Example signature path integral calculations

Jeremy Reizenstein, Centre for Complexity Science*
University of Warwick

October 2015–2018

1 Introduction

The iterated integral signature of a continuous path is a sequence of numbers which characterises it enough for the purpose of solving certain differential equations driven by the path. Here I give some concrete examples of the most basic calculations, in the manner of the introduction to [3]. The aim here is to present examples of calculations on which the theory of rough paths is based without worrying about convergence. I've settled on a slightly colourful notation and try to present things more laboriously than some other places. I'm grateful to Joscha Diehl, Jiawei Chang, Sina Nejad and Terry Lyons for advice on this stuff.

A path in $\mathbb{R}^d$ can be described by a continuous map $\gamma : [a, b] \to \mathbb{R}^d$ with $\gamma(t) = (\gamma_{\mathbf{1}}(t), \gamma_{\mathbf{2}}(t), \ldots, \gamma_{\mathbf{d}}(t))$. Its signature is a function from words written in the alphabet $\{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\}$ to $\mathbb{R}$, denoted by $X^\gamma_{a,b}$. The **signature** is defined inductively on the length of the word, where the length $|w|$ of a word w is the number of letters it contains. The signature of the empty word ϵ is defined as 1. If w is a word and i is a letter from the alphabet then $X^\gamma_{a,b}(wi)$ is defined as the Stieltjes integral $\int_a^b X^\gamma_{a,t}(w)\, d\gamma_i(t)$. In the simple case that γ_i is differentiable, this is equal to $\int_a^b X^\gamma_{a,t}\gamma_i'(t)\, dt$. The restriction of the signature to words of length m is called the mth **level** of the signature. It contains d^m values.

The signature is made up of a series of levels, one for each nonnegative integer. Level m takes values in $(\mathbb{R}^d)^{\otimes m}$, which is a d^m-dimensional real vector space, and consists of the d^m values of integrals of the form

$$\int_a^b \int_a^{t_1} \cdots \int_a^{t_{m-2}} \int_a^{t_{m-1}} d\gamma_{i_1}(t_m)\, d\gamma_{i_2}(t_{m-1})\, \ldots\, d\gamma_{i_{m-1}}(t_2)\, d\gamma_{i_m}(t_1), \tag{1}$$

where each i_j is allowed to range over values in $\{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\}$.

2 Taylor's formula in words

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a smooth function.

*Supported by the Engineering and Physical Sciences Research Council

We might want to expand f near the point a. Let $r(t) = a + tx$.

$$\begin{aligned}
f(a+x) &= f(a) + \int_0^1 \nabla f(r(t_1)) \cdot r'(t_1)\, dt_1 \\
&= f(a) + \int_0^1 \nabla f(a + t_1 x) \cdot (x)\, dt_1 \\
&= f(a) + \int_0^1 \sum_{i_1=\mathbf{1}}^{\mathbf{d}} f^{(i_1)}(a + t_1 x))(x_{i_1})\, dt_1 \\
&= f(a) + \sum_{i_1=\mathbf{1}}^{\mathbf{d}} x_{i_1} \int_0^1 f^{(i_1)}(a + t_1 x)\, dt_1 \\
&= f(a) + \sum_{i_1=\mathbf{1}}^{\mathbf{d}} x_{i_1} \int_0^1 \left[f^{(i_1)}(a) + \sum_{i_2=\mathbf{1}}^{\mathbf{d}} t_1 x_{i_2} \int_0^1 f^{(i_1,i_2)}(a + t_1 t_2 x)\, dt_2 \right] dt_1 \\
&= f(a) + \sum_{i_1=\mathbf{1}}^{\mathbf{d}} x_{i_1} \int_0^1 \left[f^{(i_1)}(a) + \sum_{i_2=\mathbf{1}}^{\mathbf{d}} t_1 x_{i_2} \int_0^1 \left\{ f^{(i_1,i_2)}(a) + \cdots \right\} dt_2 \right] dt_1 \\
&\approx f(a) + \sum_{i_1=\mathbf{1}}^{\mathbf{d}} x_{i_1} f^{(i_1)}(a) + \sum_{i_1=\mathbf{1}}^{\mathbf{d}} \sum_{i_2=\mathbf{1}}^{\mathbf{d}} \tfrac{1}{2} x_{i_1} x_{i_2} f^{(i_1,i_2)}(a) + \cdots
\end{aligned}$$

A multi-index α is a d-tuple $(\alpha_{\mathbf{1}}, \ldots, \alpha_{\mathbf{d}})$ of non negative integers. Let the coordinates be $x_{\mathbf{1}}, x_{\mathbf{2}}, \ldots, x_{\mathbf{d}}$ and define $|\alpha| := \alpha_{\mathbf{1}} + \cdots + \alpha_{\mathbf{d}}$, $\alpha! := \alpha_{\mathbf{1}}! \ldots \alpha_{\mathbf{d}}!$, $x^\alpha := x_{\mathbf{1}}^{\alpha_{\mathbf{1}}} \ldots x_{\mathbf{d}}^{\alpha_{\mathbf{d}}}$, and $D^\alpha f$ as the function $\frac{\partial^{|\alpha|} f}{\partial x_{\mathbf{1}}^{\alpha_{\mathbf{1}}} \ldots \partial x_{\mathbf{d}}^{\alpha_{\mathbf{d}}}}$ from $\mathbb{R}^d$ to $\mathbb{R}$. With these definitions, Taylor's formula around $a \in \mathbb{R}^d$ expresses f as a sum over multi-indices as follows.

$$f(x) = \sum_\alpha \frac{D^\alpha f(a)}{\alpha!} (x-a)^\alpha \tag{2}$$

There is a simple map κ from words on $\{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\}$ to multi-indices which returns the number of occurrences of each letter. For example, if $d = 4$ then $\kappa(\mathbf{312221}) = (2, 3, 1, 0)$. It has the property that $|w| = |\kappa(w)|$. By the multinomial theorem, the number of words which map to a given multi-index α is $\frac{|\alpha|!}{\alpha!}$. Taylor's formula as a sum over words looks like this:

$$f(x) = \sum_w \frac{1}{\frac{|\kappa(w)|!}{\kappa(w)!}} \frac{D^{\kappa(w)} f(a)}{\kappa(w)!} (x-a)^{\kappa(w)} = \sum_w \frac{D^{\kappa(w)} f(a)}{|w|!} (x-a)^{\kappa(w)} \tag{3}$$

Of course, Taylor series don't always have to converge.

3 Shuffle product

Given two words w_1 and w_2, their shuffle product $w_1 \sqcup\!\sqcup\, w_2$ is the multiset of words which can be formed by interleaving them, including multiplicity, which we write as a linear combination of words[1]. For example

$$\begin{aligned}
\mathbf{12} \sqcup\!\sqcup\, \epsilon &= \mathbf{12} \\
\mathbf{12} \sqcup\!\sqcup\, \mathbf{3} &= \mathbf{312} + \mathbf{132} + \mathbf{123} \\
\mathbf{12} \sqcup\!\sqcup\, \mathbf{12} &= 2\mathbf{1212} + 4\mathbf{1122} \\
\mathbf{12} \sqcup\!\sqcup\, \mathbf{13} &= 2\mathbf{1123} + 2\mathbf{1132} + \mathbf{1213} + \mathbf{1312} \\
\mathbf{1}^i \sqcup\!\sqcup\, \mathbf{1}^j\mathbf{2} &= \textstyle\sum_{k=j}^{i+j} \binom{k}{j} \mathbf{1}^k\mathbf{21}^{i+j-k}
\end{aligned}$$

The shuffle product between linear combinations of words is the extension of this linearly in its arguments. It is a commutative and associative operation. For example

$$\mathbf{12} \sqcup\!\sqcup\, (\mathbf{1} + 3\mathbf{2}) = 2\mathbf{112} + \mathbf{121} + 3\mathbf{212} + 6\mathbf{122}$$

The signature of a bounded variation path obeys the following relation, which is a complicated consequence of integration by parts or the product rule for differentiation.

$$X^\gamma_{a,b}(w_1)X^\gamma_{a,b}(w_2) = X^\gamma_{a,b}(w_1 \sqcup\!\sqcup\, w_2) \tag{4}$$

For example

$$\begin{aligned}
X^\gamma_{a,b}(\mathbf{1})X^\gamma_{a,b}(\mathbf{1}) &= \left(\int_a^b \gamma_{\mathbf{1}}'(t)\,dt\right)\left(\int_a^b \gamma_{\mathbf{1}}'(t)\,dt\right) \\
&= (\gamma_{\mathbf{1}}(b) - \gamma_{\mathbf{1}}(a))^2 \\
&= \int_a^b \frac{d}{dt}\left(\gamma_{\mathbf{1}}(t) - \gamma_{\mathbf{1}}(a)\right)^2\,dt \\
&= \int_a^b 2\left(\gamma_{\mathbf{1}}(t) - \gamma_{\mathbf{1}}(a)\right)\gamma_{\mathbf{1}}'(t)\,dt \\
&= 2\int_a^b \left(\int_a^t \gamma_{\mathbf{1}}'(t')\,dt'\right)\gamma_{\mathbf{1}}'(t)\,dt \\
&= 2X^\gamma_{a,b}(\mathbf{11}) = X^\gamma_{a,b}(2\mathbf{11}) = X^\gamma_{a,b}(\mathbf{1} \sqcup\!\sqcup\, \mathbf{1})
\end{aligned}$$

$$\begin{aligned}
X^\gamma_{a,b}(\mathbf{1})X^\gamma_{a,b}(\mathbf{2}) &= \left(\int_a^b \gamma_{\mathbf{1}}'(t)\,dt\right)\left(\int_a^b \gamma_{\mathbf{2}}'(t)\,dt\right) \\
&= (\gamma_{\mathbf{1}}(b) - \gamma_{\mathbf{1}}(a))(\gamma_{\mathbf{2}}(b) - \gamma_{\mathbf{2}}(a)) \\
&= \int_a^b \frac{d}{dt}\left[(\gamma_{\mathbf{1}}(t) - \gamma_{\mathbf{1}}(a))(\gamma_{\mathbf{2}}(t) - \gamma_{\mathbf{2}}(a))\right]\,dt \\
&= \int_a^b \left(\int_a^t \gamma_{\mathbf{1}}'(t')\,dt'\right)\gamma_{\mathbf{2}}'(t)\,dt + \int_a^b \left(\int_a^t \gamma_{\mathbf{2}}'(t')\,dt'\right)\gamma_{\mathbf{1}}'(t)\,dt \\
&= X^\gamma_{a,b}(\mathbf{12} + \mathbf{21}) = X^\gamma_{a,b}(\mathbf{1} \sqcup\!\sqcup\, \mathbf{2})
\end{aligned}$$

More generally in rough path theory, the relation 4 is the characteristic of a *geometric* rough path. We define $\sqcup\!\sqcup(\alpha)$ to be the shuffle between all the letters in the multi-index α, and $E(w)$ to be $\sqcup\!\sqcup(\kappa(w))$. For example

$$\sqcup\!\sqcup((1,3)) = E(\mathbf{1222}) = \mathbf{1} \sqcup\!\sqcup\, \mathbf{2} \sqcup\!\sqcup\, \mathbf{2} \sqcup\!\sqcup\, \mathbf{2} = 6\mathbf{1222} + 6\mathbf{2122} + 6\mathbf{2212} + 6\mathbf{2221}$$

It will then be the case that $(\gamma(b) - \gamma(a))^\alpha = X^\gamma_{a,b}(\sqcup\!\sqcup(\alpha))$.

[1] which is the same as a (non-commutative) polynomial on letters

4 Going from paths to signatures

Here I give an example of how a formula giving a path y in terms of the signature of a path x might give us a formula for the *signature* of y in terms of x as well. Let some path $y : [a, b] \to \mathbb{R}^c$ be given in terms of the signature of some path $x : [a, b] \to \mathbb{R}^d$ in the following way, which is just an arbitrary linear way.

$$y_i(t) = y_i(a) + \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_{v \in I} z_i(j, v) X^x_{a,t}(w_i(j, v)j)$$

Here I is some index set, and for each i in $\{1, \dots, c\}$, z_i is a function from $\{\mathbf{1}, \mathbf{2}, \dots, \mathbf{d}\} \times I$ to $\mathbb{R}$ and w_i is a function from $\{\mathbf{1}, \mathbf{2}, \dots, \mathbf{d}\} \times I$ to words on $\{\mathbf{1}, \mathbf{2}, \dots, \mathbf{d}\}$. Then an element of the signature of y at level 2 is

$$\begin{aligned}
X^y_{a,t}(hi) &= \int_a^t \int_a^{t'} dy_h(t')\, dy_i(t) = \int_a^t [y_h(t') - y_h(a)]\, dy_i(t) \\
&= \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_{v \in I} \sum_{j'=\mathbf{1}}^{\mathbf{d}} \sum_{v' \in I} z_h(j, v) z_i(j', v') \int_a^t \left[X^x_{a,t'}(w_h(j, v)j)\right] d\left(X^x_{a,t'}(w_i(j', v')j')\right) \\
&= \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_{v \in I} \sum_{j'=\mathbf{1}}^{\mathbf{d}} \sum_{v' \in I} z_h(j, v) z_i(j', v') \int_a^t \left[X^x_{a,t'}(w_h(j, v)j)\right] d\left(\int_a^{t'} X^x_{a,t''}(w_i(j', v') dx_{j'}(t'')\right) \\
&= \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_{v \in I} \sum_{j'=\mathbf{1}}^{\mathbf{d}} \sum_{v' \in I} z_h(j, v) z_i(j', v') \int_a^t \left[X^x_{a,t'}(w_h(j, v)j)\right] X^x_{a,t'}(w_i(j', v'))\, dx_{j'} \\
&= \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_{v \in I} \sum_{j'=\mathbf{1}}^{\mathbf{d}} \sum_{v' \in I} z_h(j, v) z_i(j', v') X^x_{a,t}\left([w_h(j, v)j \sqcup\!\sqcup\, w_i(j', v')]j'\right) \qquad (5)
\end{aligned}$$

As an aside, the expression $[w_h(j, v)j \sqcup\!\sqcup\, w_i(j', v')]j'$ is the right half shuffle of $w_h(j, v)j$ and $w_i(j', v')j'$. We thus have an example of how the right half shuffle of two signature elements has something to do with the area between them.

As a sanity check, in the special case where we want y the same as x, we could set $I = \{0\}$, $z_i(j, 0) = \delta_{ij}$, and $w_i(j, 0)$ to always be ϵ. Then (5) reads $X^y_{a,t}(hi) = X^x_{a,t}([h \sqcup\!\sqcup\, \epsilon)i) = X^x_{a,t}(hi)$.

5 Integrating a form, or $dy = f(x)\,dx$

This is an attempt to write out an explicit form of a calculation found in many places (starting in section 3.2 of [2]). It is written out explicitly in vector form in [5].

If $x : [a,b] \to \mathbb{R}^d$ is a smooth path and f is a smooth function from $\mathbb{R}^d$ to $c \times d$ matrices, then the equation $dy = f(x)\,dx$, together with the starting point $y(a) = 0$, defines a path $y : [a,b] \to \mathbb{R}^c$. Expanded, it is a system of c differential equations: $y_i'(t) = \sum_{j=\mathbf{1}}^{\mathbf{d}} f_{ij}(x(t))\,x_j'(t)$ for each i in $\{1, \ldots, c\}$. Each y_i is determined by one equation.

We can find $y_i(b)$, in terms of the signature of x, by expanding f_{ij} for each j using the multivariate Taylor's theorem (ignoring the fact that the Taylor series may not converge nicely).

$$\begin{aligned}
y_i(b) = X^y_{a,b}(i) &= \int_a^b y_i'(t)\,dt = \int_a^b \sum_{j=\mathbf{1}}^{\mathbf{d}} f_{ij}(x(t))\,x_j'(t)\,dt \\
&= \int_a^b \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \frac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} (x(t) - x(a))^{\kappa(w)} x_j'(t)\,dt \\
&= \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \frac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} \int_a^b (x(t) - x(a))^{\kappa(w)} x_j'(t)\,dt \\
&= \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \frac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} \int_a^b X^x_{a,t}\big(E(w)\big)\,x_j'(t)\,dt \\
&= \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \frac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} X^x_{a,b}\big(E(w)j\big)
\end{aligned} \tag{6}$$

where w ranges over words on $\{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\}$ and the concatenation product is indicated at the end. To find an element of the signature of the path y from the same data, we can proceed in a similar way. For a word with two letters,

$$\begin{aligned}
X^y_{a,b}(hi) &= \int_a^b \left[\int_a^t y_h'(t')\,dt'\right] y_i'(t)\,dt \\
&= \int_a^b \left[\sum_{k=\mathbf{1}}^{\mathbf{d}} \sum_v \tfrac{D^{\kappa(v)} f_{hk}(x(a))}{|v|!} X^x_{a,t}\big(E(v)k\big)\right] \sum_{j=\mathbf{1}}^{\mathbf{d}} f_{ij}(x(t))\,x_j'(t)\,dt \\
&= \sum_{k=\mathbf{1}}^{\mathbf{d}} \sum_v \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \tfrac{D^{\kappa(v)} f_{hk}(x(a))}{|v|!} \tfrac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} \int_a^b X^x_{a,t}\big(E(v)k\big)(x(t) - x(a))^{\kappa(w)}\,x_j'(t)\,dt \\
&= \sum_{k=\mathbf{1}}^{\mathbf{d}} \sum_v \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \tfrac{D^{\kappa(v)} f_{hk}(x(a))}{|v|!} \tfrac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} \int_a^b X^x_{a,t}\big(E(v)k\big) X^x_{a,t}\big(E(w)\big)\,x_j'(t)\,dt \\
&= \sum_{k=\mathbf{1}}^{\mathbf{d}} \sum_v \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \tfrac{D^{\kappa(v)} f_{hk}(x(a))}{|v|!} \tfrac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} \int_a^b X^x_{a,t}\big(E(v)k \sqcup\!\sqcup E(w)\big)\,x_j'(t)\,dt \\
&= \sum_{k=\mathbf{1}}^{\mathbf{d}} \sum_v \sum_{j=\mathbf{1}}^{\mathbf{d}} \sum_w \frac{D^{\kappa(v)} f_{hk}(x(a))}{|v|!} \frac{D^{\kappa(w)} f_{ij}(x(a))}{|w|!} X^x_{a,b}\big(\big[E(v)k \sqcup\!\sqcup E(w)\big]j\big)
\end{aligned} \tag{7}$$

where both w and v range over words on $\{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\}$. The pattern will continue for longer words. This can also be got with Equation (5).

5.1 Example: constant matrix

If f is a function which returns a constant matrix, then the only nonzero terms in the sums over words will be the one corresponding to the empty word. We recover the fact that if we transform a path by a matrix A then we transform level m of its signature by the tensor power $A^{\otimes m}$.

An example

Consider a path $x : [0,1] \to \mathbb{R}^2$ whose first dimension is time, so that $x_{\mathbf{1}}(t) = t$, and with $x_{\mathbf{2}}(0) = 0$. Under such a condition, the path can contain no backtracking, so the signature identifies the path completely. Here we use the above calculations to try to find an $x_{\mathbf{2}}(p)$, an intermediate point on the path, from the signature of the whole path.

For this, we make f be such that the path y moves like $x_{\mathbf{2}}$ when the time is much less than p, and moves hardly at all when the time is greater than p.

So it is natural to pick a positive a and set $f_{\mathit{1}\mathbf{1}}(x) = 0$ and something like $f_{\mathit{1}\mathbf{2}}(x) = 1 - \sigma(x_{\mathbf{1}} - p; a)$. The idea is that f uses a sharper sigmoid, and more like the ideal step function, as a increases.

$$
\begin{aligned}
x_{\mathbf{2}}(p) \approx y_{\mathit{1}}(1) &= \sum_{j=\mathbf{1}}^{\mathbf{2}} \sum_{w} \frac{D^{\kappa(w)} f_{\mathit{1}j}(x(0))}{|w|!} X^x_{0,1}\big(E(w)j\big) \\
&= \sum_{w} \frac{D^{\kappa(w)} f_{\mathit{1}\mathbf{2}}(x(0))}{|w|!} X^x_{0,1}\big(E(w)\mathbf{2}\big) \\
&= \sum_{k=0}^{\infty} \frac{D^{\kappa(\mathbf{1}^k)} f_{\mathit{1}\mathbf{2}}(x(0))}{|\mathbf{1}^k|!} X^x_{0,1}\big(E(\mathbf{1}^k)\mathbf{2}\big) \\
&= \sum_{k=0}^{\infty} \frac{D^{\kappa(\mathbf{1}^k)} f_{\mathit{1}\mathbf{2}}(x(0))}{k!} X^x_{0,1}\big(k!\mathbf{1}^k\mathbf{2}\big) \\
&= [1 - \sigma(-p; a)] X^x_{0,1}\big(\mathbf{2}\big) + \sum_{k=1}^{\infty} D^{\kappa(\mathbf{1}^k)} f_{\mathit{1}\mathbf{2}}(x(0)) X^x_{0,1}\big(\mathbf{1}^k\mathbf{2}\big)
\end{aligned}
$$

I can't pull this into a nice exact result, having tried both with the actual sigmoid (using the derivatives from [4]) or the error function, but this calculation shows what we might have understood intuitively (e.g. from spatial units consideration) that the best result is some kind of linear combination of the values of the signature at words containing exactly one $\mathbf{2}$.

Another idea

Consider a path $x : [0,1] \to \mathbb{R}^2$ whose first dimension is time, so that $x_{\mathbf{1}}(t) = t$, and with $x_{\mathbf{2}}(0) = 0$. Under such a condition, the path can contain no backtracking, so the signature identifies the path completely. We have $x_{\mathbf{2}}(1) = X^x_{0,1}(\mathbf{2})$ and $\int_0^1 x_{\mathbf{2}}(t)\, dt = \int_0^1 x_{\mathbf{2}}(t)\, dx_{\mathbf{1}}(t) = X^x_{0,1}(\mathbf{21})$.

Another idea

Consider a path $x : [0,1] \to \mathbb{R}^2$ whose first dimension is time, so that $x_{\mathbf{1}}(t) = t$, and with $x_{\mathbf{2}}(0) = 0$. Under such a condition, the path can contain no backtracking, so the signature identifies the path completely. Let's try to find the proportion of time $x(t) > p$ for some p. So let $f_{\mathit{1}\mathbf{1}}(x) = 0$ and we want a smooth version of $f_{\mathit{1}\mathbf{2}}(x) \approx 1_{x>p}$.

6 Solving an RDE with constant coefficients, or $dy = A(y)\,dx$

Here we present a simple differential equation driven by a path, a rough differential equation. If $x : [a,b] \to \mathbb{R}^d$ is a smooth path and A is a $c \times c \times d$ array, then we can consider the system of c differential equations

$$dy_i = \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} y_j \, dx_k, \tag{8}$$

together with a given starting point $y(a)$. It defines a path $y : [a,b] \to \mathbb{R}^c$. Unlike the previous section, each y_i is not determined by one equation, but rather all of them.

The equations can be written in integral form

$$y_i(t) - y_i(a) = \int_a^t \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} y_j(t') \, dx_k(t'). \tag{9}$$

We can try to approximate the solution by a Picard Iteration. Our first approximation will be

$$y^{(0)}(t) = y(a) \tag{10}$$

and successive approximations will come from

$$y_i^{(m+1)}(t) = y_i(a) + \int_a^t \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} y_j^{(m)}(t') \, dx_k(t'). \tag{11}$$

which gives

$$\begin{aligned}
y_i^{(1)}(t) &= y_i(a) + \int_a^t \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} y_j(a) \, dx_k(t') \\
&= y_i(a) + \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} y_j(a) \int_a^t dx_k(t') \\
&= y_i(a) + \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} y_j(a) X_{a,t}^x(k) \\
&= y_i^{(0)}(t) + \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} y_j(a) X_{a,t}^x(k) &(12) \\
y_i^{(2)}(t) &= y_i(a) + \int_a^t \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} \left[y_j(a) + \sum_{j'=1}^{c} \sum_{k'=\mathbf{1}}^{\mathbf{d}} A_{jj'k'} y_j'(a) X_{a,t}^x(k') \right] dx_k(t') \\
&= y_i^{(1)}(t) + \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} \sum_{j'=1}^{c} \sum_{k'=\mathbf{1}}^{\mathbf{d}} A_{jj'k'} \int_a^t X_{a,t}^x(k') \, dx_k(t') \\
&= y_i^{(1)}(t) + \sum_{j=1}^{c} \sum_{k=\mathbf{1}}^{\mathbf{d}} A_{ijk} \sum_{j'=1}^{c} \sum_{k'=\mathbf{1}}^{\mathbf{d}} A_{jj'k'} X_{a,t}^x(k'k) &(13)
\end{aligned}$$

Each successive approximation is formed by adding a term to the previous one. The new term looks only at the next level of the signature of x higher than looked at before. Equation (5) will work to get higher signature elements of y.

simple examples

If $y(a)$ is the origin, then y stays at the origin. If $y(a) = 1$ and $c = d = 1$ and $A_{ijk} = 1$ then we have the exponential series.

7 Solving a more general RDE, or $dy = f(y)\,dx$

If $x : [a, b] \to \mathbb{R}^d$ is a smooth path and f is a smooth function from $\mathbb{R}^c$ to $c \times d$ matrices, then the equation $dy = f(y)\,dx$, together with the starting point $y(a)$, determines a path $y : [a, b] \to \mathbb{R}^c$. Expanded, it is a system of c differential equations:

$$y_i'(t) = \sum_{k=\mathbf{1}}^{\mathbf{d}} f_{ik}(y(t))\,x_k'(t) \tag{14}$$

for each i in $\{\mathit{1}, \ldots, c\}$. The case where f is a constant function was considered before. The equations can be written in the following integral form.

$$y_i(t) - y_i(a) = \int_a^t \sum_{k=\mathbf{1}}^{\mathbf{d}} f_{ik}(y(t'))\,dx_k(t') \tag{15}$$

and $f_{ik} \circ y$ can then be expanded using the fundamental theorem of calculus using the multivariate chain rule

$$= \int_a^t \sum_{k=\mathbf{1}}^{\mathbf{d}} \left\{ f_{ik}(y(a)) + \int_a^{t'} \sum_{j=\mathit{1}}^{c} f_{ik}^{(j)}(y(t''))y_j'(t'')\,dt'' \right\} dx_k(t') \tag{16}$$

and y' can be expanded using the original equation (14)

$$= \int_a^t \sum_{k=\mathbf{1}}^{\mathbf{d}} \left\{ f_{ik}(y(a)) + \int_a^{t'} \sum_{j=\mathit{1}}^{c} f_{ik}^{(j)}(y(t'')) \sum_{k'=\mathbf{1}}^{\mathbf{d}} f_{jk'}(y(t''))\,x_{k'}'(t'')\,dt'' \right\} dx_k(t') \tag{17}$$

and the integrals can be separated and the process (fundamental theorem of calculus, expand with (14)) can be iterated

$$= \sum_{k=\mathbf{1}}^{\mathbf{d}} \left\{ f_{ik}(y(a))X_{a,t}^x(k) + \sum_{j=\mathit{1}}^{c} f_{ik}^{(j)}(y(a)) \sum_{k'=\mathbf{1}}^{\mathbf{d}} f_{jk'}(y(a))X_{a,t}^x(k'k) + \ldots \right\} \tag{18}$$

8 Integrating a form as an RDE

The same setup we discussed in section 5 can be written as an RDE. If $x : [a, b] \to \mathbb{R}^d$ is a smooth path and f is a smooth function from $\mathbb{R}^d$ to $c \times d$ matrices, then the equation $dy = f(x)\,dx$, together with the starting point $y(a) = 0$, defines a path $y : [a, b] \to \mathbb{R}^c$. Expanded, it is a system of c differential equations: $y_i'(t) = \sum_{j=\mathbf{1}}^{\mathbf{d}} f_{ij}(x(t))\,x_j'(t)$ for each i in $\{\mathit{1}, \ldots, c\}$. Each y_i is determined by one equation. Let's try to convert this to an RDE. The basic idea is to add more dimensions to y. Let B be $\{\mathit{1}, \ldots, c, \mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\}$. We want y_i to be x_i for i in $\{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\}$. Now we have the following RDE.

$$y_i'(t) = \sum_{k=\mathbf{1}}^{\mathbf{d}} 1\,x_k'(t) \qquad \text{for } i \in \{\mathbf{1}, \mathbf{2}, \ldots, \mathbf{d}\} \tag{19}$$

$$y_i'(t) = \sum_{k=\mathbf{1}}^{\mathbf{d}} f_{ik}(y_k(t))\,x_k'(t) \qquad \text{for } i \in \{\mathit{1}, \ldots, c\} \tag{20}$$

which is in the form of Equation (14).

Let's plug this into Equation (18) to see the solution for i in $\{1, \dots, c\}$.

$$\begin{aligned} y_i(t) &= y_i(t) - y_i(a) \\ &= \text{“} \sum_{k=\mathbf{1}}^{\mathbf{d}} \left\{ f_{ik}(y(a)) X^x_{a,t}(k) + \overbrace{\sum_{j=1}^{c}}^{\substack{\text{i.e.} \\ j \in B}} \overbrace{f^{(j)}_{ik}(y(a))}^{\substack{\text{only nonzero} \\ \text{if } j \text{ is } \mathbf{blue}}} \sum_{k'=\mathbf{1}}^{\mathbf{d}} \overbrace{f_{jk'}(y(a))}^{\substack{=\delta_{jk'} \\ \text{if } j \text{ is } \mathbf{blue}}} X^x_{a,t}(k'k) + \dots \right\} \text{”} \\ &= \sum_{k=\mathbf{1}}^{\mathbf{d}} \left\{ f_{ik}(x(a)) X^x_{a,t}(k) + \sum_{j=\mathbf{1}}^{\mathbf{d}} f^{(j)}_{ik}(x(a)) X^x_{a,t}(jk) + \dots \right\} \end{aligned} \tag{21}$$

This matches the part of Equation (6) for w being 0 or 1 letter long.

References

[1] Antoine Lejay. “An introduction to rough paths”. In: *Séminaire de probabilités XXXVII*. Springer, 2003, pp. 1–59.

[2] Terry Lyons. “Differential equations driven by rough signals”. In: *Revista Matemática Iberoamericana* 14 (1998), pp. 215–310.

[3] Terry Lyons, Michael Caruana, and Thierry Lévy. *Differential Equations Driven by Rough Paths*. 2007.

[4] Ali A Minai and Ronald D Williams. “On the derivatives of the sigmoid”. In: *Neural Networks* 6.6 (1993), pp. 845–853.

[5] Anastasia Papavasiliou and Christophe Ladroue. “Parameter estimation for rough differential equations”. In: *The Annals of Statistics* 39.4 (2011), pp. 2047–2073. URL: `http://arxiv.org/abs/0812.3102`.