{S}}\}$
- multiplication distributes over addition, i.e.

$$a \odot_{\mathbb{S}} (b \oplus_{\mathbb{S}} c) = (a \odot_{\mathbb{S}} b) \oplus_{\mathbb{S}} (a \odot_{\mathbb{S}} c)$$

Examples of semirings

- any commutative ring
- boolean semiring
  $(\{\mathtt{false}, \mathtt{true}\}, \vee, \wedge, \mathtt{false}, \mathtt{true})$
- min-plus ("tropical") semiring
  $(\mathbb{R} \cup \{+\infty\}, \min, +, +\infty, 0)$
- possibilistic (or Viterbi or Bayesian) semiring
  $([0, 1], \max, \cdot, 0, 1)$

Examples of semirings $(\mathbb{S}, \oplus_{\mathbb{S}}, \odot_{\mathbb{S}}, \mathbf{0}_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}})$

- semiring of subsets of a set M
  $(2^M, \cup, \cap, \emptyset, M)$
- any distributive lattice (with minimal and maximal element)
- ...

They are of huge interest in computer science / automata theory.

Corollary (DEFT '20)

Let $(\mathbb{S}, \oplus_{\mathbb{S}}, \odot_{\mathbb{S}}, \mathbf{0}_{\mathbb{S}}, \mathbf{1}_{\mathbb{S}})$ *be a commutative semiring. Then*

$$\operatorname{pool}\left(z_I : I \in \binom{[n_{\text{in}}]}{k}\right) := \bigoplus_{\mathbb{S}}{}_{i_1 < \cdots < i_k \leq n_{\text{in}}} z_{i_1}^{\odot_{\mathbb{S}} \alpha_1} \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} z_{i_k}^{\odot_{\mathbb{S}} \alpha_k},$$

is calculable in $O(n_{\text{in}})$*-time.*

Examples

- Over the ring $\mathbb{R}$

$$\sum_{i_1<\cdots<i_k} z_{i_1}^{\alpha_1}\dots z_{i_k}^{\alpha_k},$$

$\rightsquigarrow$ iterated-sums signature (quasisymmetric functions)

This has a long history.

- Graham '13 "Sparse arrays of signatures for ...".
- Lyons, Ni, Oberhauser '14 "A feature set for streams ..."
- various works by L Jin et al '15 on Chinese character recognition.
- Kiraly, Oberhauser '16 "Kernels for sequentially ordered data".
- Lyons, Oberhauser '17 "Sketching the order of events".
- D '13, D,Reizenstein '19 on invariant features.
- D,Ebrahimi-Fard,Tapia '19 "Time warping invariants".
- Kidger, Bonnier, Arribas, Salvi, Lyons '19 "Deep Signature Transforms".
- Toth, Bonnier, Oberhauser '20 "Seq2Tens".

In these works it progressively emerged that it is helpful to learn the signature-type features.

Paraphrasing

$$\rightsquigarrow \sum_{i_1 < \cdots < i_k} f_{\theta_1}(z_{i_1}) \cdot \cdots \cdot f_{\theta_k}(z_{i_k}).$$

with $f_\theta : \mathbb{R}^d \to \mathbb{R}$.

We propose to boil this down to the bare minimum needed, namely

distributivity and associativity,

to arrive at a richer set of features.

$$\rightsquigarrow \bigoplus_{i_1 < \cdots < i_k}{}_{\mathbb{S}} \, f_{\theta_1}(z_{i_1}) \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} f_{\theta_k}(z_{i_k}),$$

with $f_\theta : \mathbb{R}^d \to \mathbb{S}$.

In these works it progressively emerged that it is helpful to learn the signature-type features.

Paraphrasing

$$\rightsquigarrow \sum_{i_1 < \cdots < i_k} f_{\theta_1}(z_{i_1}) \cdot \cdots \cdot f_{\theta_k}(z_{i_k}).$$

with $f_\theta : \mathbb{R}^d \to \mathbb{R}$.

We propose to boil this down to the bare minimum needed, namely

distributivity and associativity,

to arrive at a richer set of features.

$$\rightsquigarrow \bigoplus_{i_1 < \cdots < i_k}{}_{\mathbb{S}} f_{\theta_1}(z_{i_1}) \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} f_{\theta_k}(z_{i_k}),$$

with $f_\theta : \mathbb{R}^d \to \mathbb{S}$.

Examples

- Over the tropical semiring

$$\min_{i_1 < \cdots < i_k} \{\alpha_1 \cdot z_{i_1} + \cdots + \alpha_k \cdot z_{i_k}\}$$

$\rightsquigarrow$ tropical-sums signature
(tropical quasisymmetric expressions [DEFT '20])

Leaving the strict setting of tropical-sums, we can do a learnable version of the visiting-cities example:

- Fix some embedding z_i of the visited cities in $\mathbb{R}^d$ (e.g. one-hot-encoding).
- Introduce parametrized functions $f_\theta : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$,

$$\rightsquigarrow \max_{i_1 < i_2} \left\{ f_{\theta_1}(z_{i_1}) + f_{\theta_2}(z_{i_2}) \right\},$$

and learn θ_1, θ_2.

Non-example

Not all type of sums work. For general nonlinear σ the sum

$$\sum_{i_1 < \cdots < i_k} \sigma(x_{i_1} + .. + x_{i_k}),$$

cannot be efficiently computed, since one can frame NP-complete problems in this form:

Subset sum problem: Given $x_1, \ldots, x_n \in \mathbb{Z}$ is there a subset which sums to 0?
Sub-problem: Is there a k-subset that sums to 0?

$$\sum_{i_1 < \cdots < i_k} 1_{\{0\}}(x_{i_1} + \cdots + x_{i_k}).$$

If this would only cost $O(k \cdot n)$ we would get an $O(n + 2n + \cdots + nn) = O(n^2)$ algorithm. ↯

Summary

- Expressions of the from

$$\mathrm{pool}\left(K(x_I) : I \subset \binom{n_{\mathrm{in}}}{k}\right)$$

extract meaningful, chronological information of time series. In this generality they are computationally untractable.

- Semirings provide a large class of examples that are tractable, namely

$$\bigoplus_{i_1<\cdots<i_k}^{\mathbb{S}} f_{\theta_1}(x_{i_1}) \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} f_{\theta_k}(x_{i_k}).$$

Semirings

Reason for iterated-sums

Algebraic setting

Outlook

Algebraic setting

For $z_1, z_2, \dots \in \mathbb{S}$, $s < t$, we define a collection of values in $\mathbb{S}$, indexed by words in the alphabet $\mathbb{N}$,

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle := \bigoplus_{s<i_1<\cdots<i_k<t+1}{}_{\mathbb{S}} \; z_{i_1}^{\odot_{\mathbb{S}} w_1} \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} z_{i_k}^{\odot_{\mathbb{S}} w_k}.$$

For example

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), 537 \right\rangle = \bigoplus_{s<i_1<\cdots<i_3<t+1}{}_{\mathbb{S}} \; z_{i_1}^{\odot_{\mathbb{S}} 5} \odot_{\mathbb{S}} z_{i_2}^{\odot_{\mathbb{S}} 3} \odot_{\mathbb{S}} z_{i_3}^{\odot_{\mathbb{S}} 7}$$

which in min-plus equals

$$\min_{s<i_1<i_2<i_3<t+1} \{5 \cdot z_{i_1} + 3 \cdot z_{i_2} + 7 \cdot z_{i_3}\}.$$

$\mathrm{ISS}^{\mathbb{S}}_{s,t}(z)$ is an element of $\mathbb{S}\langle\langle\mathbb{N}\rangle\rangle$, the space of formal, infinite sums of words (in the alphabet $\mathbb{N}$) with coefficients in $\mathbb{S}$:

$$\mathrm{ISS}^{\mathbb{S}}_{s,t}(z) = \sum_{w} c_w \; w,$$

with

$$c_w := \bigoplus_{s<i_1<\dots<i_k<t+1}^{\mathbb{S}} z_{i_1}^{\odot_{\mathbb{S}} w_1} \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} z_{i_k}^{\odot_{\mathbb{S}} w_k}.$$

Theorem (DEFT '20)

1. *(Quasi-shuffle identity)*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), u \right\rangle = \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \star u \right\rangle$$

2. *(Chen's identity) For* $s < t < u$,

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,u}(z), w \right\rangle = \bigoplus_{w' \cdot w'' = w}{}_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w' \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{t,u}(z), w'' \right\rangle$$

3. $\mathrm{ISS}^{\mathbb{S}}_{0,\infty}(z)$ *is invariant to inserting* $\mathbf{0}_{\mathbb{S}}$ *into* z.

Quasi-shuffle:

$32 \star 4 = 324 + 36 + 342 + 72 + 432$

Theorem (DEFT '20)

1. *(Quasi-shuffle identity)*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), u \right\rangle = \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \star u \right\rangle$$

2. *(Chen's identity) For $s < t < u$,*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,u}(z), w \right\rangle = \bigoplus_{w' \cdot w'' = w}{}_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w' \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{t,u}(z), w'' \right\rangle$$

3. $\mathrm{ISS}^{\mathbb{S}}_{0,\infty}(z)$ *is invariant to inserting* $\mathbf{0}_{\mathbb{S}}$ *into* z.

Quasi-shuffle:

$32 \star 4 = 324 + 36 + 342 + 72 + 432$

Theorem (DEFT '20)

1. *(Quasi-shuffle identity)*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), u \right\rangle = \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \star u \right\rangle$$

2. *(Chen's identity) For $s < t < u$,*

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,u}(z), w \right\rangle = \bigoplus_{w' \cdot w'' = w}{}_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w' \right\rangle \odot_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}_{t,u}(z), w'' \right\rangle$$

3. $\mathrm{ISS}^{\mathbb{S}}_{0,\infty}(z)$ *is invariant to inserting* $\mathbf{0}_{\mathbb{S}}$ *into* z.

Concatenation:

$32 \cdot 4 = 324$

Quasisymmetric functions

Using formal variables $Z_1, Z_2, \ldots,$ the expressions

$$\bigoplus_{s<i_1<\cdots<i_k<t+1} Z_{i_1}^{\odot_{\mathbb{S}} w_1} \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} Z_{i_k}^{\odot_{\mathbb{S}} w_k}$$

are **quasisymmetric expressions**.

This is the monomial basis.

Over a ring there are many bases (monomial, fundamental, ..).
This does not work over a semiring (there is no additive inverse).

In the monomial basis, the product is given by the quasi-shuffle.

Summary

- In the special case of monomial f, we are led to the iterated-sums signature over a semiring

$$\left\langle \mathrm{ISS}^{\mathbb{S}}_{s,t}(z), w \right\rangle = \bigoplus_{s<i_1<\cdots<i_k<t+1}^{\mathbb{S}} z_{i_1}^{\odot_{\mathbb{S}} w_1} \odot_{\mathbb{S}} \cdots \odot_{\mathbb{S}} z_{i_k}^{\odot_{\mathbb{S}} w_k}.$$

- This is the evaluation of quasisymmetric function expressions on the time series. Almost all properties of the classical setting survive (they mostly depend on the structure of the index set ..).

Semirings

Reason for iterated-sums

Algebraic setting

Outlook

- Log signature
- Multidimensional time series
- Controlled systems
- Dynamic programming

Log signature

There is no log signature, since there is no minus.

More concretely, over the tropical semiring

$$\left(\left\langle \mathrm{ISS}^{\mathbb{S}}(z), 1 \right\rangle\right)^{\odot_{\mathbb{S}} 2} = \left\langle \mathrm{ISS}^{\mathbb{S}}(z), 11 \right\rangle \oplus_{\mathbb{S}} \left\langle \mathrm{ISS}^{\mathbb{S}}(z), 2 \right\rangle.$$

But, knowing both

$$\left(\left\langle \mathrm{ISS}^{\mathbb{S}}(z), 1 \right\rangle\right)^{\odot_{\mathbb{S}} 2} = 2 \min_i z_i$$
$$\left\langle \mathrm{ISS}^{\mathbb{S}}(z), 2 \right\rangle = 2 \cdot \min_i z_i,$$

we can clearly not deduce the value of

$$\left\langle \mathrm{ISS}^{\mathbb{S}}(z), 11 \right\rangle = \min_{i_1 < i_2} \{ z_{i_1} + z_{i_2} \}.$$

Q: How to extract the "minimal" information contained in the signature?

Multidimensional time series

Multidimensional time series can be treated as usual, by projecting the time series to coordinates before calculating the iterated-sums.
In the semiring setting a more interesting approach seems possible, by considering a time series as taking values in a larger semiring.

One example is via the map

$$\begin{aligned} \mathbb{R}^d &\to \text{ bounded convex polytopes} \\ x &\mapsto \{x\}. \end{aligned}$$

The resulting time series can then be considered in the semiring of polytopes (compare *Borinsky - 2020 - Tropical Monte Carlo quadrature for Feynman integrals*).

Q: In what semirings to embed a time series?

Controlled systems

The iterated-integrals signature has close relation to controlled ODEs, and iterated-sums over a ring appear in discretized dynamic systems.

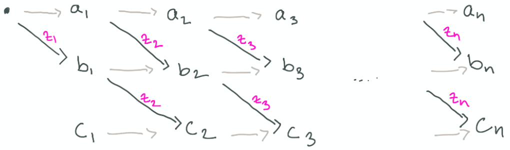
There is a vast literature on discrete control over semirings.

> ***Q: Is there a relation of the*** $\mathrm{ISS}^{\mathbb{S}}$ ***to discrete control theory in a semiring?***

Dynamic programming

We can embed the iterated-sums in such a framework:

Let $z_1, .., z_n$ be a time series in a semiring $\mathbb{S}$. Consider



where all horizontal edges have weight $\mathbf{1}_{\mathbb{S}}$.

$$W(m) := \text{the sum of weight of all paths from 0 to } m.$$

Then

$$
\begin{aligned}
W(c_2) &= z_1 \odot_{\mathbb{S}} z_2 \\
W(c_3) &= z_1 \odot_{\mathbb{S}} z_2 \odot_{\mathbb{S}} \mathbf{1}_{\mathbb{S}} \oplus_s z_1 \odot_{\mathbb{S}} \mathbf{1}_{\mathbb{S}} \odot_{\mathbb{S}} z_3 \oplus_{\mathbb{S}} \mathbf{1}_{\mathbb{S}} \odot_{\mathbb{S}} z_2 \odot_{\mathbb{S}} z_3 \\
&= z_1 \odot_{\mathbb{S}} z_2 \oplus_s z_1 \odot_{\mathbb{S}} z_3 \oplus_{\mathbb{S}} z_2 \odot_{\mathbb{S}} z_3 \\
W(c_n) &= \left\langle \mathrm{ISS}^{\mathbb{S}}(z), 11 \right\rangle .
\end{aligned}
$$

Dynamic programming

Q: Is there a deeper connection to dynamic programming?

Research assistant (3 years) with N. Sugiura (JAMSTEC), Japan.

"We are looking for an **applied mathematician** who is willing to research **real-world applications**, or a geophysicist who is willing to incorporate new mathematical theories. She/he is expected to enhance geophysical research through novel data-driven approaches. Specifically, she/he will be responsible for the **data analysis**, as well as relevant theories, of geophysical data that appear as vertical profiles and time series, by assembling and coding machine learning algorithms with the use of the **path signature**."

Please go to `https://diehlj.github.io` for the link.

Deadline: **March, 22th**.

Thank you!