



UNIVERSITAT DE  
BARCELONA

Facultat de Matemàtiques  
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

---

EL MÈTODE DE LA  
SIGNATURA I APLICACIONS  
A DADES DE TIPUS  
FINANCER

---

Autor: David Cots Mañas

Director: Dr. José Manuel Corcuera Valverde

Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 17 de gener de 2024

# Abstract

The signature captures a non-parametric characteristic of a data stream that has been previously transformed into a path through an embedding algorithm.

In this work, we will introduce the signature method, which consists of embedding a data stream into a path to then use its signature as a feature in machine learning problems.

First, we will define the signature of a path and present its most important properties.

Then, we will explore different ways of embedding data streams into a continuous path and examine the application of this method to financial time series data.

Finally, we will see how we can implement a supervised machine learning model based on signatures.

# Resum

La signatura extrau una característica no paramètrica d'un flux de dades que ha sigut previàment transformat en un camí mitjançant un algorisme d'incrustació. A aquest treball introduïrem el mètode de la signatura, aquest consisteix en incrustar un flux de dades a un camí i utilitzar la signatura d'aquest com a característica en problemes d'aprenentatge automàtic. Primer definirem la signatura d'un camí en donarem les seves propietats més importants.

Després explorarem diferents formes d'incrustar un flux de dades en un camí continu i veurem l'aplicació d'aquest mètode a dades de sèrie temporal de tipus financer.

Finalment veurem com podem implementar un model d'aprenentatge automàtic supervisat basat en signatures.

## Agraïments

En primer lloc, vull agrair al Dr. José Manuel Corcuera per tutoritzar i proposar el tema d'aquest treball.

M'agradaria també agrair als meus pares per haver-me donat la possibilitat de cursar aquests estudis i per donar-me tot el seu suport.

M'agradaria fer també un agraïment general a tots els docents del grau que a través de la seva passió i dedicació inspiren a molts alumnes.

Finalment, voldria agrair també a tots els amics i companys, així com a la meva germana, que m'han acompanyat i donat suport durant aquest llarg camí.

# Índex

<b>1</b>	<b>Introducció</b>	<b>1</b>
<b>2</b>	<b>La signatura d'un camí</b>	<b>2</b>
2.1	Conceptes previs . . . . .	2
2.2	Definició de signatura . . . . .	5
2.3	Intuïció geomètrica dels primers nivells . . . . .	10
2.3.1	Primer nivell . . . . .	10
2.3.2	Termes d'ordre superior . . . . .	10
2.3.3	Segon nivell (àrea) . . . . .	10
2.3.4	Tercer nivell (àrea de segon ordre) . . . . .	11
2.3.5	Signe de l'àrea signada . . . . .	11
2.4	Propietats de la signatura . . . . .	11
2.4.1	Invariància respecte a reparametritzacions del temps . . . . .	12
2.4.2	Linealitat: Producte mescla . . . . .	12
2.4.3	Identitat de Chen . . . . .	13
2.4.4	Inversió en el temps . . . . .	14
2.4.5	Log-signatura . . . . .	14
2.4.6	Iteracions de Picard: origen de la signatura . . . . .	15
2.4.7	Declivi Factorial . . . . .	17
2.4.8	Unicitat de la signatura . . . . .	17
2.4.9	Relació amb la Teoria de <i>Rough Paths</i> . . . . .	18
<b>3</b>	<b>Transformació d'un flux de dades en camí</b>	<b>19</b>
3.1	Incrustació mitjançant interpolació lineal a trossos . . . . .	19
3.2	Incrustació mitjançant interpolació rectilínia . . . . .	20
3.3	Suma cumulativa . . . . .	20
3.4	Incrustació mitjançant augmentacions . . . . .	21
3.5	Exemple explícit de càlcul . . . . .	25
<b>4</b>	<b>Signatura de fluxos de dades de caire financer</b>	<b>26</b>
4.1	La transformació <i>time-joined</i> i la signatura d'una sèrie temporal . . . . .	26
4.2	La signatura i els moments estadístics . . . . .	27
4.3	Variació quadràtica . . . . .	27
4.4	La suma cumulativa d'una successió . . . . .	28

<b>5</b>	<b>La signatura en l'aprenentatge automàtic</b>	<b>30</b>
5.1	Aprenentatge supervisat . . . . .	30
5.2	Les signatures com a característiques . . . . .	31
5.3	Model d'aprenentatge automàtic basat en signatures . . . . .	31
5.4	Implementació del model a un cas pràctic . . . . .	33
<b>6</b>	<b>Conclusions</b>	<b>35</b>
<b>A</b>	<b>Codi que computa la transformació Lead-Lag</b>	<b>38</b>
<b>B</b>	<b>Codi que computa la transformació time-joined</b>	<b>39</b>
<b>C</b>	<b>Codi que implementa el model de la secció 5</b>	<b>40</b>
<b>D</b>	<b>Conjunts de dades usats per a provar el model</b>	<b>46</b>

# 1 Introducció

La signatura d'un camí és una col·lecció infinita de nombres reals que captura certes propietats geomètriques d'aquest. Tot i que un camí no està totalment determinat per la seva signatura, la signatura determina completament les propietats geomètriques d'un camí sota certes condicions de no degeneració. Això la fa útil per a caracteritzar fluxos de dades, incrustats en camins prèviament. En això últim consisteix precisament el mètode de la signatura: donat un flux de dades, l'incrustem a un camí continu i posteriorment usem la signatura com a característica d'aquest dins del marc de l'aprenentatge automàtic.

## Estructura de la Memòria

En el primer capítol donem la definició de signatura d'un camí, així com les dels conceptes previs necessaris. A la mateixa secció hi trobem certes propietats importants de la signatura, així com exemples i la intuïció geomètrica darrera de la signatura. En el segon capítol expliquem primer diferents maneres d'incrustar un flux de dades a un camí. Al següent capítol explorem la signatura aplicada a dades de caire financer, o més en general, dades en forma de sèrie temporal. Finalment, a l'últim capítol discutim un model d'aprenentatge automàtic basat en la signatura.

## 2 La signatura d'un camí

### 2.1 Conceptes previs

A continuació revisem una sèrie de nocions bàsiques d'anàlisi matemàtic.

**Definició 2.1.** *Un espai normat  $(V, \|\cdot\|)$  és un parell format per un espai vectorial  $V$  sobre un cos  $\mathbb{K}$  i una norma  $\|\cdot\| : V \rightarrow \mathbb{K}$ .*

Les normes induïxen una funció distància invariant per traslacions anomenada la *distància canònica* o *distància induïda per la norma*, definida  $\forall u, v \in V$  com:

$$d(u, v) := \|x - y\| = \|y - x\|.$$

Podem formar, per tant, l'espai mètric  $(V, d)$ .

**Definició 2.2.** *Una successió  $\{v_i\}_{i \in \mathbb{N}}$  en un espai mètric  $(V, d)$  és diu de Cauchy si per a cada  $r > 0$  real,  $\exists N \in \mathbb{N} \cup \{0\}$  tal que  $\forall m, n > N$*

$$d(v_n, v_m) < r$$

**Definició 2.3.** *Un espai mètric  $(V, d)$  es diu complet si es compleix alguna de les següents condicions equivalents:*

1. *Tota successió de Cauchy de punts en  $V$  té un límit en  $V$ .*
2. *Tota successió de Cauchy en  $V$  convergeix a  $V$ .*
3. *Tota successió decreixent de subespais no buits de  $V$ , amb diàmetres que tendeixen a 0, té una intersecció no buida: si  $F_n$  es tancat i no buit,  $F_{n+1} \subseteq F_n$  per a cada  $n$ , i  $\text{diam}(F_n) \rightarrow 0$ , aleshores hi ha un punt  $v \in V$  comú a tots els conjunts  $F_n$ .*

**Definició 2.4.** *Un espai de Banach  $V$  és un espai normat i complet  $(V, \|\cdot\|)$  en la mètrica induïda per la seva norma.*

**Exemple 2.5.** L'espai euclidià  $(\mathbb{R}^d, \|\cdot\|_2)$  és un espai de Banach.

En la següent secció treballarem sobre l'espai euclidià  $\mathbb{R}^d$ , per major simplicitat, però notem que tot el que s'exposa a continuació es pot estendre a un espai de Banach  $V$  qualsevol.

**Definició 2.6.** *Un camí en un espai euclidià  $\mathbb{R}^d$ , és una aplicació contínua  $X : [a, b] \rightarrow \mathbb{R}^d$ , on  $[a, b] \subset \mathbb{R}$ .*

Notarem  $X_t = X(t)$  per a denotar la dependència del temps  $t \in [a, b]$ . Per a  $X_t \in \mathbb{R}^d$ , representarem la parametrització del camí de la forma:

$$X_t = \{X_t^1, \dots, X_t^d\}$$

D'ara en endavant asumirem que tots els camins que considerem són diferenciables a trossos. Quan parlem de camí *suau* voldrem dir que té derivades de tots els ordres.

**Exemple 2.7** (Figura 1). Un exemple de camí suau a  $\mathbb{R}^2$  és el següent:

$$X_t = \{X_t^1, X_t^2\} = \{t, t^2\}, t \in [-4, 4]$$

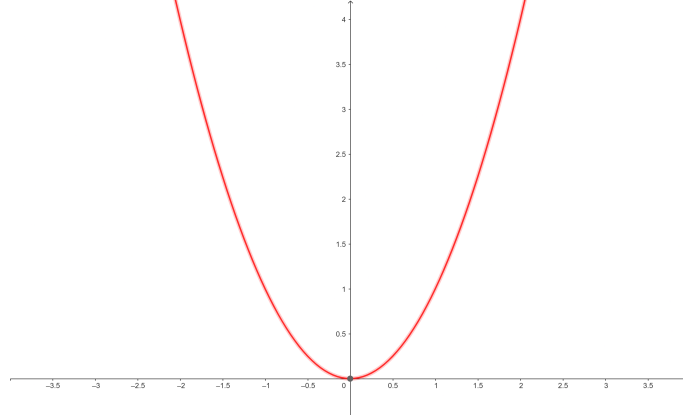


Figura 1: Exemple d'un camí bidimensional suau

**Exemple 2.8.** Un exemple d'un camí diferenciable a trossos pot ser el que observem a la figura 2

$$X_t = \{X_t^1, X_t^2\} = \{t, f(t)\}, t \in [0, 1]$$

on  $f$  és una funció lineal a trossos. Per exemple,  $f$  podria ser el preu d'un actiu a temps  $t$ . Aquest tipus de camins no suaus poden representar dades seqüencials, com per exemple, sèries temporals. A la Figura 2 observem un exemple de camí estocàstic bidimensional no suau.



Figura 2: BTC/EUR període: 05/05/2023-05/06/2023.

**Definició 2.9.** Siguin  $X : [a, b] \rightarrow \mathbb{R}$  un camí 1-dimensional i una funció  $f : \mathbb{R} \rightarrow \mathbb{R}$ , la integral de línia de  $X$  respecte  $f$  es defineix com

$$\int_a^b f(X_t) dX_t = \int_a^b f(X_t) \dot{X}_t dt, \quad (2.1)$$

on la última integral és de Riemann d'una funció continua acotada i on  $\dot{X}_t = dX_t/dt$ .

**Exemple 2.10.** Per al camí constant  $Y_t = 1, \forall t \in [a, b]$ . La integral de línia de  $Y_t$  respecte qualsevol camí  $X : [a, b] \rightarrow \mathbb{R}$  és l'increment de  $X$  :

$$\int_a^b dX_t = \int_a^b \dot{X}_t dt = X_b - X_a$$



**Exemple 2.11.** Per al camí constant  $X_t = t$ ,  $\forall t \in [a, b]$ ,  $\dot{X}_t = 1$  i la integral de línia per a qualsevol  $Y : [a, b] \mapsto \mathbb{R}$  és simplement la integral de Riemann de  $Y$ :

$$\int_a^b Y_t dX_t = \int_a^b Y_t dt.$$

**Definició 2.12.** Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$  un camí  $d$ -dimensional. Diem que  $X$  té una  $p$ -variació finita o acotada per cert  $p \geq 1$  si la  $p$ -variació de  $X$  es defineix com

$$\|X\|_{p,[a,b]} = \left( \sup_{D \subset [a,b]} \sum_l \|X_{t_l} - X_{t_{l-1}}\|^p \right)^{1/p} < \infty$$

on el suprem s'agafa sobre totes les possibles particions finites  $D = t_{il}$  de l'interval  $[a, b]$ . Denotarem  $\mathcal{V}^p([a, b], \mathbb{R}^d)$  el conjunt de tots els camins continus  $X : [a, b] \rightarrow \mathbb{R}^d$  de  $p$ -variació finita.

De vegades, als camins de 1-variació finita els anomenarem simplement de variació acotada o finita.

**Exemple 2.13.** Un moviment brownià fins a temps  $T$ , indicat per  $\mathcal{B}[0, T]$ , té  $p$ -variació finita quasi segur si i només si  $p > 2$ .

A continuació presentem alguns conceptes previs d'anàlisi estocàstic.

**Definició 2.14** (Procés estocàstic). Donat un espai de probabilitats  $(\Omega, \mathcal{F}, \mathbb{P})$ . Un procés estocàstic  $X$  és una aplicació:

$$X : (\omega, t) \in \Omega \times \mathbb{T} \rightarrow X_t(\omega) \in \mathbb{R}$$

mesurable, és a dir, tal que

$$X^{-1}(B) \in \mathcal{F} \otimes \mathcal{B}(\mathbb{T})$$

per a tot  $B \in \mathcal{B}(\mathbb{R})$ . On  $\mathbb{T}$  és un conjunt que mesura el temps. En temps continu, és habitual tenir  $\mathbb{T} = [0, T]$  o  $\mathbb{T} = [0, \infty)$ . En temps discret,  $\mathbb{T} := \{0, 1, \dots, N\}$  o  $\mathbb{T} = \mathbb{N}$ .

**Definició 2.15** (Reticulat o norma d'una partició). Donada una partició d'un interval  $[a, b]$

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$$

definim la norma de la partició o el reticulat com la longitud

$$\max\{|x_i - x_{i-1}| : i = 1, \dots, n\}.$$

És a dir, la longitud més gran de les longituds dels subintervalls  $[x_{i-1}, x_i]$ .

**Definició 2.16** (variació quadràtica). *Sigui  $X_t$  un procés estocàstic real sobre un espai de probabilitats  $(\Omega, \mathcal{F}, \mathbb{P})$  i amb l'índex real  $t \in \mathbb{R}_{\geq 0}$ . La seva variació quadràtica, denotada  $[X]_t$  es defineix com*

$$[X]_t = \lim_{\|P\| \rightarrow 0} \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^2$$

on  $P$  recorre sobre particions de l'interval  $[0, t]$  i la norma de la partició  $P$  és el reticulat. Aquest límit, d'existir, és la noció de límit en probabilitats.

La variació quadràtica és un cas particular de la covariància creuada.

**Definició 2.17** (Covariació). *La covariació o variació creuada de dos processos  $X$  i  $Y$  es defineix com*

$$[X, Y]_t = \lim_{\|P\| \rightarrow 0} \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})(Y_{t_k} - Y_{t_{k-1}})$$

**Observació 2.18.** La covariació es pot escriure en funció de la variació degut a la identitat de polarització:

$$[X, Y]_t = \frac{1}{2}([X + Y]_t - [X]_t - [Y]_t)$$

La variació quadràtica també es pot escriure  $\langle X \rangle_t$  o  $\langle X, X \rangle_T$ , així com  $QV(X)$ .

## 2.2 Definició de signatura

Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$  un camí,  $\forall i \in \{1, \dots, d\}$ , definim la quantitat

$$S(X)_{a,t}^i = \int_{a < s < t} dX_s^i = X_t^i - X_a^i \quad (2.2)$$

que és l'increment de la  $i$ -èssima coordenada de camí a temps  $t \in [a, b]$ . Observem que  $S(X)_{a,\cdot}^i : [a, b] \rightarrow \mathbb{R}$  és en si mateixa un camí 1-dimensional.

**Definició 2.19.** *Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$  un camí,  $\forall i, j \in \{1, \dots, d\}$ , definim la integral doble iterada com*

$$S(X)_{a,t}^{i,j} = \int_{a < s < t} S(X)_{a,s}^i dX_s^j = \int_{a < r < s < t} dX_r^i dX_s^j \quad (2.3)$$

on  $S(X)_{a,s}^i$  es donat per (2.2) i el límits d'integració són:

$$a < r < s < t = \begin{cases} a < r < s \\ a < s < t. \end{cases} \quad (2.4)$$

Anàlogament,  $\forall i, j, k \in \{1, \dots, d\}$  podem definir la integral triple iterada com

$$S(X)_{a,t}^{i,j,k} = \int_{a < s < t} S(X)_{a,s}^{i,j} dX_s^k = \int_{a < q < r < s < t} dX_q^i dX_r^j dX_s^k \quad (2.5)$$

Podem continuar recursivament:

Per a qualsevol  $k \geq 1$  i  $i_1, \dots, i_k \in \{1, \dots, d\}$  definim la integral  $k$  vegades iterada de  $X$  recorrent els índexos  $i_1, \dots, i_k$  com

$$S(X)_{a,t}^{i_1, \dots, i_k} = \int_{a < s < t} S(X)_{a,s}^{i_1, \dots, i_{k-1}} dX_s^{i_k} = \int_{a < t_k < t} \dots \int_{a < t_1 < t_2} dX_{t_1}^{i_1} \dots dX_{t_k}^{i_k} \quad (2.6)$$

**Definició 2.20** (Signatura). La signatura d'un camí  $X : [a, b] \rightarrow \mathbb{R}^d$  és la col·lecció (sèrie infinita) de totes les integrals iterades de  $X$ . Formalment, és la successió de nombres reals

$$S(X)_{a,b} = (1, S(X)_{a,b}^1, \dots, S(X)_{a,b}^d, S(X)_{a,b}^{1,1}, S(X)_{a,b}^{1,2}, \dots) \quad (2.7)$$

on els superíndexos recorren el conjunt de tots els multi-índexos

$$W = \{(i_1, \dots, i_k) \mid k \geq 1, i_1, \dots, i_k \in \{1, \dots, d\}\}. \quad (2.8)$$

El conjunt  $W$  es sol anomenar el conjunt de paraules de l'alfabet  $A = \{1, \dots, d\}$  de  $d$  lletres.

De vegades denotarem simplement  $S(X) = (1, S(X)^1, S(X)^2, \dots)$ , per a la signatura d'un camí  $X : [a, b] \mapsto \mathbb{R}^d$ , ometent el subíndex per tal d'alleugerar notació.

**Exemple 2.21.** Si considerem l'alfabet de tres lletres  $\{1, 2, 3\}$ , tenim un nombre infinit de paraules que podem fer amb aquest alfabet:  $(1, 2, 3, 11, 12, 13, 21, 22, 23, 31, 32, 33, 111, 112, 113, 121, \dots)$ .

**Observació 2.22.** Observem que les integrals iterades d'un camí són independents del punt inicial de  $X$ . És a dir, si per algun  $x \in \mathbb{R}^d$ , definim el camí  $\tilde{X}_t = X_t + x$ , aleshores  $S(\tilde{X})_{a,b}^{i_1, \dots, i_k} = S(X)_{a,b}^{i_1, \dots, i_k}$ .

A continuació explorem una definició més completa de signatura.

**Definició 2.23** (Sèrie de Potències Formal). Siguin  $e_1, \dots, e_d$   $d$  indeterminats formals. L'àlgebra de sèries de potències formals (no commutatives) en  $d$  indeterminats és l'espai vectorial de totes les sèries de la forma

$$\sum_{k=0}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \lambda_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k}, \quad (2.9)$$

**Definició 2.24.** Un polinomi formal (no commutatiu) és una sèrie de potències formal per a la que només un nombre finit de coeficients  $\lambda_{i_1, \dots, i_k}$  és no nul. Els termes  $e_{i_1}, \dots, e_{i_k}$  s'anomenen monomis.

El terme que correpon a  $k = 0$  és un nombre real  $\lambda_0$ . L'espai de les sèries formals es sol anomenar l'àlgebra tensorial de  $\mathbb{R}^d$ , i el podem escriure  $T(\mathbb{R}^d)$ .

Observem que si definim la suma i el producte escalar de la següent manera

$$\left( \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \lambda_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \right) + \left( \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \mu_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \right) = \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} (\lambda_{i_1, \dots, i_k} + \mu_{i_1, \dots, i_k}) e_{i_1} \dots e_{i_k} \quad (2.10)$$

i

$$c \left( \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \lambda_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \right) = \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} c \lambda_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \quad (2.11)$$

Podem definir, a més, el producte  $\otimes$  entre monomis ajuntant multi-índexos.

$$e_{i_1} \dots e_{i_k} \otimes e_{j_1} \dots e_{j_k} = e_{i_1} \dots e_{i_k} e_{j_1} \dots e_{j_k} \quad (2.12)$$

El producte  $\otimes$  s'exten llavors de forma única i lineal a totes les sèries de potències formals. Observem-ne els primers termes:

$$\begin{aligned} \left( \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \lambda_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \right) \otimes \left( \sum_{k=1}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \mu_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \right) = \\ \lambda_0 \mu_0 + \sum_{i=1}^d (\lambda_0 \mu_i + \lambda_i \mu_0) e_i + \sum_{i=1}^d (\lambda_0 \mu_{i,j} + \lambda_i \mu_j + \lambda_{i,j} \mu_0) e_i e_j + \dots \end{aligned} \quad (2.13)$$

L'espai de sèries formals forma un àlgebra en l'espai vectorial que queda definit junt amb  $\otimes$ .

Hom pot observar que el conjunt d'índexos dels monomis  $e_{i_1}, \dots, e_{i_K}$ , coincideix amb el conjunt d'índexos dels termes de la signatura d'un camí  $X : [a, b] \rightarrow \mathbb{R}^d$ , més en particular la col·lecció de tots els multi-índexos  $(i_1, \dots, i_K)$  on  $i_1, \dots, i_k \in \{1, \dots, d\}$ . Podem observar doncs que una forma d'expressar la signatura del camí  $X$  és amb una sèrie de potències formal en que el coeficient de cada monomi  $e_{i_1}, \dots, e_{i_k}$  és defineix com  $S(X)_{a,b}^{i_1, \dots, i_k}$ , és a dir:

$$S(X)_{a,b} = \sum_{k=0}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} S(X)_{a,b}^{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \quad (2.14)$$

Per tant, podem definir alternativament la signatura d'una camí de la forma que s'exposa a continuació:

Definim la  $n$ -èsima potència tensorial de  $\mathbb{R}^d$  com:

$$(\mathbb{R}^d)^{\otimes n} := \underbrace{\mathbb{R}^d \otimes \mathbb{R}^d \dots \otimes \mathbb{R}^d}_n$$

on  $\otimes$  és el producte tensorial. Definim l'àlgebra tensorial com

$$T(\mathbb{R}^d) := \{(a_0, a_1, a_2, \dots) : a_n \in (\mathbb{R}^d)^{\otimes n} \forall n \geq 0\},$$

on prenem  $(\mathbb{R}^d)^{\otimes 0} = \mathbb{R}$  per conveni. A més, definim l'àlgebra tensorial truncada com

$$T(\mathbb{R}^d) := \bigoplus_{i=0}^n (\mathbb{R}^d)^{\otimes i}.$$

Donats dos elements  $a = (a_0, a_1, \dots), b = (b_0, b_1, \dots) \in T(\mathbb{R}^d)$  introduïm la suma i producte de  $a$  i  $b$  com

$$a + b := (a_0 + b_0, a_1 + b_1, \dots)$$

$$a \otimes b = ab := \left( \sum_{i=0}^n a_i \otimes b_{n-i} \right)_{n \geq 0}.$$

Que es defineix de forma anàloga en l'àlgebra tensorial truncada. Definim doncs:

**Definició 2.25** (Signatura). *La signatura d'un camí  $X : [a, b] \rightarrow \mathbb{R}^d$  és la seqüència*

$$S(X)_{a,b} = (1, X_{a,b}^1, X_{a,b}^2, \dots) \in T(\mathbb{R}^d) \quad (2.15)$$

on

$$X_{a,b}^n := \int \cdots \int_{a < t_1 < t_2 < \dots < t_n < b} dX_{t_1} \otimes \dots \otimes dX_{t_n} \in (\mathbb{R}^d)^{\otimes n} \quad \forall n \geq 0 \quad (2.16)$$

De nou, de vegades ometrem els subíndexos per a alleugerar notació.

**Exemple 2.26.** 1. Quan  $X$  té 1-variació finita, la integració es pot entendre en el sentit de la integral de Stieltjes.

2. Quan  $X$  és un moviment brownià, la seva signatura es pot definir en el sentit de la integral d'Itô o la integral de Stratonovich.

També ens interessa la següent definició:

**Definició 2.27** (Nivell  $k$ -èssim de la signatura). *Definim el nivell  $k$ -èssim de la signatura o signatura truncada d'ordre  $k$  d'un camí  $X : [a, b] \rightarrow \mathbb{R}^d$ , com la col·lecció finita de tots els termes  $S(X)_{a,b}^{i_1, \dots, i_k}$ , on el multi-índex és de longitud  $k$ .*

Així, per exemple, el primer nivell de la signatura és la col·lecció de  $d$  nombres reals  $S(X)_{a,b}^1, \dots, S(X)_{a,b}^d$  i el segon nivell és la col·lecció de  $d^2$  nombres reals

$$S(X)_{a,b}^{1,1}, \dots, S(X)_{a,b}^{1,d}, S(X)_{a,b}^{2,1}, \dots, S(X)_{a,b}^{2,d}.$$

En termes de la segona definició que hem ofert de signatura la *signatura truncada* de nivell  $k$  és

$$S(X)_{a,b}^k = (1, X^1, \dots, X^k).$$

**Exemple 2.28.** Considerem un camí unidimensional  $X : [a, b] \rightarrow \mathbb{R}$ ,  $X_t = t$ , en aquest cas l'alfabet és  $A = \{1\}$  i el conjunt de paraules és  $W = \{(1, \dots, 1) \mid k \geq 1\}$ , on 1 apareix  $k$  vegades en  $(1, \dots, 1)$ . La signatura és doncs

$$S(X)_{a,b}^1 = X_b - X_a, \quad (2.17)$$

$$S(X)_{a,b}^{1,1} = \frac{(X_b - X_a)^2}{2!},$$

$$S(X)_{a,b}^{1,1,1} = \frac{(X_b - X_a)^3}{3!},$$

$\vdots$

**Observació 2.29.** Es por demostrar que aquesta última expressió és certa per a qualsevol camí  $X : [a, b] \rightarrow \mathbb{R}$ . Per a camins unidimensionals la signatura només depèn de l'increment  $X_b - X_a$ .

**Exemple 2.30.** Considerem un camí bidimensional. Tenim l'alfabet  $A = \{1, 2\}$  com a conjunt de índexos, i el conjunt de paraules o multi-índexos  $W = \{(i_1, \dots, i_K) \mid k \geq 1, i_1, \dots, i_k \in \{1, 2\}\}$ . Considerem el següent camí a  $\mathbb{R}^d$ :

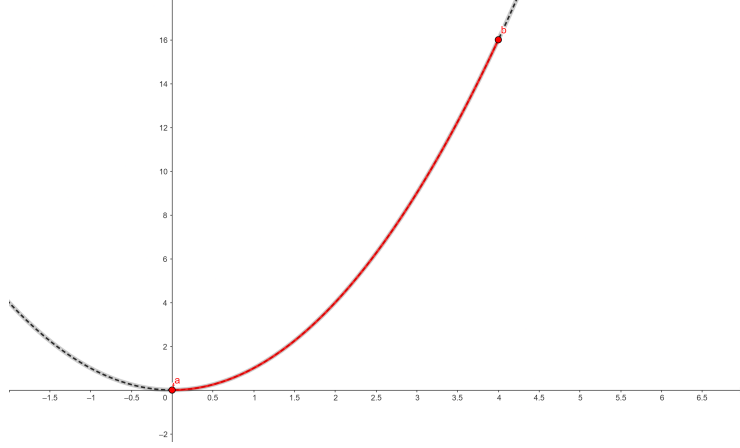


Figura 3: Camí bidimensional  $X_t, t \in [0, 4]$ .

$$\begin{aligned} X_t &= \{X_t^1, X_t^2\} = \{t, t^2\}, t \in [0, 4]. \\ dX_t &= \{dX_t^1, dX_t^2\} = \{dt, 2tdt\}. \end{aligned} \quad (2.18)$$

Computem la signatura

$$\begin{aligned} S(X)_{0,4}^1 &= \int_{0 < t < 4} dX_t^1 dt = \int_0^4 dt = X_4^1 - X_0^1 = 4, \\ S(X)_{0,4}^2 &= \int_{0 < t < 4} dX_t^2 dt = \int_0^4 2tdt = X_4^2 - X_0^2 = 16, \\ S(X)_{0,4}^{1,1} &= \iint_{0 < t_1 < t_2 < 4} dX_{t_1}^1 dX_{t_2}^1 dt = \int_0^4 \left[ \int_0^{t_2} dt_1 \right] dt_2 = 8, \\ S(X)_{0,4}^{1,2} &= \iint_{0 < t_1 < t_2 < 4} dX_{t_1}^1 dX_{t_2}^2 dt = \int_0^4 \left[ \int_0^{t_2} dt_1 \right] 2t_2 dt_2 = \frac{128}{3}, \\ S(X)_{0,4}^{2,1} &= \iint_{0 < t_1 < t_2 < 4} dX_{t_1}^2 dX_{t_2}^1 dt = \int_0^4 \left[ \int_0^{t_2} 2t_1 dt_1 \right] dt_2 = \frac{64}{3}, \\ S(X)_{0,4}^{2,2} &= \iint_{0 < t_1 < t_2 < 4} dX_{t_1}^2 dX_{t_2}^2 dt = \int_0^4 \left[ \int_0^{t_2} 2t_1 dt_1 \right] 2t_2 dt_2 = 128, \\ S(X)_{0,4}^{1,1,1} &= \iiint_{0 < t_1 < t_2 < t_3 < 4} dX_{t_1}^1 dX_{t_2}^1 dX_{t_3}^1 = \int_0^4 \left[ \int_0^{t_3} \left[ \int_0^{t_2} dt_1 \right] dt_2 \right] dt_3 = \frac{32}{3}, \\ &\vdots \end{aligned} \quad (2.19)$$

Si seguim de la mateixa manera, podem calcular cada terme  $S(X)_{0,5}^{i_1, \dots, i_K}$  de la signatura per a cada multi-índex  $(i_1, \dots, i_k)$  amb  $i_1, \dots, i_k \in \{1, 2\}$ .

## 2.3 Intuïció geomètrica dels primers nivells

A continuació donem el significat geomètric dels dos primers nivells de la signatura.

### 2.3.1 Primer nivell

El primer nivell, tal i com hem comentat a 2.17  $(1, X^1, \dots, X^d)$ , és l'increment del camí  $X : [a, b] \rightarrow \mathbb{R}^d$ , és a dir que

$$X^{(i)} = X_b^i - X_a^i$$

### 2.3.2 Termes d'ordre superior

Les integrals iterades d'ordre superior poden ser interpretades com polinomis generalitzats de camins. La regressió lineal sobre aquests polinomis generalitzats es pot interpretar com a regressió polinomial directe sobre els camins.

No obstant, no tots els termes són necessaris per a la representació única d'un camí en termes de les seves integrals iterades doncs certes integrals iterades estan determinades per altres integrals iterades. Per exemple:

$$\begin{aligned} X^{(i,i)} &= \frac{(X^{(i)})^2}{2} = \frac{(X_b^i - X_a^i)^2}{2}, \quad i \in \{1, \dots, d\} \\ X^{(i,j)} &= X^{(i)}X^{(j)} - X^{(j,i)}, \quad i, j \in \{1, \dots, d\} \end{aligned} \tag{2.20}$$

Aquestes últimes igualtats són conseqüència de la identitat del producte mescla (Teorema 2.8) que veiem més endavant a la subsecció 2.4.2.. La *log-signatura* que es presenta a la subsecció 2.4.5 presenta la mínima informació que determina únicament totes les integrals iterades d'un camí a través d'operacions no lineals.

### 2.3.3 Segon nivell (àrea)

Observem doncs que  $X_{a,b}^{(i,i)}$  és  $(X_b^i - X_a^i)^2/2$ . Per al terme  $X_{a,b}^{(i,j)}$  per  $i \neq j$ . Per  $1 \leq i \leq j \leq d$ , els termes

$$A_{a,b}^{i,j} := \frac{1}{2} \left( \int_{a < t_1 < t_2 < b} dX_{t_1}^1 dX_{t_2}^2 - \int_{a < t_1 < t_2 < b} dX_{t_1}^2 dX_{t_2}^1 \right) = \frac{1}{2} \left( X_{a,b}^{(i,j)} - X_{a,b}^{(j,i)} \right)$$

determinen l'àrea signada entre la corba  $t \mapsto (X_t^i, X_t^j)$ , per  $t \in [a, b]$  i la corda que uneix els punts  $(X_a^i, X_a^j)$  i  $(X_b^i, X_b^j)$ . Aquesta àrea es la que es coneix com *àrea de Lévy*.

En la figura 4, pel camí bidimensional  $X_t = \{X_t^1, X_t^2\}$  les àrees signades  $A_-$  i  $A_+$  són negativa i positiva respectivament i  $\Delta X_t^1$  i  $\Delta X_t^2$  són els increments en cada coordenada.

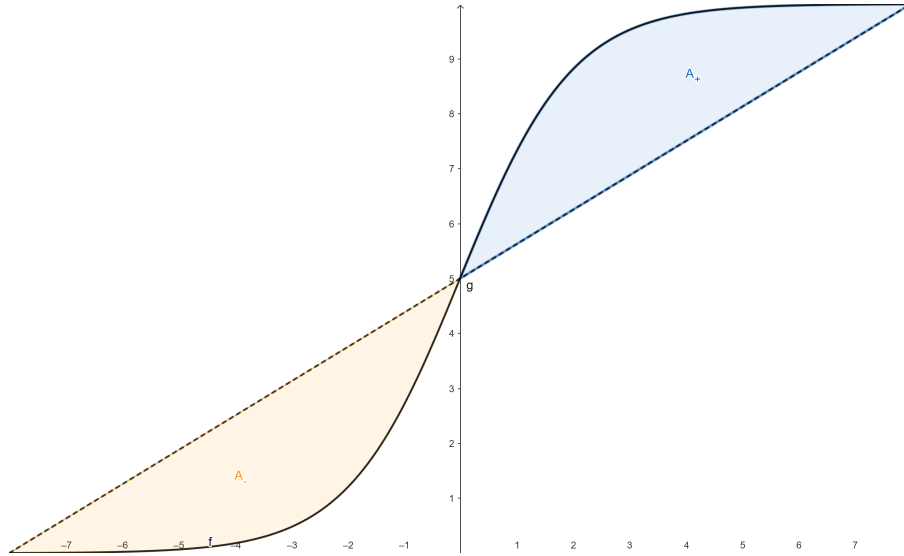


Figura 4: Camí bidimensional  $X_t$ ,  $t \in [0, 4]$ .

### 2.3.4 Tercer nivell (àrea de segon ordre)

La gràfica de la figura 4 es pot dividir en dues trajectòries en dues dimensions. Totes dues trajectòries tenen increments unitaris en ambdues components i àrees signades nul·les. L'àrea de segon ordre es calcula mitjançant la fórmula

$$A_{1,(1,2)}^{a,b} := \frac{1}{2} \left( \int_{a < t_1 < t_2 < b} dX_{t_1}^1 dA_{b,t_2}^{1,2} - \int_{a < t_1 < t_2 < b} dA_{b,t_1}^{1,2} dX_{t_2}^1 \right) = \frac{1}{2} (X_{a,b}^{(1,1,2)} - X_{s,t}^{(1,2,1)}),$$

l'últim pas s'infereix de la propietat que es presenta a la subsecció 2.4.2 (una demostració d'això es pot trobar a [14]).

Mitjançant arguments recursius similars, es poden identificar àrees d'ordre superior i representar-les en termes de combinacions lineals d'integrals iterades.

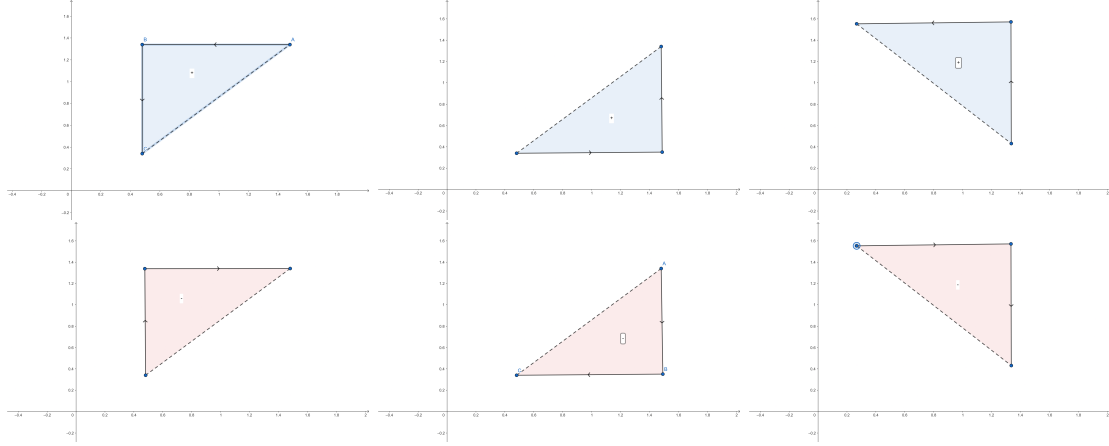
### 2.3.5 Signe de l'àrea signada

A la figura 5 observem les sis possibilitats en funció de la direcció del camí amb el que estem treballant: les tres de dalt són positives i les tres d'abaix negatives. Això té relació directa amb l'índex del camí (vegis [15]).

## 2.4 Propietats de la signatura

A continuació presentem algunes propietats de la signatura, en deixem diverses sense demostrar però les demostracions es poden trobar en la bibliografia citada





### 2.4.1 Invariància respecte a reparametritzacions del temps

Anomenem *reparametrització* a una aplicació  $\psi : [a, b] \rightarrow [a, b]$  exhaustiva, contínua i no decreixent. D'ara en endavant considerarem reparametritzacions suaus només.

Considerem dos camins  $X, Y : [a, b] \rightarrow \mathbb{R}$  i una reparametrització  $\psi : [a, b] \rightarrow [a, b]$ . Definim els camins  $\tilde{X}, \tilde{Y} : [a, b] \rightarrow \mathbb{R}$  com  $\tilde{X}_t = X_{\psi(t)}$  i  $\tilde{Y}_t = Y_{\psi(t)}$ . Notem que  $\dot{\tilde{X}}_t = \dot{X}_{\psi(t)}\dot{\psi}(t)$  del que es segueix que

$$\int_a^b \tilde{Y}_t d\tilde{X}_t = \int_a^b Y_{\psi(t)} \dot{X}_{\psi(t)} \dot{\psi}(t) dt = \int_a^b Y_u dX_u, \quad (2.21)$$

on a la última igualtat hem usat la substitució  $u = \psi(t)$ . És a dir, que les integrals de línia són invariants sota reparametritzacions del temps dels dos camins.

Considerem ara un camí  $X : [a, b] \rightarrow \mathbb{R}^d$  i una reparametrització  $\psi : [a, b] \rightarrow [a, b]$ . Com abans,  $\tilde{X}_t = X_{\psi(t)}$ . Com que cada terme de la signatura  $S(X)_{a,b}^{i_1, \dots, i_k}$  és una integral de línia iterada de  $X$ , es dedueix de (2.19) que

$$S(\tilde{X})_{a,b}^{i_1, \dots, i_k} = S(X)_{a,b}^{i_1, \dots, i_k}, \forall k \geq 0, i_1, \dots, i_k \in \{1, \dots, d\}. \quad (2.22)$$

### 2.4.2 Linealitat: Producte mescla

Una permutació  $\sigma$  de  $\{1, \dots, k+m\}$  s'anomena  $(k, m)$ -mescla si  $\sigma^{-1}(1) < \dots < \sigma^{-1}(k)$  i  $\sigma^{-1}(k+1) < \dots < \sigma^{-1}(k+m)$ . La llista  $(\sigma(1), \dots, \sigma(k+m))$  també s'anomena una mescla de  $(1, \dots, k)$  i  $(k+1, \dots, k+m)$ . Denotarem  $M(k, m)$  la col·lecció de totes les  $(k, m)$ -mescles.

**Definició 2.31** (Producte Mescla). *Considerem dos multi-índexos  $I = (i_1, \dots, i_k)$  i  $J = (j_1, \dots, j_m)$  amb  $i_1, \dots, i_k, j_1, \dots, j_m \in \{1, \dots, d\}$ . Considerem el multi-índex*

$$(r_1, \dots, r_k, r_{k+1}, \dots, r_{k+m}) = (i_1, \dots, i_k, j_1, \dots, j_m). \quad (2.23)$$

*El producte mescla de  $I$  i  $J$ , que denotem  $I \sqcup J$ , és un conjunt finit de multi-índexos de longitud  $k+m$  de la següent manera*

$$I \sqcup J = \{(r_{\sigma(1)}, \dots, r_{\sigma(k+m)}) \mid \sigma \in M(k, m)\} \quad (2.24)$$

**Teorema 2.32** (Identitat del producte mescla). *Donat un camí  $X : [a, b] \rightarrow \mathbb{R}^d$  i dos multi-índexos  $I = (i_1, \dots, i_k)$  i  $J = (j_1, \dots, j_m)$  amb  $i_1, \dots, i_k, j_1, \dots, j_m \in \{1, \dots, d\}$  aleshores*

$$S(X)_{a,b}^I S(X)_{a,b}^J = \sum_{K \in I \sqcup J} S(X)_{a,b}^K. \quad (2.25)$$

El producte mescla implica, en particular, que el producte de dos termes de la signatura es pot expressar com a una combinació lineal de termes d'ordre superior, cosa que serà útil de cara a les aplicacions pràctiques.

### 2.4.3 Identitat de Chen

**Definició 2.33** (Concatenació). *Per a dos camins  $X : [a, b] \rightarrow \mathbb{R}^d$  i  $Y : [b, c] \rightarrow \mathbb{R}^d$ , definim la seva concatenació com el camí  $X * Y : [a, c] \rightarrow \mathbb{R}^d$  de la següent forma*

$$(X * Y)_t = \begin{cases} X_t & t \in [a, b] \\ X_b + (Y_t - Y_b) & t \in [b, c] \end{cases}$$

**Teorema 2.34** (Identitat de Chen). *Sigui  $p \in [1, 2)$ ,  $X : [a, b] \rightarrow \mathbb{R}^d$  i  $Y : [b, c] \rightarrow \mathbb{R}^d$  camins de  $p$ -variació finita. Aleshores*

$$S(X * Y)_{a,c} = S(X)_{a,b} \otimes S(Y)_{b,c} \quad (2.26)$$

*Demostració.* Sigui  $Z = X * Y$  i  $S(Z)$  la seva signatura. Prenem  $n \geq 1$  i computem el  $n$ -èssim terme de la signatura

$$\begin{aligned} Z_{a,c}^n &= \int_{a < u_1 < \dots < u_n < c} dZ_{u_1} \otimes \dots \otimes dZ_{u_n} \\ &= \sum_{k=0}^n \int_{a < u_1 < \dots < u_k < b < u_{k+1} < \dots < u_n < c} dZ_{u_1} \otimes \dots \otimes dZ_{u_n} \end{aligned}$$

Aplicant el Teorema de Fubini, això últim equival a

$$\begin{aligned} Z_{a,c}^n &= \sum_{k=0}^n \int_{a < u_1 < \dots < u_k < b} dX_{u_1} \otimes \dots \otimes dX_{u_k} \otimes \int_{b < u_{k+1} < \dots < u_n < c} dY_{u_{k+1}} \otimes \dots \otimes dY_{u_n} \\ &= \sum_{k=0}^n X_{a,b}^k \otimes Y_{b,c}^{n-k}, \end{aligned}$$

Pel que, com volíem veure,  $S(Z) = S(X) \otimes S(Y)$ . □

La identitat, ens diu que la signatura transforma el producte  $*$ , o concatenació en el producte  $\otimes$ .

**Exemple 2.35.** Sigui  $X$  un camí lineal a trossos de valor real, és a dir,  $X$  és la concatenació d'un nombre finit de camins lineals, o sigui que existeixen un enter positiu  $l$  i  $X_1, X_2, \dots, X_l$  camins lineals tals que

$$X = X_1 * X_2 * \dots * X_l$$

La identitat de Chen ens dona un mètode per a calcular la signatura d'un camí lineal a trossos, en primer lloc, per la Identitat de Chen és te que

$$S(X) = \bigotimes_{i=1}^l S(X_i)$$

Suposem que cada  $X_i$  està definit sobre l'interval  $[t_i, t_{i+1}]$ , com que  $X_i$  és un camí lineal es pot obtenir  $S(X_i)_{t_i, t_{i+1}}$  de l'exemple 2.17.

#### 2.4.4 Inversió en el temps

**Definició 2.36** (Inversió en el temps). *Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$ , definim la seva inversió en el temps com el camí  $\overleftarrow{X} : [a, b] \rightarrow \mathbb{R}^d$  de forma que  $\overleftarrow{X}_t = X_{a+b-t}$ ,  $\forall t \in [a, b]$ .*

Tenim la següent propietat algebraica de la signatura:

**Teorema 2.37** (Signatura invertida en el temps). *Donat un camí  $X : [a, b] \rightarrow \mathbb{R}^d$ , aleshores*

$$S(X)_{a,b} \otimes S(\overleftarrow{X})_{a,b} = 1. \quad (2.27)$$

#### 2.4.5 Log-signatura

Donada una sèrie de potències formal,

$$x = \sum_{k=0}^{\infty} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \lambda_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k} \quad (2.28)$$

amb  $\lambda_0 > 0$ , definim el seu logaritme com la sèrie de potències

$$\log x = \log \lambda_0 + \sum_{n \geq 1} \frac{(-1)^n}{n} \left(1 - \frac{x}{\lambda_0}\right)^{\otimes n}, \quad (2.29)$$

on amb  $\otimes n$  denotem la n-èsima potència respecte el producte  $\otimes$ .

D'aquesta manera, per exemple, per  $\lambda \in \mathbb{R}$  i la sèrie:

$$x = 1 + \sum_{k \geq 1} \frac{\lambda^k}{k!} e_1^{\otimes k}, \quad (2.30)$$

aleshores, es pot verificar que

$$\log x = \lambda e_1. \quad (2.31)$$

**Observació 2.38.** En general,  $\log x$  és una sèrie amb una quantitat infinita de termes, emperò per a cada multi-índex  $(i_1, \dots, i_k)$ , el coeficient de  $e_{i_1} \dots e_{i_k}$  en  $\log x$  només depèn dels coeficients de  $x$  de la forma  $\lambda_{j_1, \dots, j_m}$  amb  $m \leq k$ , dels quals n'hi ha un nombre finit, de forma que  $\log x$  és ben definit sense que calgui considerar la convergència de la sèrie infinita.

**Definició 2.39** (Log-signatura). *Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$  un camí, definim la log-signatura de  $X$  com la sèrie de potències  $\log S(X)_{a,b}$ .*

**Definició 2.40** (Parèntesi de Lie). *Donades  $x$  i  $y$  dues sèries de potències formals, definim el parèntesi de Lie per*

$$[x, y] = x \otimes y - y \otimes x \quad (2.32)$$

*Els primers termes de la log-signatura s'expressen:*

$$\log S(X)_{a,b} = \sum_{i=1}^d S(X)_{a,b}^i e_1 + \sum_{1 \leq i < j \leq d} \frac{1}{2} (S(X)_{a,b}^{i,j} - S(X)_{a,b}^{j,i}) [e_i, e_j] + \dots \quad (2.33)$$

Observem, de fet, que el coeficient del segon terme és precisament, l'àrea de Lévy.

El següent teorema, demostrat per Chen [6], generaliza la fórmula de Camper-Baker-Hausdorff.

**Teorema 2.41.** *Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$  un camí. Existeixen nombres reals  $\lambda_{i_1, \dots, i_k}$  tal que*

$$\log S(X)_{a,b} = \sum_{k \geq 1} \sum_{i_1, \dots, i_k \in \{1, \dots, d\}} \lambda_{i_1, \dots, i_k} [e_{i_1}, [e_{i_2}, \dots [e_{i_{k-1}}, e_{i_k}] \dots]]. \quad (2.34)$$

**Observació 2.42.** Els coeficients  $\lambda_{i_1, \dots, i_k}$  en general no són únics ja que els polinomis de la forma  $[e_{i_1}, [e_{i_2}, \dots, [e_{i_{k-1}}, e_{i_k}] \dots]]$  no són linealment independents.

## 2.4.6 Iteracions de Picard: origen de la signatura

El primer cop que es va estudiar la signatura d'un camí va ser a mans del geometre Kuo-Tsai Chen a l'any 1957 en l'estudi de la integració de camins ([6]). Les integral iterades de camins apareixen a l'hora d'aplicar iteracions de Picard per a resoldre equacions diferencials ordinàries controlades de la forma

$$dY_t = g(Y_t) dX_t, \quad Y_a = y_a \quad (2.35)$$

Considerem el camí  $X \in \mathcal{V}^1([a, b], \mathbb{R}^d)$ . Anomenem  $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$  a l'espai vectorial de les aplicacions lineals de  $\mathbb{R}^d$  a  $\mathbb{R}^e$ . Equivalentment,  $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$  es pot considerar com l'espai vectorial de les matrius reals de mida  $d \times e$ . Donat un camí  $Z : [a, b] \rightarrow \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$ , observem que podem definir la integral

$$\int_a^b Z_t dX_t \quad (2.36)$$

com un element de  $\mathbb{R}^e$  de la mateixa manera que la integral de línia usual. Donada una funció  $g : \mathbb{R}^e \rightarrow \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$  i un camí  $Y : [a, b] \rightarrow \mathbb{R}^e$ , diem que  $Y$  resol l'equació diferencial controlada 2.35 quan  $\forall t \in [a, b]$  resol l'equació de Volterra

$$Y_t = y_a + \int_a^t g(Y_s) dX_s. \quad (2.37)$$

La funció  $g$  en l'expressió anterior sovint s'anomena una col·lecció de camps vectorials conduïts, el camí  $X$  s'anomena control o conductor, i  $Y$  s'anomena solució o resposta.

Un procediment típic per a obtenir una solució de 2.37 és mitjançant iteracions de Picard. Donat un camí arbitrari  $Y : [a, b] \rightarrow \mathbb{R}^e$ , definim un nou camí  $F(Y) : [a, b] \rightarrow \mathbb{R}^e$  mitjançant

$$F(Y)_t = y_a + \int_a^t g(Y_s) dX_s. \quad (2.38)$$

$Y$  és una solució de 2.37 si i només si és un punt fix de  $F$ . Considerem la successió de camins  $Y_t^n = F(Y^{n-1})_t$  amb un camí arbitrari inicial  $Y_t^0$  (sovint pres com el camí constant  $Y_t^0 = y_a$ ). Sota certes condicions adequades, es pot demostrar que  $F$  té un únic punt fix  $Y$  i que  $Y_t^n \rightarrow Y$  per  $n \rightarrow \infty$  en la norma  $\|\cdot\|_{1,[a,b]}$  (*teorema del Picard-Lindelöf*).

Suposem ara que  $g : \mathbb{R}^d \rightarrow \mathcal{L}(\mathbb{R}^d, \mathbb{R}^e)$  és una aplicació lineal (també podem tractar  $g$  com a una aplicació lineal  $\mathbb{R}^d \rightarrow \mathcal{L}(\mathbb{R}^e, \mathbb{R}^e)$ , on  $\mathcal{L}(\mathbb{R}^e, \mathbb{R}^e)$  és l'espai de totes les matrius reals  $e \times e$ ). Comencem les iteracions de Picard amb el camí constant inicial  $Y_t^0 = y_a$  per a tot  $t \in [a, b]$ . Denotem per  $I_e$  la matriu identitat a  $\mathcal{L}(\mathbb{R}^e, \mathbb{R}^e)$ , podem expressar les iteracions de  $F$  de la següent manera:

$$\begin{aligned} Y_t^0 &= y_a, & (2.39) \\ Y_t^1 &= y_a + \int_a^t g(Y_s^0) dX_s = \left[ \int_a^t dg(X_s) ds + I_e \right] (y_a), \\ Y_t^2 &= y_a + \int_a^t g(Y_s^1) dX_s = \left[ \int_a^t \int_a^s dg(X_u) dg(X_s) + \int_a^t dg(X_s) ds + I_e \right] (y_a), \\ &\vdots \\ Y_t^n &= y_a + \int_a^t g(Y_s^{n-1}) dX_s = \left[ \sum_{k=1}^n \int_{a < t_1 < \dots < t_k < t} dg(X_{t_1}) \dots dg(X_{t_k}) + I_e \right] (y_a), \\ &\vdots \end{aligned}$$

Com que  $\mathcal{L}(\mathbb{R}^e, \mathbb{R}^e)$  és un àlgebra de matrius, cada quantitat

$$\int_{a < t_1 < \dots < t_k < t} dg(X_{t_1}) \dots dg(X_{t_k}) \quad (2.40)$$

es pot definir naturalment com un element de  $\mathcal{L}(\mathbb{R}^e, \mathbb{R}^e)$ , que, com es pot comprovar, queda completament determinat (de manera lineal) la signatura truncada de nivell  $k$   $S(X)_{a,t}$  amb  $t \in [a, b]$ .

Pel que en podem extreure que la solució  $Y_t$  queda completament determinada per la signatura  $S(X)_{a,t}$  per a cada  $t \in [a, b]$ . En particular, si les signatures de dos conductors  $X_1$  i  $X_2$  coincideixen en el temps  $t \in [a, b]$ , és a dir,  $S(X_1)_{a,t} = S(X_2)_{a,t}$ ,

llavors les solucions corresponents a 2.37 també coincideixen en el temps  $t$  per a qualsevol elecció dels camps vectorials lineals  $g$ .

Un resultat menys evident, és que el mateix és cert per als camps vectorials no lineals  $g$ . Aquest resultat va ser obtingut per primer cop per K.T. Chen per a una certa classe de camins suaus a trossos, i recentment ha estat ampliat per Hambly i Lyons [8] per a camins de variació finita, i per Boedihardjo, Geng, Lyons i Yang [15] per a camins geomètrics rugosos (*rough paths*) en un marc de camins completament no suaus en què la signatura encara està ben definida. Aquesta darrera classe de camins és d'un interès especial en l'anàlisi estocàstic.

### 2.4.7 Declivi Factorial

Els ordinadors només poden emmagatzemar un nombre finit de termes de la seqüència, de manera que només podem treballar amb la signatura truncada. Per tant, tot i que la signatura d'un camí descriu bé aquest camí, la signatura truncada podria no fer-ho, ja que els termes que s'eliminen podrien contenir una gran quantitat d'informació. No obstant, el resultat següent mostra que aquest no és el cas, que els termes de la signatura disminueixen factorialment en tamany a mesura que l'ordre augmenta.

Aquest teorema el podem trobar demostrat com a lema a [14].

**Teorema 2.43.** *Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$  un camí continu de variació acotada. Aleshores, donats  $1 \leq i_1 \leq \dots \leq i_n \leq d$ ,*

$$\left\| \int_{a < t_1 < \dots < t_n < b} dX_{t_1}^{i_1} \dots dX_{t_n}^{i_n} \right\| \leq \frac{\|X\|_1^n}{n!}$$

on

$$\|X\|_1 := \sup_{t_i} \sum_i |X_{t_{i+1}} - X_{t_i}|,$$

i el suprem es pren sobre totes les particions de  $[a, b]$ .

Per tant, si la signatura d'un camí el descriu bé, la signatura truncada també ho farà i pot ser usada com a característica.

### 2.4.8 Unicitat de la signatura

Una qüestió que ens podem plantejar és si un camí està totalment determinat per la seva signatura, es despren tant de les identitats de Chen i del producte mescla com de les propietats d'inversió en el temps i d'invariància sota reparametritzacions, que la resposta en general és no. Així, per exemple, per la propietat d'invariància respecte parametritzacions, no es pot recuperar de la signatura d'un camí, la velocitat a la que s'estava recorrent, tampoc podem, per exemple, diferenciar la signatura

d'un camí constant trivial de la d'un camí concatenat amb el seu invers en el temps.

No obstant, aquesta és fonamentalment la única informació que es perd del camí. Per a un camí  $X$  que no es creua a si mateix, la signatura descriu completament les propietats geomètriques del camí, més concretament, els punts pels que passa i en quin ordre passa per ells. Això últim es formalitzat a [8] amb la noció de camí *Tree-Like*.

**Teorema 2.44.** *Sigui  $X : [a, b] \rightarrow \mathbb{R}^d$  un camí continu de variació acotada. Aleshores  $S(X)_{a,b}$  determina  $X$  fins a una equivalència *Tree-Like*.*

Això, vol dir que la signatura determina camins fins a seccions en les que el camí es creua a si mateix. Per tant, la signatura únicament determina camins amb almenys una component monòtona, ja que aquesta impedeix que un camí es creui a si mateix.

**Exemple 2.45.** Un camí de la forma  $Y : [a, b] \rightarrow \mathbb{R}^d$ ,  $Y(t) = (t, X_t)$  queda únicament determinat per la seva signatura de la manera en que hem descrit anteriorment.

#### 2.4.9 Relació amb la Teoria de *Rough Paths*

La nostra discussió de la signatura s'ha limitat fins ara a camins que són diferenciables a trossos o més en general de variació acotada. Aquesta restricció era necessària per a assegurar-se que les integrals iterades del camí existissin com a integrals de Riemann-Stieltjes. De manera més general, es poden definir les integrals iterades d'un camí utilitzant la integral de Young per a qualsevol camí de variació finita  $p$  amb  $1 \leq p < 2$ . La integral de Young va més enllà de la definició clàssica de la integral i ja és capaç de cobrir una classe de camins substancialment més irregulars que aquells de variació acotada. Un problema que sorgeix per als camins de variació  $p$ -infinita per a tot  $p < 2$  (que és una situació d'un gran interès en l'anàlisi estocàstica a causa del fet que els camins de mostra del moviment Brownià tenen gairebé segur variació  $p$  finita si i només si  $p > 2$ ) és que es pot demostrar que no hi ha una noció ben definida d'una integral iterada per a aquests camins. Cal dir, però, que això no és degut a cap limitació tècnica de la integral de Young, sinó més aviat al fet que no hi ha un candidat únic per a les integrals iterades. És a dir, no hi ha necessàriament només una manera única de definir les integrals iterades.

Una de las observacions clau de Terry Lyons en la seva introducció de *rough paths* va ser que si es defineixen les primeres  $p$  integrals iterades d'un camí  $X$  de variació finita  $p$ , aleshores hi ha una manera única d'obtenir la resta d'integrals iterades i, per tant, la signatura de  $X$ .

### 3 Transformació d'un flux de dades en camí

Suposem que tenim un conjunt de dades (o flux de dades), el que volem és transformar aquesta successió discreta en una aplicació contínua, és a dir, en un camí, incrustant els punts (dades) en ella. A continuació presentem diferents maneres d'incrustar dades en camins continus. A la pràctica, quina d'aquestes maneres s'emplea depèn del que ens interessa més en cada cas.

Primer donem una definició més formal dels següents conceptes.

**Definició 3.1.** *Un conjunt (o flux) de dades  $\Omega \subset \mathbb{R}^d$  és una seqüència, és a dir, un element de*

$$\mathcal{S}(\mathbb{R}^d) := \{(x_1, \dots, x_k) \mid k \in \mathbb{N}, x_i \in \mathbb{R}^d, i \in \{1, \dots, k\}\}. \quad (3.1)$$

*Per a cada  $i \in \{1, \dots, k\}$ , diem que  $x_i$  és una dada.*

**Definició 3.2.** *La longitud d'una seqüència  $\Sigma \in \mathcal{S}(\mathbb{R}^d)$  és la seva cardinalitat:  $\#(\Sigma)$ .*

**Definició 3.3.** *Un conjunt de dades etiquetat  $\Omega \subset \mathbb{R}^d \times \mathbb{R}$  és una seqüència (o flux), és a dir, un element de*

$$\mathcal{S}(\mathbb{R}^d \times \mathbb{R}) := \{((x_1, y_1), \dots, (x_k, y_k)) \mid k \in \mathbb{N}, (x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}, i \in \{1, \dots, k\}\}. \quad (3.2)$$

*Per a cada  $i \in \{1, \dots, k\}$ , diem que  $x_i$  és una dada i  $y_i$  la seva etiqueta.*

#### 3.1 Incrustació mitjançant interpolació lineal a trossos

En aquest cas, donat el conjunt de dades, unim les dades amb línies rectes, és a dir amb el camí amb distància euclidiana menor, en un determinat ordre.

Si les dades són ordenades, per exemple, sèries temporals, tenim  $\Sigma = ((t_1, x_1), \dots, (t_k, x_k))$ , amb  $(t_i, x_i) \in \mathbb{R} \times \mathbb{R}^{d-1}$ ,  $1 \leq i \leq k$ . Aleshores,  $\forall t_i, t_{i+1} \in \{t_1, \dots, t_k\}$  interpolem amb el polinomi de Newton

$$f(t|t_i; t_{i+1}) = f(t_i) + \frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}(t - t_i), \quad (3.3)$$

on  $f(t_i) = x_i$  i  $f(t_{i+1}) = x_{i+1}$ . És a dir, el graf de l'aplicació és el camí (geomètric) que uneix cada punt  $(t_i, x_i)$  amb el punt  $(t_{i+1}, x_{i+1})$  en línia recta.

**Exemple 3.4.** Sigui  $\{(t, X)\} = \{(t_i, X_i)\}_{i=0}^4 = \{(0, 1), (1, 4), (2, 2), (3, 6)\}$  a la figura 5 podem observar el resultat d'incrustar aquest flux amb interpolació lineal a trossos



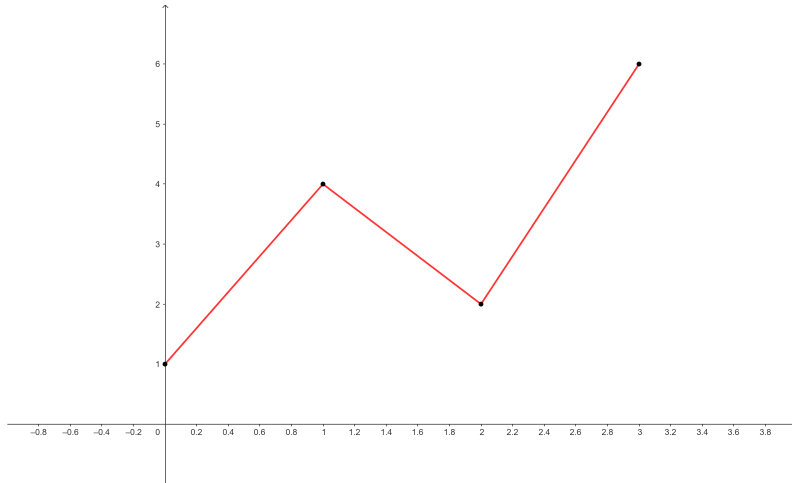


Figura 5: Interpolació lineal a trossos de  $\{(t_i, X_i)\}_{i=0}^4$

### 3.2 Incrustació mitjançant interpolació rectilínia

En aquest cas unim cada parell de punts  $\{(t_i, x_i), (t_{i+1}, x_{i+1})\}$  amb el camí més curt (de menor longitud) amb distància Manhattan, és a dir, la distància  $d(\cdot)$  tal que  $d(x, y) = d((x_1, \dots, x_d), (y_1, \dots, y_d)) = \sum_{i=1}^d |x_i - y_i|$ ,  $\forall x, y \in \mathbb{R}^d$ , recorrent sempre primer la recta inscrita a l'hiperplà  $\{t = t_i\}$  (o a qualsevol altre, sempre i quan sigui el mateix durant tot el procés).

**Exemple 3.5.** Prenent el mateix flux que a l'últim exemple, a la figura 6 podem observar el resultat de la incrustació mitjançant interpolació rectilínia.

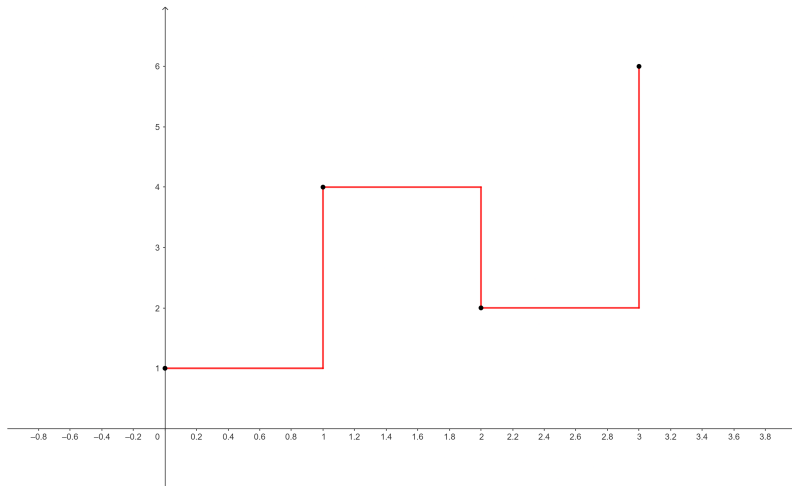


Figura 6: Interpolació rectilínia de  $\{(t_i, X_i)\}_{i=0}^4$

### 3.3 Suma cumulativa

Per a algunes aplicacions numèriques el flux de dades original, en ocasions, pot ser graficat de formes diferents. Una d'aquestes és la suma cumulativa o successió de

sumes parcials.

**Definició 3.6.** Donat  $\{X_i\}_{i=0}^n$  definim la seva suma cumulativa o successió de sumes parcials com

$$CS(\{X_i\}_{i=0}^n) = \{S_1, \dots, S_n\}$$

on  $S_k = \sum_{i=0}^k X_i$ , per a  $k \in \{1, \dots, n\}$ .

**Exemple 3.7.** Tornem a prendre el flux dels dos últims exemples. Aleshores  $CS(X) = \{1, 5, 7, 13\}$ . A la figura 7 podem veure el resultat d'aplicar les dues interpolacions anteriors a  $\{(t, CS(X))\}$

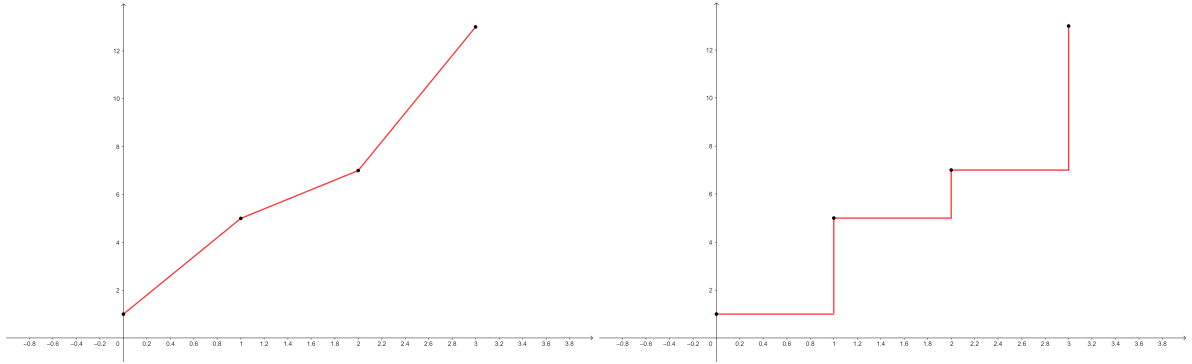


Figura 7: Interpolacions lineal a trossos i rectilinia de  $\{(t, CS(X))\}$

### 3.4 Incrustació mitjançant augmentacions

Un altre tipus de transformacions més complexes passen per augmentar el flux  $\Omega \subset \mathbb{R}^d$  abans d'interpol·lar entre els punts resultants per a incrustar-los en un camí. Més precisament, considerem una aplicació fixada  $\theta : \mathbb{R}^d \rightarrow \mathbb{R}^e$ , i interpo·lem el conjunt de dades augmentat  $\theta(\Omega) \subset \mathbb{R}^e$ . Algunes de les augmentacions explícites més comunment utilitzades són: l'augmentació *Lead-Lag*, l'augmentació *Time*, l'augmentació de *No Future Pause* i l'augmentació *Invisibility Reset*.

**Definició 3.8.** L'augmentació *Lead-Lag* és l'aplicació  $T : S(\mathbb{R}^d) \rightarrow S(\mathbb{R}^d \times \mathbb{R}^d)$  definida per

$$T(x_1, \dots, x_d) := ((x_1, x_1), (x_2, x_1), (x_2, x_2), \dots, (x_j, x_j), (x_{j+1}, x_j), \dots, (x_d, x_d)) \quad (3.4)$$

L'augmentació *Lead-Lag* pren una seqüència de longitud  $d$  i la converteix en una seqüència de longitud  $2d$ , dividint la original en dues còpies etiquetades, *flux del futur* i *flux de passat*. Hi ha un retard entre quan s'actualitza el futur i quan s'actualitza el passat posteriorment. La longitud del retard pot variar, donant lloc a un *Lead-Lag* on el nombre de passos entre l'actualització del futur i la del passat és d'almenys dos. A més, es poden enregistrar més d'una seqüència del passat, sent cada una d'elles actualitzada amb un nombre diferent de passos de retard entre l'actualització de la seqüència del futur i la seva pròpia actualització.

**Exemple 3.9.** L'augmentació Lead-Lag, amb 2 passos de retard és l'aplicació  $T : S(\mathbb{R}^d) \rightarrow S(\mathbb{R}^d \times \mathbb{R}^d)$  definida per

$$T(x_1, \dots, x_d) := ((x_1, x_1), (x_2, x_1), (x_3, x_1), (x_3, x_2), (x_4, x_2), (x_5, x_2), \dots, (0, x_d)) \quad (3.5)$$

L'augmentació *Lead-Lag* definida en primer lloc és l'augmentació *Lead-Lag* amb un pas de retard.

**Definició 3.10.** Donats  $k$  fluxos del passat, definim l'augmentació *No Future Pause* com l'aplicació  $T : S(\mathbb{R}^d) \rightarrow S(\mathbb{R}^{d(k+1)})$  que envia el flux  $(x_1, \dots, x_d) \in S(\mathbb{R}^d)$  a

$$\left\{ \begin{pmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} x_2 \\ x_1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} x_{k+1} \\ x_k \\ x_{k-1} \\ \vdots \\ x_2 \\ x_1 \end{pmatrix}, \begin{pmatrix} x_{k+2} \\ x_{k+1} \\ x_k \\ \vdots \\ x_3 \\ x_2 \end{pmatrix}, \dots, \begin{pmatrix} x_d \\ x_{d-1} \\ x_{d-2} \\ \vdots \\ x_{d-k+1} \\ x_{d-k} \end{pmatrix}, \begin{pmatrix} 0 \\ x_d \\ x_{d-1} \\ \vdots \\ x_{d-k+2} \\ x_{d-k+1} \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ x_d \\ x_{d-1} \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ x_d \end{pmatrix} \right\} \quad (3.6)$$

a  $S(\mathbb{R}^{d(k+1)})$ .

**Exemple 3.11.** Considerem el flux  $\Omega = \{1, 4, 2, 5, 3\}$ , l'augmentació *Lead-Lag* envia a aquest flux al flux

$$\{(1, 1), (4, 1), (4, 4), (2, 4), (2, 2), (5, 2), (5, 5), (3, 5), (3, 3)\} \in S(\mathbb{R}^2)$$

Si el retard és de dos passos, obtenim el flux

$$\{(1, 1), (4, 1), (2, 1), (2, 4), (5, 4), (3, 4), (3, 2), (0, 2), (0, 2), (0, 5), (0, 5), (0, 5), (0, 3)\} \in S(\mathbb{R}^2)$$

Podem doncs, interpolat aquests fluxos en camins a  $\mathbb{R}^2$ .

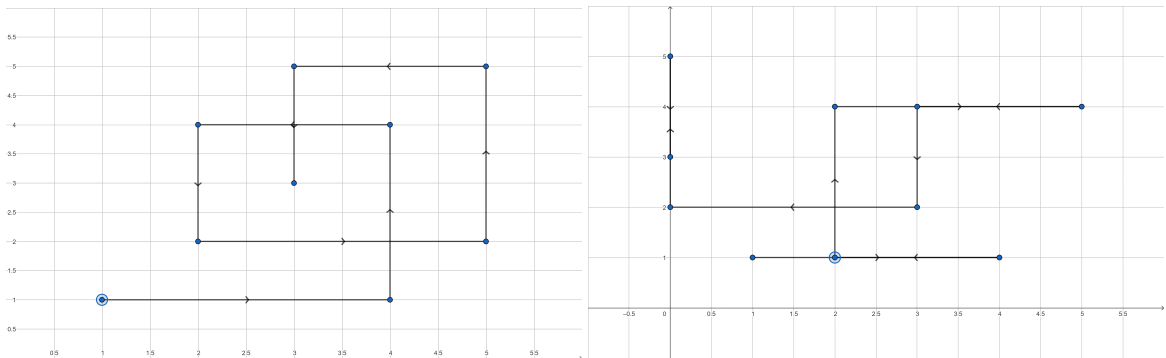


Figura 8: Transformacions Lead-Lag amb 1 i 2 passos de retard resp.

També es poden ajuntar les dues augmentacions en una mateixa augmentació, enviant el flux  $\Omega \subset S(\mathbb{R}^2)$  a flux de  $S(\mathbb{R}^3)$

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 5 \\ 2 \end{pmatrix}, \begin{pmatrix} 5 \\ 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} \right\}$$

Si apliquem la augmentació *No Future Past* amb  $k = 2$ , obtenim el flux

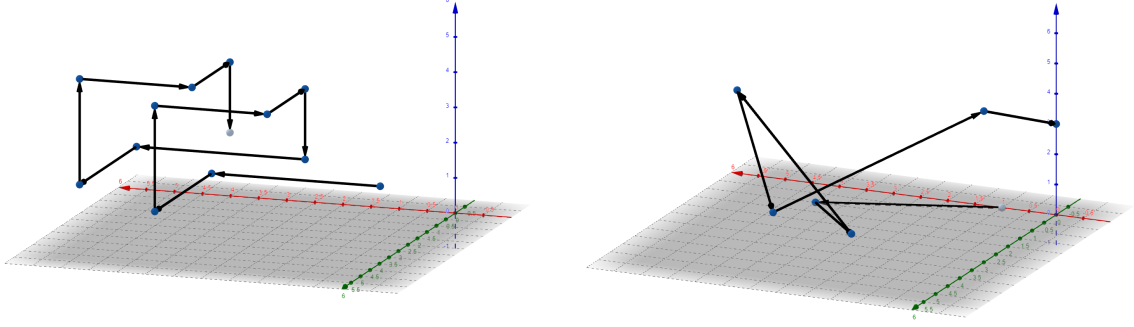


Figura 9: Exemples d'interpolacions Lead-Lag i No Future Past

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 5 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} \right\} \in S(\mathbb{R}^3)$$

**Definició 3.12.** Definim l'augmentació *Time* com l'aplicació  $T : S(\mathbb{R}^d) \rightarrow S(\mathbb{R}^d \times \mathbb{R})$  definida com

$$T(x_1, \dots, x_d) := ((x_1, t_0), \dots, (x_j, t_j), \dots, (x_d, t_d)) \quad (3.7)$$

per a una seqüència estrictament creixent de temps  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$ .

**Observació 3.13.** La seqüència estrictament creixent de temps  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$  assegura que el camí resultant té una component estrictament monotona.

Una altra variant d'aquesta augmentació és l'augmentació *Time-difference*:

**Definició 3.14.** L'augmentació *Time-difference* és l'aplicació  $T : S(\mathbb{R}^d) \rightarrow S(\mathbb{R}^d \times \mathbb{R})$  definida com

$$T(x_1, \dots, x_d) := ((x_1, t_0), \dots, (x_j, t_j - t_{j-1}), \dots, (x_d, t_d - t_{d-1})) \quad (3.8)$$

per a una seqüència estrictament creixent de temps  $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$ .

**Observació 3.15.** Observem que tant l'augmentació *Time* com l'augmentació *Time-difference* envien fluxos de longitud  $k$  a fluxos de longituds  $k$ .

**Exemple 3.16.** Considerem el flux  $\Omega = \{1, 5, 2, 9, 7, 6\} \in S(\mathbb{R})$  i la seqüència estrictament creixent  $0, 1, 3, 6, 8, 12$ . Aleshores

$$T_{time}(\Omega) = \{(1, 0), (5, 1), (2, 3), (9, 6), (7, 8), (6, 12)\} \in S(\mathbb{R}^d) \quad (3.9)$$

i

$$T_{time-diff}(\Omega) = \{(1, 0), (5, 1), (2, 2), (9, 3), (7, 2), (6, 4)\} \in S(\mathbb{R}^d) \quad (3.10)$$

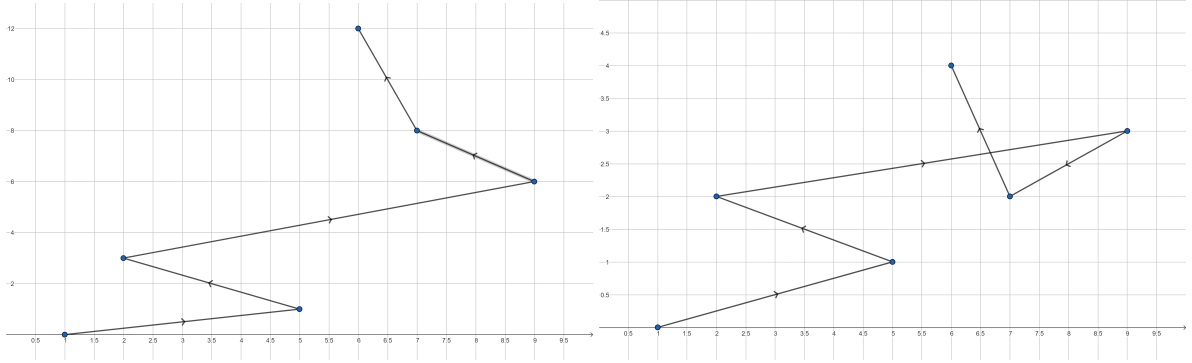


Figura 10: Interpolacions time i time-difference

La última augmentació que definirem és la següent:

**Definició 3.17.** L'augmentació *Invisibility Reset* és l'aplicació definida com  $T : S(\mathbb{R}^d) \rightarrow S(\mathbb{R}^d \times \mathbb{R})$  definida com

$$T(x_1, \dots, x_d) := ((x_1, 1), \dots, (x_j, 1), \dots, (x_d, 1), (x_d, 0), (0, 0)) \quad (3.11)$$

**Observació 3.18.** L'augmentació *Invisibility Reset* pren un flux de longitud  $d$  i l'envia a un de longitud  $d+2$ . El nou flux resultant conté la informació de la norma dels elements del flux original  $x_1, \dots, x_d$  de forma invariant per translació.

**Exemple 3.19.** Donats el fluxos  $\Omega_1 = \{1, 3, 4, 8, 9\} \in S(\mathbb{R})$  i  $\Omega_2 = \{(1, 2), (3, 4), (4, 6), (5, 9), (7, 10)\} \in S(\mathbb{R}^2)$ . Aplicant l'augmentació *Invisibility Reset* obtenim:

$$T(\Omega_1) = \{(1, 1), (3, 1), (4, 1), (8, 1), (9, 1), (9, 0), (0, 0)\} \in S(\mathbb{R}^2) \quad (3.12)$$

$$T(\Omega_2) = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 4 \\ 6 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 \\ 9 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ 10 \\ 1 \end{pmatrix}, \begin{pmatrix} 7 \\ 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\} \in S(\mathbb{R}^3) \quad (3.13)$$

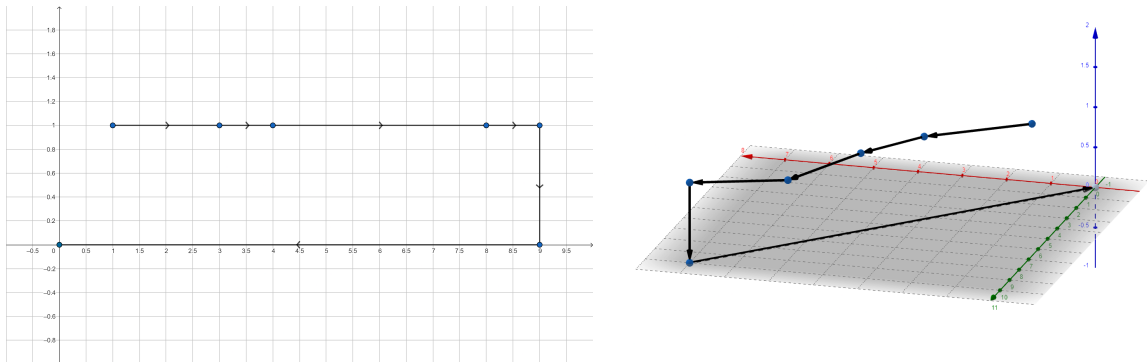


Figura 11: Interpolacions Invisibility Reset

### 3.5 Exemple explícit de càlcul

A continuació calculem algunes signatures basant-nos en la primera definició de signatura.

Considerem el camí de la figura 5, podem veure el seu increment i la seva àrea

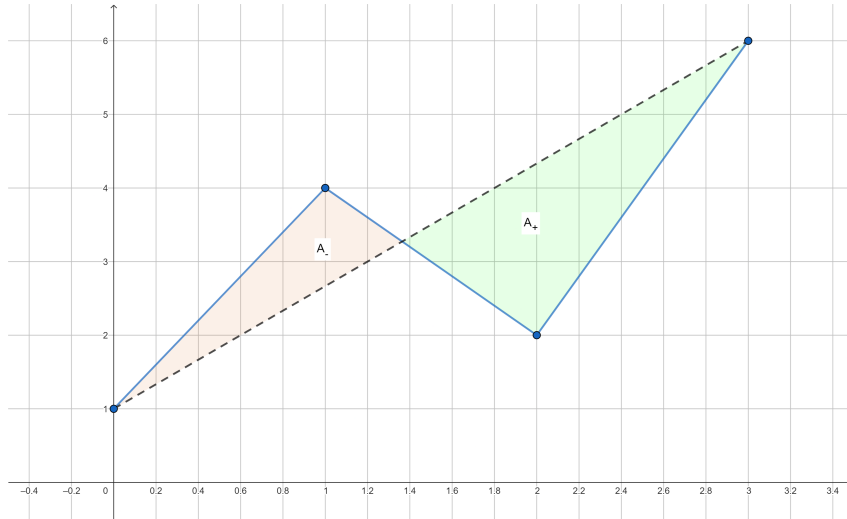


Figura 12: Camí amb l'àrea signada

signada a la figura 12. La seva signatura truncada de nivell 2 és:

$$S(X) = (1, 3, 5, 4.5, 8.5, 6.5, 12.5) = (1, S^{(1)}, S^{(2)}, S^{(1,1)}, S^{(1,2)}, S^{(2,1)}, S^{(2,2)})$$

i

$$\log S(X) = (3, 5, 1) = (S^{(1)}, S^{(2)}, S^{(1,2)})$$

l'últim terme de  $\log S(X)$  es calcula de  $\frac{1}{2}(S^{(1,2)} - S^{(2,1)})$  i és l'àrea total entre els extrems. Observem que l'àrea  $A_+$  és més gran que l'àrea  $A_-$  doncs l'àrea total es positiva. A la figura 12 podem entendre el significat geomètric dels termes de segon ordre  $S^{(1,2)}$  i  $S^{(2,1)}$ .

Si canviem l'ordre d'integració sobre el camí en els termes  $S^{(1,2)}$  i  $S^{(2,1)}$  obtenim dues àrees que es completen l'una a l'altra i sumen l'àrea total d'un rectangle amb longituds laterals  $X^1$  i  $X^2$ . Aquest senzill significat geomètric és la relació de producte mescla:

$$S^{(1)} \cdot S^{(2)} = S^{(1,2)} + S^{(2,1)}$$

$$3 \cdot 5 = 8.5 + 6.5.$$

La interpretació geomètrica dels termes d'ordre superior és menys intuïtiva pel que la ometem.

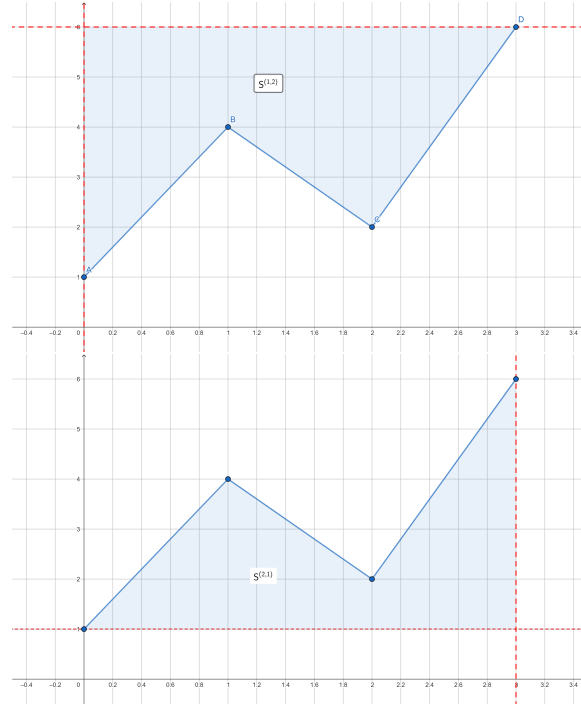


Figura 13: àrees  $S^{(1,2)} = 8.5$  i  $S^{(2,1)} = 6.5$

## 4 Signatura de fluxos de dades de caire financer

En el mon de les finances sovint es presenten dades en forma de sèries temporals, es a dir presenten trajectories indexades per un nombre finit de punts en el temps:  $(\hat{X}_{t_i})_{i=0}^N$  a  $\mathbb{R}^d$ .

Una forma de definir la signatura d'un flux dades d'aquest tipus es amb les integrals iterades del camí obtingut per interpolació lineal a trossos, com hem vist a l'apartat 3.1.. És a dir, introduïm el camí continu

$$X_u = \hat{X}_{t_i} + \frac{u - t_i}{t_{i+1} - t_i} (\hat{X}_{t_{i+1}} - \hat{X}_{t_i}), \quad u \in [t_i, t_{i+1}],$$

i definim la signatura del flux  $(\hat{X}_{t_i})_{i=0}^N$  com

$$S(\hat{X})_{t_0, t_N} := S(X)_{0, 2N}.$$

### 4.1 La transformació *time-joined* i la signatura d'una sèrie temporal

Una altra manera de transformar un flux de dades  $(\hat{X}_{t_i})_{i=0}^N = (t_i, \hat{X}_{t_i})_{i=0}^N$  en un camí continu és utilitzar la transformació *time-joined* que es defineix com  $R : [0, 2N] \rightarrow \mathbb{R}^+ \times \mathbb{R}$  tal que:

$$X_t^{TJ} = R(t) = \begin{cases} (t_0, \hat{X}_{t_0} t) & \text{per a } t \in [0, 1], \\ (t_i + (t_{i+1} - t_i)(t - 2i - 1), \hat{X}_{t_i}) & \text{per a } t \in [2i + 1, 2i + 2], \\ (t_{i+1}, \hat{X}_{t_i} + (\hat{X}_{t_{i+1}} - \hat{X}_{t_i})(t - 2i - 2)) & \text{per a } t \in [2i + 2, 2i + 3]. \end{cases}$$

per a  $i = 0, 1, \dots, N - 1$ . A l'annex B trobem un programa que computa aquesta transformació per a un flux de dades bidimensional donat.

**Observació 4.1.** La funció contínua  $R$  el que fa simplement és mantenint el valor  $\hat{X}_{t_i}$  durant l'interval de temps  $[t_i, t_{i+1})$ . Quan les noves dades  $\hat{X}_{t_{i+1}}$  arriben al temps  $t_{i+1}$ , hi ha un salt instantani de  $\hat{X}_{t_i}$  a  $\hat{X}_{t_{i+1}}$ . Afegim un punt més, 0, al temps  $t_0$ , a la sèrie temporal  $(\hat{X}_i)_{i=0}^N$  per tal de convertir-la en una nova sèrie temporal, de manera que la signatura de  $(\hat{X}_i)_{i=0}^N$  pugui determinar unívocament  $(\hat{X}_i)_{i=0}^N$ .

La signatura de  $\{(t_i, \hat{X}_i)\}_{i=0}^N$  es defineix com la signatura de la transformació *time-joined* de  $\{(t_i, \hat{X}_i)\}_{i=0}^N$ ,  $N \in \mathbb{N}$ . Es pot demostrar que la signatura d'una sèrie temporal la determina completament [9]

**Definició 4.2.** Sigui  $\{(t_i, \hat{X}_i)\}_{i=0}^N$  una sèrie temporal incrustada en el camí *time-joined*  $R$ . La signatura de la sèrie temporal  $\{(t_i, \hat{X}_i)\}_{i=0}^N$  es defineix com la signatura del camí  $\{R(s)\}_{s \in [0, 2N]}$ , i la denotem per  $S(\{(t_i, \hat{X}_i)\}_{i=0}^N)$ , on  $1 \leq m < n \leq N$  i  $m, n \in \mathbb{N}$ .

## 4.2 La signatura i els moments estadístics

El fet de combinar diferents fluxos de dades  $X \equiv \{X_i\}$  en un sol ens permet calcular moments estadístics i correlacions entre els diferents fluxos.

A continuació veiem la relació entre els moments estadístics d'un conjunt de dades  $X$  en termes de la seva signatura.

## 4.3 Variació quadràtica

Tot i que els termes de la signatura mesuren diferents quantitats associades al camí, la variació quadràtica del procés no es captura directament. Aquesta quantitat és molt rellevant en aplicacions al camp de les finances, és per tant d'interès incorporar-la en la signatura. Això últim es pot fer mitjançant la augmentació *Lead-Lag* presentada a la secció 3.

Donat un flux  $(\hat{X}_{t_i})_{i=0}^N \subset S(\mathbb{R}^d)$  el seu flux *Lead-transformat*  $(\hat{X}_j^{lead})_{j=0}^{2N}$  és defineix com

$$\hat{X}_j^{lead} = \begin{cases} \hat{X}_{t_i}, & \text{si } j = 2i \\ \hat{X}_{t_i}, & \text{si } j = 2i - 1 \end{cases}$$

Definim també el seu flux *Lag-transformat*,  $(\hat{X}_j^{lag})_{j=0}^{2N}$ , com

$$\hat{X}_j^{lag} = \begin{cases} \hat{X}_{t_i}, & \text{si } j = 2i \\ \hat{X}_{t_i}, & \text{si } j = 2i + 1 \end{cases}$$

Definim doncs el flux *Lead-Lag-transformat*, que pren valors a  $\mathbb{R}^{2d}$  com

$$(\hat{X}_j^{lead-lag})_{j=0}^{2N} = (\hat{X}_j^{lead}, \hat{X}_j^{lag})_{j=0}^{2N}$$



Observem que  $X^{lag}$  correspon a una reparametrització del temps de camí  $X$ :

$$S(\hat{X})_{t_0, t_N} = S(\hat{X}^{lag})_{0, 2N}$$

Similarment

$$S(\hat{X})_{t_0, t_N} = S(\hat{X}^{lead})_{0, 2N}$$

A més l'àrea signada entre la  $i$ -èssima component del camí *lead*-transformat i la component  $j$ -èssima del camí *lag*-transformat equival a la covariació de les trajectòries  $\hat{X}^i$  i  $\hat{X}^j$ . A la pràctica s'utilitza una transformació *Lead-Lag* parcial, prenent la transformació *lead* del flux d'entrada i l'ajuntem amb la transformació *lag* d'aquells components pels quals la variació quadràtica és rellevant.

Una altra forma de definir més explícitament la transformació *Lead-Lag* d'un flux de dades  $(\hat{X})_{i=0}^N = \{(t_i, \hat{X}_{t_i})_{i=0}^N\}$  és la següent:

$$X_t^{Lead-Lag} = \begin{cases} (X_{t_i}, X_{t_{i+1}}) & \text{per a } t \in [2i, 2i + 1], \\ (X_{t_i}, X_{t_{i+1}} + 2(t - (2i + 1))(X_{t_{i+2}} - X_{t_{i+1}})) & \text{per a } t \in [2i + 1, 2i + \frac{3}{2}], \\ (X_{t_{i+2}} + 2(t - (2i + \frac{3}{2}))(X_{t_{i+1}} - X_{t_i}), X_{t_{i+2}}) & \text{per a } t \in [2i + \frac{3}{2}, 2i + 2]. \end{cases}$$

per a  $t \in [0, 2N]$ . A l'annex A trobem programa que computa aquesta transformació per a un flux de dades bidimensional donat.

**Exemple 4.3.** Sigui

$$[X]_t = \sum_{i=0}^{N-1} (X_{t_{i+1}} - X_{t_i})^2$$

la variació quadràtica del camí construït amb el flux  $(\hat{X}_{t_i})_{i=0}^N \subset S(\mathbb{R}^2)$ . Aleshores, podem escriure

$$A_{lead-lag} = \frac{1}{2}[X]_t$$

## 4.4 La suma cumulativa d'una successió

Vegem algunes propietats de camins obtinguts incrustant punts utilitzant sumes cumulatives.

La suma considerem una seqüència unidimensional  $X = \{\hat{X}_i\}_{i=0}^N$ . La seva suma cumulativa és

$$\tilde{X} = S_1, \dots, S_N$$

On  $S_k = \sum_{i=0}^k \hat{X}_i$  amb  $k \in \{1, \dots, N\}$ .

Ajuntant-la amb una component temporal  $\{t_i\}_{i=0}^N$  obtenim un camí bidimensional  $X_t$  amb interpolació rectilínia.

Si augmentem la seqüència original  $\{\hat{X}_i\}$  afegint el valor 0 al principi i trunquem la signatura de la nova seqüència  $\{\tilde{X}\}$  al nivell  $L$ , podem determinar els moments

estadístics d'ordre  $k \leq L$  de la seqüència original.

És a dir passem de  $\{\hat{X}_i\}_{i=1}^N$  a  $\{0, \{\tilde{X}_i\}_{i=0}^N\}$  a  $CS(\{\hat{X}_i\}_{i=0}^N) = \{\tilde{X}\}$ , li podem aplicar la augmentació *Lead-Lag* i calcular fins al segon nivell de la signatura, obtenint quantitats proporcionals als moments de primer i segon ordre, és a dir, la mitjana i la variància.

$$\Delta\tilde{X} = \sum_{i=0}^N \hat{X}_i$$

$$QV(\tilde{X}) = \sum_{i=0}^{N-1} (\tilde{X}_{i+1} - \tilde{X}_i)^2 = \sum_{i=0}^N (\hat{X}_i)^2$$

I podem observar fàcilment que la mitjana i la variància (per a dades uniformement distribuïdes) respectivament són:

$$\bar{X} = E[X] = \frac{1}{N} \sum_{i=0}^N \hat{X}_i = \frac{\Delta\tilde{X}}{N}$$

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2 = \frac{1}{N} \left( QV(\tilde{X}) - \frac{1}{N} (\Delta\tilde{X})^2 \right)$$

La Signatura trucada de nivell 2 del camí serà

$$S(\tilde{X})|_{L=2} = \left( 1, S^{(1)}(\tilde{X}), S^{(2)}(\tilde{X}), S^{(1,1)}(\tilde{X}), S^{(1,2)}(\tilde{X}), S^{(2,1)}(\tilde{X}), S^{(2,2)}(\tilde{X}) \right)$$

on

$$S^{(1)}(\tilde{X}) = S^{(2)}(\tilde{X}) = \tilde{X}_N - \tilde{X}_0 = \tilde{X}_N = \sum_{i=0}^N \hat{X}_i$$

$$S^{(1,1)}(\tilde{X}) = S^{(2,2)}(\tilde{X}) = \frac{(\tilde{X}_N - \tilde{X}_0)^2}{2} = \frac{(\tilde{X}_N)^2}{2} = \frac{1}{2} \left( \sum_{i=1}^N X_i \right)^2$$

$$\begin{aligned} S^{(1,2)}(\tilde{X}) &= \int_0^N \tilde{X}_t d\tilde{X}_t + \sum_{i=0}^N \int X_i dX_i = \int_0^N \sum_{i=0}^t X_i d \left( \sum_{i=0}^t X_i \right) + \tilde{X}_t + \sum_{i=0}^N \int X_i dX_i \\ &= \frac{1}{2} \left[ \left( \sum_{i=1}^N X_i \right)^2 + \sum_{i=1}^N X_i^2 \right] \end{aligned}$$

$$\begin{aligned} S^{(2,1)}(\tilde{X}) &= \int_0^N \tilde{X}_t d\tilde{X}_t - \sum_{i=0}^N \int X_i dX_i = \int_0^N \sum_{i=0}^t X_i d \left( \sum_{i=0}^t X_i \right) - \tilde{X}_t + \sum_{i=0}^N \int X_i dX_i \\ &= \frac{1}{2} \left[ \left( \sum_{i=1}^N X_i \right)^2 - \sum_{i=1}^N X_i^2 \right] \end{aligned}$$

On hem utilitzat 2.17 per a calcular  $S^{(1)}$  i  $S^{(2)}$  i 2.20 per a  $S^{(1,1)}$  i  $S^{(2,2)}$  i 2.3 de per a  $S^{(1,2)}$  i  $S^{(2,1)}$ .

Per tant, podem veure que a partir de la signatura truncada de nivell  $L = 2$  del camí resultant de la suma cumulativa i la incrustació *Lead-Lag* es poden obtenir els dos primers moments estadístics empírics, la variància i la mitjana:

$$\bar{X} = \frac{1}{N} S^{(1)}(\tilde{X})$$

$$Var(X) = -\frac{1}{N+1} \left( \frac{1}{N^2} S^{(1,2)}(\tilde{X}) + \frac{N-1}{N^2} S^{(2,1)}(\tilde{X}) \right)$$

Es pot demostrar també que els nivells d'ordre superior de la signatura determinen els moments d'ordre superior.

## 5 La signatura en l'aprenentatge automàtic

La idea principal d'utilitzar la transformació de la signatura per a problemes d'aprenentatge automàtic es dona per la seva capacitat d'extraure característiques distintives de les dades. Incrustar les dades en un camí i calcular la seva signatura ens proporciona informació important sobre les dades originals.

Les característiques extretes es poden utilitzar per a diversos tipus d'aplicacions a l'aprenentatge automàtic (o *machine learning*), incloent tant l'aprenentatge supervisat com l'aprenentatge no supervisat. Per exemple, es poden classificar sèries temporals o distingir clústers de dades. Un dels avantatges de l'extracció de característiques amb el mètode de la signatura és que aquesta és sensible a la forma geomètrica d'un camí, això últim, per exemple, ha portat a una aplicació exitosa del mètode de la signatura en el problema de reconeixement de caràcters xinesos [10]. Una de les aplicacions més conegudes i naturals del mètode de la signatura és en finances quantitatives, concretament en l'anàlisi de dades de sèries temporals. Les dades de sèries temporals representen dades seqüencials ordenades, que són un candidat ideal per crear un camí a partir de les dades, seguit de calcular la signatura i aplicar algorismes d'aprenentatge automàtic per a un anàlisi posterior. Qualsevol tipus de dades seqüencials ordenades en el temps encaixa naturalment en el marc de treball de la signatura. A més, si les dades d'entrada provenen de diverses fonts paral·leles, això donarà com a resultat un camí multidimensional. Un exemple d'aquest tipus de dades són les dades de panell (en econometria) o les dades longitudinals (en medicina, psicologia, bioestadística, etc.) que involucren observacions repetides de la mateixa quantitat durant períodes de temps.

### 5.1 Aprenentatge supervisat

Una classe molt àmplia de problemes en l'aprenentatge automàtic és l'aprenentatge supervisat. L'objectiu de l'aprenentatge supervisat és aprendre una funció que assigni una entrada (*input*) a una sortida (*output*). Això es fa utilitzant un conjunt de dades d'entrenament amb parells coneguts d'entrada-sortida (o *input-output*). És crucial fer una bona selecció d'entrades, conegudes com a *característiques*, per

construir un model a partir d'aquestes característiques. Aquestes característiques haurien de descriure l'objecte que s'està estudiant de manera precisa, però al mateix temps, característiques amb una dimensió molt alta provoquen problemes computacionals i de sobreajust.

Suposem, per exemple, que volguessim construir un classificador que predigués si una persona és d'una o d'una altra ètnicitat determinada. La nacionalitat, pot ser una bona característica que podríem utilitzar, ja que els diferents països solen tenir alguna etnicitat majoritària, però el sexe obviament no ho és, ja que per a una determinada etnicitat, hi ha aproximadament, el mateix nombre d'homes que de dones. Per tant, el sexe no és una característica que descriu de manera precisa la etnicitat d'un grup d'individus.

## 5.2 Les signatures com a característiques

La pregunta que ens plantegem aquí és: poden les signatures ser característiques efectives que es puguin utilitzar per a crear un model d'aprenentatge automàtic? La resposta es troba en la secció 2 d'aquest treball.

Sabem que donat un camí de variació acotada podem construir la seqüència d'integrals iterades que defineixen la signatura. A la subsecció 2.4.7 hem vist que, tot i que donada una d'aquestes seqüències d'integrals iterades no existeix un únic camí amb variació acotada la signatura del qual sigui aquesta seqüència, si hi ha una certa noció d'unicitat sota certes condicions de regularitat del camí, concretament, per aquells camins que no es creuen a si mateixos. Per això, la signatura és una bona descripció del camí.

No obstant, donada la impossibilitat de computar la signatura (ja que és una sèrie infinita), cal assegurar que la signatura truncada proporciona també una bona descripció del camí, per la propietat de declivi factorial esmentada a la subsecció 2.4.6 deduïm que això és així, ja que la importància dels termes de la signatura decau factorialment respecte al seu ordre.

## 5.3 Model d'aprenentatge automàtic basat en signatures

Podem formular els efectes dels fluxos de dades com a variable dependent d'un problema de regressió, amb una variable explicativa que és un camí a  $\mathcal{V}^p(J, \mathbb{R}^d)$ ,  $J \subset \mathbb{R}$  compacte. Suposem que tenim un conjunt d'entrenament de parells d'entrada-sortida  $\{(X_i, Y_i)\}_{i=0}^N$  on  $X_i$  és un camí de variació acotada i amb  $Y_i \in \mathbb{R}$ . Volem aprendre la funció que assigna cada entrada a la sortida corresponent. Per tant, assumirem que les entrades i les sortides estan relacionades mitjançant una funció desconeguda  $f$  de la forma

$$Y_i = f(X_i) + \varepsilon_i \quad (5.1)$$

On  $X_i \in \mathcal{V}^p(J, \mathbb{R}^d)$ ,  $Y_i \in \mathcal{V}^p([0, T], \mathbb{R}^e)$ ,  $\mathbb{E}[\varepsilon_i | X_i] = 0$ .

El següent teorema assegura que n'hi ha prou amb buscar funcions lineals.

**Teorema 5.1.** *Sigui  $\mathcal{V}^1([0, T], \mathbb{R}^d)$  l'espai de camins  $d$ -dimensionals de variació acotada definit sobre l'interval  $[0, T]$ . Sigui  $S(\mathcal{V}^1([0, T], \mathbb{R}^d)) := \{S(X) : X \in \mathcal{V}^1([0, T], \mathbb{R}^d)\}$  i  $S_1 \subset S(\mathcal{V}^1([0, T], \mathbb{R}^d))$  un conjunt compacte. Aleshores, donat  $\varepsilon > 0$  i una funció continua  $g : S_1 \rightarrow \mathbb{R}$ , existeix una funció lineal continua  $L$  tal que*

$$\|g(x) - L(x)\| < \varepsilon \quad \forall x \in S_1$$

La demostració d'aquest teorema es pot trobar a [11].

Per tant, la funció  $f$  pot ser aproximada per funcions lineals. Com que la signatura truncada es pot interpretar com a un vector, podriem aleshores aplicar-hi regressió lineal per a resoldre aquest problema.

Usant tècniques clàssiques de regressió no paramètrica per al cas de dimensionalitat finita, cal identificar conjunts de característiques específiques de les dades d'entrada (sortida) observades per a linearitzar la relació funcional entre elles. Com hem vist a la secció 2.4.7 la signatura és en certa forma única i la usarem com a característica, per tant buscarem  $f$  tal que

$$S(Y_i) = f(S(X_i)) + \varepsilon_i$$

Més en general, definim el següent model:

**Definició 5.2** (Model de la Signatura Esperada). *Siguin  $X$  i  $Y$  dos processos estocàstics que prenen valors a  $\mathbb{R}^d$  i  $\mathbb{R}^e$ , respectivament. Suposem que les signatures de  $X$  i  $Y$ , denotades per  $S(X)$  i  $S(Y)$ , estan ben definides q.s. Assumim que*

$$S(Y) = f(S(X)) + \varepsilon,$$

on  $\mathbb{E}[\varepsilon|X] = 0$  i  $f$  és una funció lineal que fa una correspondència de  $T(\mathbb{R}^d)$  a  $T(\mathbb{R}^e)$ .

Sota el model de la signatura esperada, donat un nombre de mostres  $\{X_i, Y_i\}_{i=1}^N$ , l'estimació de la signatura truncada esperada de  $Y$  de l'ordre  $m$  condicionada per  $S(X)$ , és a dir  $\rho_m(\mathbb{E}[Y|X])$ , o en altres paraules, la funció lineal  $\rho_m \circ f$ , resulta ser el problema estàndard de regressió lineal [9]; les integrals iterades coordenades de  $S(Y)$  són regressants multidimensionals mentre que les integrals iterades coordenades de  $S(X)$  són variables explicatives. Com hem discutit anteriorment, a la pràctica, hem de considerar la signatura truncada de  $X$  d'un ordre determinat en lloc de la signatura completa, ja que el nombre de variables explicatives ha de ser finit. Com que la regressió és lineal, hi ha molts mètodes estàndard de regressió lineal que es poden emprar. Podem utilitzar, a més, tècniques de regularització o selecció de variables com ara LASSO o SVD per evitar la colinearitat de la matriu de disseny i el problema de sobreajustament. Anomenem aquest mètode de calibració l'aproximació de signatura esperada.

Per mesurar la bondat de l'ajustament del model, utilitzem l'error quadrat mitjà dels residus  $\{a_i\}_{i=1}^N$ , on

$$a_i = S(Y_i) - \hat{f}(S(X_i)), \quad \forall i = 1, \dots, N.$$

Alternativament, podem utilitzar  $R^2$  o  $R^2$  ajustat com a indicador de la bondat de l'ajust.

## 5.4 Implementació del model a un cas pràctic

En resum, tenim el següent model:

1. Donat un conjunt d'entrenament  $\{(P_i, Y_i)\}_{i=0}^N$  de parells d'entrada-sortida, on  $P_i = \{(t_{ij}, p_{ij})\}_j$  és un flux de dades, construïm un nou flux  $\{(X_i, Y_i)\}_{i=0}^N$  incrustant el flux  $P_i$  en un camí continu com hem explicat anteriorment.
2. Utilitzant el nou conjunt d'entrenament  $\{(X_i, Y_i)\}_{i=0}^N$  calculem les corresponents signatures truncades d'ordre  $n \in \mathbb{N}$ ,  $\{(S^n(X_i), Y_i)\}_{i=0}^N$ .
3. Aplicar una regressió lineal a la signatura truncada d'un cert ordre utilitzant un algorisme adequat.

Podem trobar als annexos una implementació del model en *Python*.

En aquesta secció, apliquem el model anteriorment descrit per a predir, a quin país pertany l'acció una empresa utilitzant l'evolució del preu de tancament diari de les seves accions i el volum negociat per a un període de temps donat, la idea i la base del codi l'hem pres i modificat de [16].

Òbviament, la manera com el preu d'una acció canvia amb el temps varia d'empresa a empresa. Si una empresa és rendible i els inversors creuen que seguirà sent rendible en el futur, és probable que el preu de les seves accions augmenti. Per una altra banda, si una empresa està a prop de la fallida, és bastant probable que el preu baixi.

No obstant això, de manera intuïtiva té sentit esperar que els factors externs que afecten al país en conjunt també tinguin un impacte en el preu d'una acció. Per tant, és raonable pensar que els preus de les accions que es negocien en el mateix país tindran certes similituds intrínseques que van més enllà del rendiment de cada empresa en particular.

Com que només volem mostrar la signatura en aquest context, limitarem el problema a tres països: els Estats Units, el Regne Unit i el Japó.

El que farem doncs, carregar la llista de tuples  $(P_i, V_i)_{i=0}^N$  per a un període de  $N$  dies, on  $P_i$  i  $V_i$  són el preu de tancament i volum negociat de les seves accions el  $i$ -èssim dia, junt amb la categoria (països) a la que pertany l'empresa.

Per a això amb l'ajuda del portal web [finance.yahoo.com](http://finance.yahoo.com) hem exportat fulls de càlcul amb les primeres 100 empreses de cada un dels tres països ordenades per capitalització de mercat de major a menor (per tal de seleccionar aquelles accions més representatives de cada país), hem eliminat tota la informació que no siguin els *tickers* i afegit el país per a ser llegit després pel programa, després usant el paquet de Python *yfinance* podem obtenir les dades històriques d'una acció a través del seu *ticker*. Assignem també les tres categories (països) a punts linealment independents  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ .

k	Precisió
1	0.71
2	0.77
3	0.78
4	0.79
5	0.84

Posteriorment, calcularem la signatura del camí (signatura truncada d'un cert nivell) incrustat de les evolucions dels preus de tancament i volum negociats del període inserit (Les dates usades són des de la data 01/12/2023 fins al 01/01/2024), per a fer-ho usem el paquet de Python *iisignature* que computa la signatura. Això ho farem per a diferents ordres de la signatura truncada.

Un cop tenim això barrejem les dades inserides i seleccionem el primer 70% de les dades per a entrenar el model i l'altre 30% com a test. Per a entrenar el model usarem regressió lineal amb LASSO, per a això utilitzem el paquet de Python *scikit learn* que és un paquet que conté diverses funcions per a l'aprenentatge automàtic. Cal incidir en que no pretenem aquí comparar aquest mètode amb un altre, sinó mostrar la potència que la signatura pot tenir com a característica en problemes d'aprenentatge automàtic.

Després d'una execució obtenim que la precisió (és a dir, la tasa d'encerts: (prediccions encertades)÷(prediccions totals)) en funció de l'ordre la signatura truncada ( $k$ ) que podem observar a la taula superior.

Tot i que els resultats del model estan lluny de ser perfectes, ens son prou per a poder observar que la signatura és de fet una característica acceptable capaç de predir la majoria de classes correctament. Observem, a més que la precisió augmenta a mesura que el nivell  $k$  de la signatura truncada ho fa.

Hem provat a aplicar aquest model per a predir el sector al que pertany una empresa d'un mateix país (els Estats Units) basant-nos també en els preus de tancament diaris de les seves accions i el volum negociat. En aquest cas hem pres les primeres 50 empreses amb més capitalització de mercat del sectors de les finances, la salut i la tecnologia de manera similar al exemple anterior. El model en aquest cas no ha donat tan bons resultats, és probable que el problema no estigui ben condicionat o que les accions dels diferents sectors estiguin massa correlacionades entre si, que el preu de tancament i volum negociat diaris no siguin per si sols una característica prou rellevant, o bé que es degui a que la mostra és més petita, cal afegir que en aquest cas la hem pres més petita ja que donant més de 50 empreses per sector apareixien valors "NaN" per al volum negociat o preu de tancament. En la taula d'abaix podem veure els resultats obtinguts per a  $k \in \{1, \dots, 5\}$  el nivell de la signatura truncada.

A l'annex C està el programa en Python que aplica aquest model i a l'annex D els conjunts de dades que hem utilitzat per a provar-lo.

k	Precisió
1	0.63
2	0.65
3	0.68
4	0.7
5	0.72

## 6 Conclusions

Aquest treball hem donat una idea global de que és i quines són algunes de les propietats més interessants de la signatura d'un camí. Hem explicat el mètode de la signatura per a un flux de dades, i la utilitat d'aquest en dades de sèrie temporal i per tant també en dades de tipus financer i hem presentat un exemple simple de com usar aquest mètode per a construir un model d'aprenentatge automàtic o *machine learning* que pugui classificar amb una certa precisió una dada.

Per a aquesta comesa hem hagut d'anar aprenent diferents coneixements durant el desenvolupant del treball. S'ha realitzat una tasca de recerca bibliogràfica important, tant en format digital com en format paper.

Ha quedat explicada la capacitat de la signatura de capturar les propietats geomètriques llevat de reparametritzacions d'un camí, així com que determinades transformacions adequades poden ser útils per a calcular i capturar els moments estadístics d'un flux de dades.

Tot i que els resultats del model implementats a l'últim apartat estan lluny de ser perfectes, ens son prou per a poder observar que la signatura és, de fet, una característica acceptable capaç de predir la majoria de classes correctament. Els resultats són pitjors de l'esperat, però raonables.

No obstant, això no significa que la signatura sigui una mala característica, revisant la literatura al respecte, part d'ella en la bibliografia citada a aquest treball, hi han numerosos exemples en que s'ha aconseguit amb prou éxit utilitzar la signatura com a característica en problemes d'aprenentatge automàtic aplicats a dades de tipus financers, essent capaços amb aquest mètode de classificar correctament més del 90% de classes (p. ex. [9]). És clar doncs, més enllà d'aquest experiment anecdòtic que la signatura és una eina molt potent per a classificar fluxos de dades en problemes d'aprenentatge automàtic.



## Referències

- [1] Ilya Chevyrev; Andre y Komilitzin: A Primer on the Signature Method, [arXiv:1603.03788](#), març de 2016.
- [2] Terry Lions; Zhongmin Qian: System Control and Rough Paths, 2002.
- [3] Bourbaki Nicolas (1987): Topological Vector Spaces: Chapters 1–5. *Éléments de mathématique*. Translated by Eggleston, H.G.; Madan, S. Berlin New York: Springer-Verlag.
- [4] Terry Lions; Andrew D. McLeod: Signature Methods in Machine Learning, [arXiv:2206.14674v1](#), juny de 2022.
- [5] Kuo-Tsai Chen: Iterated Integrals and Exponential Homomorphisms, *Prod. London Math. Soc.* 4 (1954).
- [6] Kuo-Tsai Chen: Integrations of paths, geometric invariants and a generalized Baker-Hausdorff formula. *Ann. of Math. (2)* 65:163-178, 1957
- [7] Paul Wilmott, *Machine Learning: An Applied Mathematics Introduction*, Panda Ohana Publishing (2019).
- [8] Ben Hambly; Terry Lyons: Uniqueness for the signature of a path of bounded variation and the reduced path group, *Ann. of Math., Volume 171*, pp.109-167, 2010.
- [9] Jonathann Field; Lajos Gergely Gyurkó; Mark Kontkowski; Terry Lyons: Extracting information from the signature of a financial data stream. [arXiv:1307.7244](#), July 2014. Preprint.
- [10] Benjamin Graham. Sparse arrays of signatures for online character recognition. [arXiv:1308.0371](#), December 2013. Preprint.
- [11] Daniel Levin, Terry Lyons, and Hao Ni. Learning from the past, predicting the statistics for the future, learning an evolving system. [arXiv:1309.0260](#), September 2015. Preprint.
- [12] Protter, Philip E. (2004), *Stochastic Integration and Differential Equations* (2nd ed.), Springer
- [13] [2] Király, F. J. and Oberhauser, H. (2016) "Kernels for sequentially ordered data, [arXiv: 1601.08169](#)"
- [14] Terry Lyons, Michael Caruana, and Thierry Lévy, *Differential Equations Driven by Rough Paths*, Ecole d'Été de Probabilités de Saint-Flour XXXIV - 2004, Lecture Notes in Mathematics, Springer, 2007.
- [15] Horatio Boedihardjo, Xi Geng, Terry Lyons, and Danyu Yang. The signature of a rough path: Uniqueness. [arXiv:1406.7871](#), August 2014. Preprint.

- [16] Arribas Pérez, Imanol, Rough Path Theory and Signatures Applied To Quantitative Finance, <https://www.quantstart.com/articles/rough-path-theory-and-signatures-applied-to-quantitative-finance-part-1/>
- [17] Chevyrev, I; Oberhauser, H 2022, 'Signature moments to characterize laws of stochastic processes', Journal of Machine Learning Research. arXiv:1810.10971

## A Codi que computa la transformació Lead-Lag

```
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import matplotlib.patches as patches

X=[[1,1], [4,1], [4,4], [2,4], [2,2], [5,2], [5,5], [3,5], [3,3]]
def leadlag(X):
    '''
    Torna la transformacio lead-lag de X

    Arguments:
        X: llista
    '''

    l=[]

    for j in range(2*(len(X))-1):
        i1=j//2
        i2=j//2
        if j%2!=0:
            i1+=1
        l.append((X[i1][1], X[i2][1]))

    return l

def plotLeadLag(X, diagonal=True):
    '''
    Grafica el cami lead-lag

    Arguments:
        X: llista de tuples (X^lead, X^lag)
        diagonal es una variable booleana inicialitzada com a certa.
        Si es certa, es grafica una linia que junta els punts d'inici i c
    '''

    for i in range(len(X)-1):
        plt.plot([X[i][1], X[i+1][1]], [X[i][0], X[i+1][0]],
                 color='k', linestyle='-', linewidth=2)

    # Grafica la diagonal
    if diagonal:
        plt.plot([min(min([p[0] for p in X]), min([p[1]
                for p in X])), max(max([p[0] for p in X]),
                max([p[1] for p in X]))], [min(min([p[0]

```

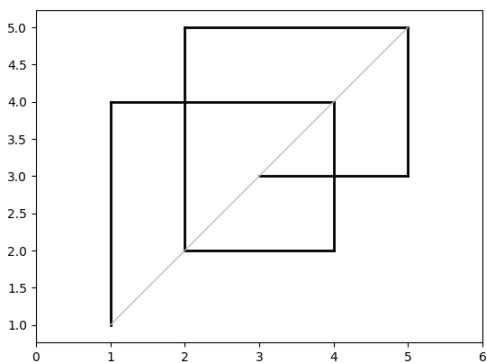
```

        for p in X]), min([p[1] for p in X])),
        max(max([p[0] for p in X]), max([p[1] for
p in X]))], color='#BDBDBD', linestyle='-',
        linewidth=1)

axes=plt.gca()
axes.set_xlim([min([p[1] for p in X])-1, max([p[1] for
p in X])+1])
axes.set_ylim([min([p[0] for p in X])-1, max([p[0] for
p in X])+1])
axes.get_yaxis().get_major_formatter().set_useOffset(False)
axes.get_xaxis().get_major_formatter().set_useOffset(False)
axes.set_aspect('equal', 'datalim')
plt.show()

print(leadlag(X))
plotLeadLag(X,)

```



## B Codi que computa la transformació time-joined

```

import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import matplotlib.patches as patches

```

```

X=[[1,1], [4,2], [6,3], [5,4], [2,5], [5,6], [10,7], [3,8], [3,9]]
def timejoined(X):
    , ,

```

```

        Torna la transformació time-joined de X (punts)

```

```

    , ,

```

```

    X.append(X[-1])

```

```

    l=[]

```

```

for j in range(2*(len(X))+1+2):
    if j==0:
        l.append((X[j][0], 0))
        continue
    for i in range(len(X)-1):
        if j==2*i+1:
            l.append((X[i][0], X[i][1]))
            break
        if j==2*i+2:
            l.append((X[i+1][0], X[i][1]))
            break
return l

def plottimejoined(X):
    '''
    Grafica la transformacio time-joined de X

    Arguments:
        X: llista de tuples (t,X)
    '''

    for i in range(len(X)-1):
        plt.plot([X[i][0], X[i+1][0]], [X[i][1], X[i+1][1]],
                color='k', linestyle='-', linewidth=2)

    axes=plt.gca()
    axes.set_xlim([min([p[0] for p in X]), max([p[0] for p
        in X])+1])
    axes.set_ylim([min([p[1] for p in X]), max([p[1] for p
        in X])+1])
    axes.get_yaxis().get_major_formatter().set_useOffset(False)
    axes.get_xaxis().get_major_formatter().set_useOffset(False)
    axes.set_aspect('equal', 'datalim')
    plt.show()

print(timejoined(X))
plottimejoined(X,)

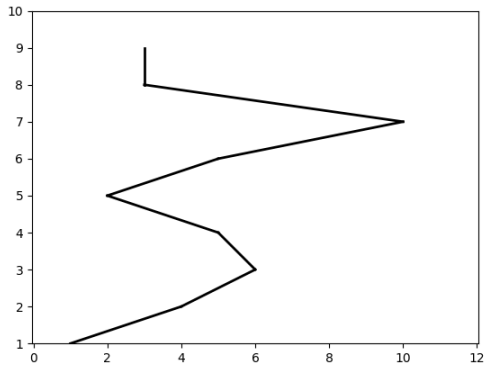
```

## C Codi que implementa el model de la secció 5

```

\begin{lstlisting}
import pandas as pd
import yfinance as yf
import numpy as np

```



```

from sklearn import linear_model
import numbers
import random
from iisignature import sig
from datetime import datetime
from sklearn import preprocessing

```

```

class SigLearn:

```

```

    def __init__(self, order, alpha=0.1):
        if not isinstance(order, numbers.Integral) or order < 1:
            raise NameError('L'ordre ha de ser un enter positiu')
        if not isinstance(alpha, numbers.Real) or alpha <= 0.0:
            raise NameError('Alpha ha de ser un valor real positiu')

```

```

        self.order=int(order)
        self.reg=None
        self.alpha=alpha

```

```

    def train(self, x, y):
        '''

```

```

        Entrena el model utilitzant signatures.

```

```

        x: llista de inputs, on cada element de la llista
        es una llista de tuples.

```

```

        y: llista of outputs.
        '''

```

```

        # Comprovem que x i y son del tipus apropiat
        if x is None or y is None:
            return
        if not (type(x) is list or type(x) is tuple)

```

```

    or not (type(y) is list or type(y) is tuple):
        raise NameError('Les entrades i les sortides
            han de ser tuples')
    if len(x)!=len(y):
        raise NameError('El nombre de entrades i de
            sortides ha de coincidir.')
###

X=[list(sig(np.array(stream), self.order,0)) for stream in x]

self.reg = linear_model.Lasso(alpha = self.alpha,
max_iter=1000000, tol=0.5)
self.reg.fit(X, y)

def predict(self, x):
    """
    Prediu els outputs dels inputs x usant el model pre-entrenat.

    x: llist d'inputs, de tuples.

    Retorna:

    llista d'outputs predits.
    """
    if self.reg is None:
        raise NameError('El model no esta entrenat')

    X=[list(sig(np.array(stream), self.order,0)) for stream in x]

    return self.reg.predict(X)

# Dates de començament i final.
start="2023-12-01"
end="2024-01-01"

#Llegim un full de calcul amb els tickers
#de cada accio i el pais al que pertany.
df = pd.read_excel("C:\Pyth\fin\USJPNUK.xlsx", header=None)

tickers = df.values
class Stock:
    """
    Classe amb la informacio d'una accio
    """

    def __init__(self, data, sector):

```

```

# Caracteristica de la classe que guarda el flux de dades
self.data=np.array(data ,dtype='float32 ')

# Caracteristica de la classe que guarda el pais de l'accio.
self.sector=sector

#El output per a entrenar ha de ser un vector
#assignem a cada pais un punt
#amb la següent funcio

self.point=sector_to_point (sector)

def accuracy(predictions , y):
    p=0

    for i in range(len(y)):
        if round(predictions [ i ][0])==y [ i ][0]
        and round(predictions [ i ][1])==y [ i ][1]
        and round(predictions [ i ][2])==y [ i ][2]:
            #print (predictions [ i ][0] , predictions [ i ][1])
            p+=1

    return p/float (len(y))
def sector_to_point (sector):
    '''
    Converteix un pais en un punt
    '''
    dictionary={"Japan": (1,0,0), "US": (0, 1,0), "UK":(0,0,1)}
    return dictionary [sector]
def getData(ticker , start , end):
    '''
    Pren les dades del ticker llegit , per a un periode
    de temps determinat
    '''
    dat=[]
    p=yf.Ticker (ticker)
    stock=p.history (str=start ,nd=end) [[ 'Close' , 'Volume' ]]
    values=stock [[ "Close" , "Volume" ]].reset_index ().values
    for i in range (len (values)):
        values [ i ][0]=values [ i ][0].strftime ("%Y%m%d%H%M%S")
        values [ i ][0]=i

    return values

```



```

data=[]

for country in tickers:
    print(" Carregant "+country+"...")
    companyData=getData(country[1], start, end)

    # Si la companyia no te dades, la saltem.
    if len(companyData)==0: continue

    data.append(Stock(companyData, country[0]))
    print(" Done.")

# Divim les dades en dos subconjunts:
#el conjut d'entenament training_set, amb el 70% de les dades,
#i el conjunt de test testing_set, amb les dades restants.
random.shuffle(data)

training_set=data[0:int(0.7*len(data))]
testing_set=[company for company in data
              if company not in training_set]
# Construim els inputs i els outputs per a entrenar el model.
inputs=[company.data for company in training_set]
outputs=[company.point for company in training_set]

# Construim els inputs i els outputs per a testejar el model.
inputsTEST=[company.data for company in testing_set]
outputsTEST=[company.point for company in testing_set]
#Apliqueu el model per a signatures de ordre 1 a 8.
for signature_order in range(1, 9):
    # S'entrena el model
    model=SigLearn(order=signature_order)
    model.train(inputs, outputs)

    # Calculem les prediccions.
    predictions=model.predict(inputsTEST)

    # Mirem l'exactitut de les prediccions i l'imprimim.
    print(accuracy(predictions, outputsTEST))
    j=0
    e=0
    r=0
    #j1=0
    #e1=0
    #r1=0
    for i in range(len(predictions)):

```

```

    if round(predictions [ i ][0])==1
    and round(predictions [ i ][1])==0:
        j+=1
    if round(predictions [ i ][0])==0
    and round(predictions [ i ][1])==1:
        e+=1
    if round(predictions [ i ][0])==0
    and round(predictions [ i ][2])==1:
        r+=1
    #if round(outputsTest [ i ][0])==1:#
    and round(predictions [ i ][1])==0:
        #j+=1
    #if round(outputsTest [ i ][0])==-1:
    #and round(predictions [ i ][1])==0:
        #e+=1
    #if round(outputsTest [ i ][0])==0:#
    and round(predictions [ i ][1])==1:
        #r+=1
print(" predits japo:")
print(j)
#print(" reals japo:" j1)
print(" predits EEUU:" "+e+")
print(e)
#print(" reals EEUU:" e1)
print(" predits RU:" "+r+")
print(r)
#print(" reals RU:" r1)

```

## D Conjunts de dades usats per a provar el model

UK	CCL.L
UK	ANTO.L
UK	ANTOL.XC
UK	ABECD.XC
UK	STANL.XC
UK	ABEAD.XC
UK	OQFP.IL
UK	IMB.L
UK	IMBL.XC
UK	WPM.L
UK	ABF.L
UK	ABFL.XC
UK	ORIH.IL
UK	OHD6.IL
UK	VOD.L
UK	VODL.XC
UK	NWGL.XC
UK	ORYA.IL
UK	SSE.L
UK	SSEL.XC
UK	NWG.L
UK	STAN.L
UK	TSCO.L
UK	TSCOL.XC
UK	AHTL.XC
UK	AHT.L
UK	BARC.L
UK	BARCL.XC
UK	PRU.L
UK	PRUL.XC
UK	III.L
UK	ORUK.IL
UK	FLTRL.XC
UK	FLTR.L
UK	OR22.IL
UK	AALL.XC
UK	AAL.L
UK	RR.L
UK	RRL.XC
UK	APCD.XC
UK	MSFD.XC
UK	EXPNL.XC
UK	EXPN.L
UK	LLOYL.XC
UK	OR2V.IL
UK	OQYP.IL
UK	OZ4S.IL
UK	FERG.L
UK	FERGL.XC
UK	OQIU.IL

UK	HLN.L
UK	HLNL.XC
UK	LLOY.L
UK	WDS.L
UK	WDSL.XC
UK	NOVOBC.XC
UK	BAL.XC
UK	BA.L
UK	CRHL.XC
UK	CRH.L
UK	OM69.IL
UK	CPGL.XC
UK	CPG.L
UK	GACB.L
UK	NGL.XC
UK	NG.L
UK	RKTL.XC
UK	RKT.L
UK	SANB.L
UK	GACA.L
UK	LSEG.L
UK	BNC.L
UK	BATSL.XC
UK	BATS.L
UK	GLEN.L
UK	GLENL.XC
UK	RELL.XC
UK	REL.L
UK	DGEL.XC
UK	DGE.L
UK	GSKL.XC
UK	GSK.L
UK	BPL.XC
UK	BP.L
UK	RIO.L
UK	RIOL.XC
UK	BC94.L
UK	ULVR.L
UK	ULVRL.XC
UK	BP-B.L
UK	BP-A.L
UK	HSBAL.XC
UK	HSBA.L
UK	BHP.L
UK	BHPL.XC
UK	SHELL.XC
UK	SHELL
UK	AZN.L
UK	AZNL.XC
UK	TYT.L

US	DIS
US	HDB
US	DHR
US	C-PJ
US	WFC
US	INTU
US	CMCSA
US	ACGBF
US	ACGBY
US	PCCYF
US	BABA
US	BABAF
US	WFC-PL
US	SAPGF
US	SAP
US	WFC-PR
US	TMUS
US	MTSUY
US	WFC-PY
US	PDD
US	ABT
US	LIN
US	INTC
US	CSCO
US	RYDAF
US	SHEL
US	TMO
US	HESAY
US	HESAF
US	MCD
US	AZN
US	NFLX
US	AZNCF
US	NVSEF
US	NVS
US	IDCBY
US	IDCBF
US	ACN
US	PEP
US	FMX
US	AMD
US	RHHVF
US	RHHBY
US	RHHBF
US	BAC-PB
US	BML-PJ
US	LRLCF
US	LRLCY
US	KO
US	CRM

US	TOYOF
US	TM
US	ADBE
US	CVX
US	ASML
US	ASMLF
US	ABBV
US	BML-PL
US	BAC-PL
US	BML-PH
US	ORCL
US	BAC-PE
US	MRK
US	COST
US	BML-PG
US	NSRGY
US	NSRGF
US	BAC-PK
US	BAC
US	JPM-PD
US	JPM-PC
US	LTMAY
US	TCEHY
US	HD
US	PG
US	TCTZF
US	LVMUY
US	LVMHF
US	JNJ
US	XOM
US	MA
US	WMT
US	NONOF
US	NVO
US	UNH
US	AVGO
US	TSM
US	V
US	JPM
US	LLY
US	TSLA
US	BRK-A
US	BRK-B
US	META
US	NVDA
US	AMZN
US	GOOGL
US	GOOG
US	AAPL
US	MSFT

Japan	2801.T
Japan	9104.T
Japan	5020.T
Japan	7832.T
Japan	7309.T
Japan	7532.T
Japan	4507.T
Japan	8309.T
Japan	7270.T
Japan	8830.T
Japan	4523.T
Japan	9843.T
Japan	1928.T
Japan	8604.T
Japan	4684.T
Japan	7201.T
Japan	3659.T
Japan	9101.T
Japan	9735.T
Japan	6701.T
Japan	8630.T
Japan	1605.T
Japan	6326.T
Japan	4307.T
Japan	7733.T
Japan	4612.T
Japan	6762.T
Japan	8802.T
Japan	4452.T
Japan	8267.T
Japan	2502.T
Japan	1925.T
Japan	2802.T
Japan	8750.T
Japan	9613.T
Japan	7269.T
Japan	8725.T
Japan	6971.T
Japan	5401.T
Japan	4503.T
Japan	8113.T
Japan	6920.T
Japan	4578.T
Japan	7011.T
Japan	9020.T
Japan	8591.T
Japan	6752.T
Japan	8015.T
Japan	6594.T
Japan	8801.T

Japan	4689.T
Japan	9022.T
Japan	6301.T
Japan	4901.T
Japan	4543.T
Japan	7751.T
Japan	6857.T
Japan	6201.T
Japan	6702.T
Japan	8053.T
Japan	8002.T
Japan	6146.T
Japan	5108.T
Japan	6954.T
Japan	6723.T
Japan	6178.T
Japan	6503.T
Japan	3382.T
Japan	7182.T
Japan	6273.T
Japan	6981.T
Japan	7741.T
Japan	8411.T
Japan	6902.T
Japan	2914.T
Japan	4502.T
Japan	6367.T
Japan	8766.T
Japan	7267.T
Japan	4568.T
Japan	8031.T
Japan	9434.T
Japan	4661.T
Japan	9984.T
Japan	7974.T
Japan	8001.T
Japan	4519.T
Japan	8316.T
Japan	6098.T
Japan	8058.T
Japan	9433.T
Japan	6501.T
Japan	4063.T
Japan	9983.T
Japan	8035.T
Japan	8306.T
Japan	9432.T
Japan	6861.T
Japan	6758.T
Japan	7203.T

Health	NTOIF
Health	DVN
Health	FANG
Health	CVE
Health	HAL
Health	IMO
Health	BKR
Health	MPLX
Health	LNG
Health	WDS
Health	KMI
Health	WOPEF
Health	TRP
Health	OKE
Health	WMB
Health	SU
Health	HES
Health	VLO
Health	TNCAF
Health	ET
Health	OXY
Health	PXD
Health	E
Health	EIPAF
Health	PEXNY
Health	PSX
Health	EPD
Health	MPC
Health	EOG
Health	CNQ
Health	SLB
Health	EBBNF
Health	ENB
Health	SNPMF
Health	STOHF
Health	EQNR
Health	CUAEF
Health	CSUAY
Health	BPAQF
Health	BP
Health	PBR-A
Health	PBR
Health	COP
Health	TTE
Health	TTFNF
Health	PCCYF
Health	RYDAF
Health	SHEL
Health	CVX
Health	XOM

Fin	PNC
Fin	BN
Fin	LDNXF
Fin	DBSDF
Fin	DBSDY
Fin	UNCRY
Fin	LNSTY
Fin	UNCFE
Fin	SMFG
Fin	USB
Fin	PYPL
Fin	SAN
Fin	BCDRF
Fin	ITUB
Fin	SMFNF
Fin	SBKFF
Fin	IVSXF
Fin	IVSBF
Fin	MCO
Fin	CME
Fin	ICE
Fin	KKR
Fin	AXAHY
Fin	AXAHF
Fin	USB-PP
Fin	ZURVY
Fin	ZFSVF
Fin	MS-PA
Fin	PBCRF
Fin	GS-PD
Fin	PBCRY
Fin	MS-PE
Fin	MS-PF
Fin	MS-PI
Fin	MS-PK
Fin	GS-PA
Fin	CILJF
Fin	GS-PK
Fin	IBN
Fin	BNPQY
Fin	BNPQF
Fin	USB-PH
Fin	PIAIF
Fin	PNGAY
Fin	AAIGF
Fin	AAGIY
Fin	CB
Fin	UBS
Fin	MMC
Fin	CIHKY

Fin	PGR
Fin	ALIZY
Fin	MBFJF
Fin	MUFG
Fin	ALIZF
Fin	TD
Fin	BLK
Fin	SCHW
Fin	WFC-PC
Fin	C
Fin	CBAUF
Fin	AXP
Fin	BACHF
Fin	CMWAY
Fin	SPGI
Fin	RY
Fin	BACHY
Fin	GS
Fin	BX
Fin	CICHY
Fin	CICHF
Fin	HBCYF
Fin	HSBC
Fin	MS
Fin	HDB
Fin	C-PJ
Fin	WFC
Fin	ACGBF
Fin	ACGBY
Fin	WFC-PL
Fin	WFC-PR
Fin	WFC-PY
Fin	IDCBY
Fin	IDCBF
Fin	BAC-PB
Fin	BML-PJ
Fin	BML-PL
Fin	BAC-PL
Fin	BML-PH
Fin	BAC-PE
Fin	BML-PG
Fin	BAC-PK
Fin	BAC
Fin	JPM-PD
Fin	JPM-PC
Fin	MA
Fin	V
Fin	JPM
Fin	BRK-A
Fin	BRK-B

Tech	NTOIF
Tech	DVN
Tech	FANG
Tech	CVE
Tech	HAL
Tech	IMO
Tech	BKR
Tech	MPLX
Tech	LNG
Tech	WDS
Tech	KMI
Tech	WOPEF
Tech	TRP
Tech	OKE
Tech	WMB
Tech	SU
Tech	HES
Tech	VLO
Tech	TNCAF
Tech	ET
Tech	OXY
Tech	PXD
Tech	E
Tech	EIPAF
Tech	PEXNY
Tech	PSX
Tech	EPD
Tech	MPC
Tech	EOG
Tech	CNQ
Tech	SLB
Tech	EBBNF
Tech	ENB
Tech	SNPMF
Tech	STOHF
Tech	EQNR
Tech	CUAEF
Tech	CSUAY
Tech	BPAQF
Tech	BP
Tech	PBR-A
Tech	PBR
Tech	COP
Tech	TTE
Tech	TTFNF
Tech	PCCYF
Tech	RYDAF
Tech	SHEL
Tech	CVX
Tech	XOM