The lead-lag transformation

Remy Messadene

DataSig

A rough path between
mathematics and data science

University of Oxford
The Alan Turing Institute
Imperial College London
UCL
UKRI Engineering and Physical Sciences Research Council

Outline

1. Definitions
2. Levy area
   - Defining the Levy area between paths
   - Levy area as a measure of auto-correlation
   - Levy area and signature
3. Recovering statistical properties of $\hat{\gamma}$ using lead-lag transform and signatures
4. Signatures as statistical properties

Definitions

Let Λ denote the lead-lag transform. We will see three variations of it:

1. **Discrete-discrete lead-lag**
   Λ_{dd} : {Sequences in $\mathbb{R}^d$} $\to$ {Sequences in $\mathbb{R}^2 d$}
2. **Discrete-continuous lead-lag**
   Λ_{dc} : {Sequences in $\mathbb{R}^d$} $\to$ {Continuous paths in $\mathbb{R}^2 d$}
3. **Continuous-continuous lead-lag**
   Λ_{cc} : {Sequences in $\mathbb{R}^d$} $\to$ {Continuous paths in $\mathbb{R}^2 d$}

Definition (p-lead lag transformation (Discrete-discrete version))

Let $\hat{\gamma} := \{x_{t_i}\}_{i\in\{1,\ldots,n\}}$ be a sequence of n points in $\mathbb{R}^d$. Define by the **discrete-discrete p-lead-lag transform** $\Lambda_{dd}(\hat{\gamma}; p)$ of $\hat{\gamma}$ the following sequence of points,

$$\Lambda_d(\hat{\gamma}, p) = \left\{(x_{t_{s+p-1}}, x_{t_s}), (x_{t_{s+p}}, x_{t_s})\right\}_{s\in\{1,\ldots,n-p\}} \tag{1}$$

for $p \in \{1, \ldots, n-1\}$.

Remark (Length and dimension)

$\Lambda_{dd}(\hat{\gamma}; p)$ transform a sequence of n points in $\mathbb{R}^d$ into one of $2n - 2p$ points in $\mathbb{R}^{2d}$.

Examples (Discrete-discrete version)

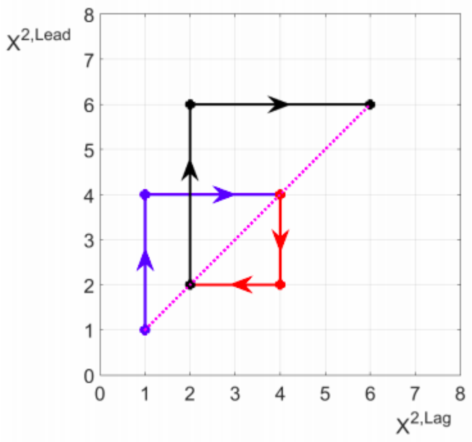
1. Take $\hat{\gamma} = (1, 4, 2, 6)$. Then,

$$\Lambda_{dd}(\hat{\gamma}, 1) = ((1,1),(4,1),(4,4),(2,4),(2,2),(6,2),(6,6)) \tag{2}$$

$$\Lambda_{dd}(\hat{\gamma}, 2) = ((4,1),(2,1),(2,4),(6,4),(6,2)) \tag{3}$$

$$\Lambda_{dd}(\hat{\gamma}, 3) = ((2,1),(6,1),(6,4)) \tag{4}$$

Examples (Discrete-discrete version)



Definition (continuous)

Definition (p-lead lag transformation (Discrete-continuous version))

Let $\hat{\gamma} := \{x_{t_i}\}_{i \in \{1,\dots,n\}}$ be a sequence of n points in $\mathbb{R}^d$. Define by the **discrete-continuous** p**-lead-lag transform** $\Lambda_{dc}(\hat{\gamma}; p)$ of $\hat{\gamma}$ the following continuous path in $\mathbb{R}^{2d}$,

$$\Lambda_{dc}(\hat{\gamma}; p)(t) := \tag{5}$$

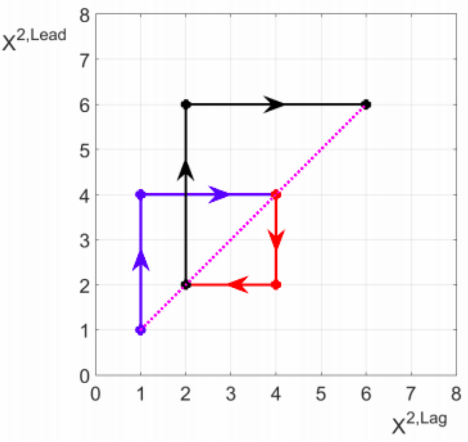
$$\begin{cases} (x_{t_{i+p-1}}, x_{t_i}), & t \in [i, i + \frac{1}{3}) \\ (x_{t_{i+p-1}} + 3(t - (i + \frac{1}{3}))(x_{t_{i+p}} - x_{t_{i+p-1}}), x_{t_i}), & t \in [i + \frac{1}{3}, i + \frac{2}{3}) \\ (x_{t_{i+p}}, x_{t_i} + 3(t - (i + \frac{2}{3}))(x_{t_{i+1}} - x_{t_i})), & t \in [i + \frac{2}{3}, i + 1) \end{cases} \tag{6}$$

for $p \in \{1, \dots, n-1\}$ and $t \in [1; n-p]$.

Remark (Trivial recovery)

The discrete-continuous lead lag path $\Lambda_{dc}(\hat{\gamma}; p)(t)$ evaluated at the times $\{k + \frac{s}{3}\}_{k \in \{1,\dots,n\}, s \in \{1,2\}}$ exactly recovers the discrete-discrete lead lag $\Lambda_{dd}(\hat{\gamma})$.

Examples (Discrete-continuous version)



Definition (continuous-continuous)

Definition (p-lead lag transformation (continuous version))

Let $\gamma : [a, b] \to \mathbb{R}^d$ be a path in $\mathbb{R}^d$. The **continuous-continuous** $p-$**lead-lag path** Λ_{cc} of γ is simply

$$\Lambda_{cc}(\gamma; p)(t) := (\gamma(t + p), \gamma(t)) \tag{7}$$

for $p \in [a; b]$ and $t \in [a, b - p]$.

Remark (Λ_{cc} is the limit of Λ_{dc} for large n)

Let $\hat{\gamma} := \{x_{t_i}\}_{i \in \{1,\dots,n\}}$ be a set of n points and consider its linear interpolation γ_L^n. If there exists a path $\gamma : [a,b] \to \mathbb{R}^d$ such that $\hat{\gamma}$ is the evaluation of γ on the set of times $\{t_i\}_{i \in \{1,\dots,n\}}$, then

$$\lim_{n\to\infty} \Lambda_{dc}(\hat{\gamma}_n; \lfloor \frac{p_n}{n} \rfloor)(\frac{t}{n}) = \Lambda_{cc}\big(\gamma; (b-a)c + a\big)\big((b-a)t + a\big), \ \forall t \in [0;1] \tag{8}$$

where p_n is such that $\frac{p_n}{n} = c$ for a constant $c \in [0,1]$, $n \in \mathbb{N}$.

Outline

1. Definitions
2. Levy area
   - Defining the Levy area between paths
   - Levy area as a measure of auto-correlation
   - Levy area and signature
3. Recovering statistical properties of $\hat{\gamma}$ using lead-lag transform and signatures
4. Signatures as statistical properties

Levy area and convexity

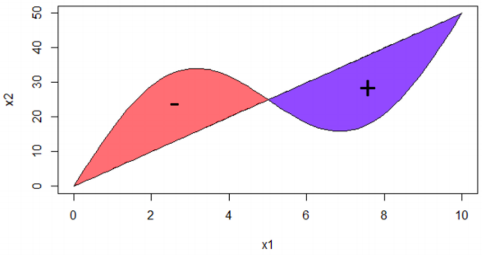
- One can think of the **correlation of γ'' with itself in terms of the Levy area A of its lead-lag path** with respect to the straight line linking endpoints.

Levy area

Definition

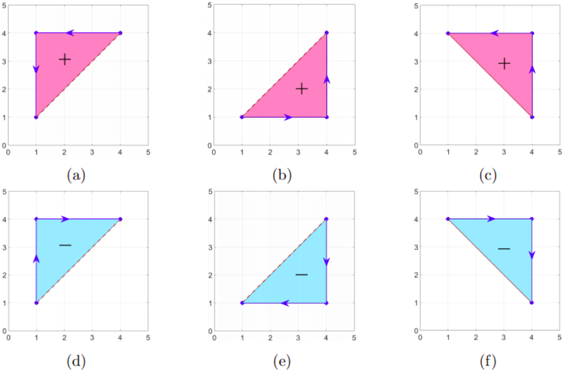
Denote by $A_{[0,t]}$ the Levy area drawn by (X_1, X_2) between $[0, t]$ wrt. straight dashed line linking start and end points,

$$A_{[0,t]} := \frac{1}{2}\left(\int_0^t X_1(s)dX_2(s) - \int_0^t X_2(s)dX_1(s)\right) \tag{9}$$



Signs of the Levy area in terms of the windig number

Using **Green's theorem** (to express A as a double integral with respect to the area) **and then Kelvin-Stokes theorem** (to express the integral of the signed area w.r.t. to windig number of the curve around the areas) **gives the following sign rules,**

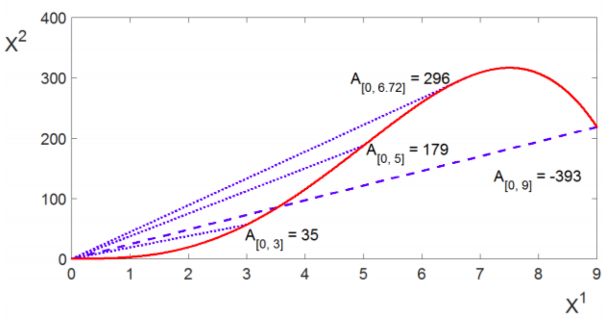


(a) (b) (c)
(d) (e) (f)

Levy area and auto-correlation

Let $\gamma : [a, b] \to \mathbb{R}$ be a path in $\mathbb{R}$.

Consider the p-lead-lag transform $\Lambda_{cc}(\gamma; p) : [a; b - p] \to \mathbb{R}^{2d}$ of γ whose components are denoted as $\Lambda_{cc}(\gamma; p) = (X_1, X_2)$.



(Plot from[1])

[1] Ilya Chevyrev and Andrey Kormilitzin. "A primer on the signature method in machine learning". In: *arXiv preprint arXiv:1603.03788* (2016).

Levy area and auto-correlations between second derivatives:

$$\frac{d}{dt}A_{[0,t]} \stackrel{\geq}{\leq} 0 \quad \Longrightarrow \quad sign(\gamma''(t)\gamma''(t-p)) = \pm 1 \tag{10}$$

if both second-derivatives are non-null and well-defined.

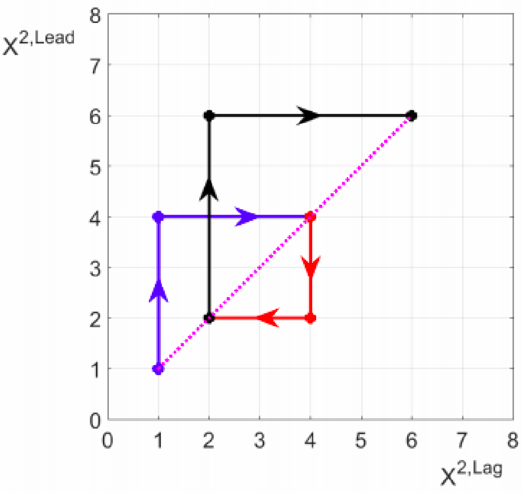
Finally, Levy area and signature are related.
By definition 7, the Levy area A^{ij} between the components X_i and X_j of a path γ can be expressed in terms of the 2-order signature terms as

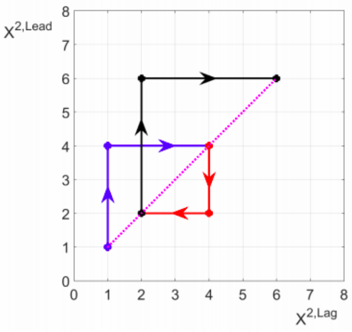
$$A^{ij}_{[0,t]} = \frac{1}{2}\left(S(\gamma)^{2;ij} - S(\gamma)^{2;ji}\right) \tag{11}$$

Outline

1. Definitions
2. Levy area
   - Defining the Levy area between paths
   - Levy area as a measure of auto-correlation
   - Levy area and signature
3. Recovering statistical properties of $\hat{\gamma}$ using lead-lag transform and signatures
4. Signatures as statistical properties

Recall the following lead-lag path





DataSig

The total Levy area $A_{[0,T]}$ of the rectangles is

$$|A_{[0,T]}| = \frac{1}{2}[(X_2^2 - X_1^2)(X_2^2 - X_1^2) + (X_3^2 - X_2^2)(X_3^2 - X_2^2) + 2) \tag{12}$$

$$+ (X_4^2 - X_3^2)(X_4^2 - X_3^2)] \tag{13}$$

$$= \frac{1}{2}[(4-1)^2 + (2-4)^2 + (6-2)^2] \tag{14}$$

Definition (Quadratic variation (discrete version))

Let $\hat{\gamma} := \{X_i\}_{i \in \{1,\dots,n\}}$ be a set of n points in $\mathbb{R}^d$. Then, the quadratic variation $QV(\hat{\gamma})$ of $\hat{\gamma}$ is

$$QV(\hat{\gamma}) := \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2 \tag{15}$$

For a lead-lag path γ, the Levy area and its quadratic variation are linked,

$$A_{[0,T]} = \frac{1}{2} QV(X) \tag{16}$$

Cumulative lead-lag transform

We now present two transformations that will be used together to recover statistical properties of $\hat{\gamma}$:

1. the $k-$cumulative sum transformation
2. the zero augmentation.

Lead lag and cumulative sum

Definition (k-cumulative sum transformation)

Let $\hat{\gamma} := \{X_i\}_{i \in \{1,\dots,n\}}$ be a set of n points in $\mathbb{R}^d$. Define the k-cumulative sum transformation C_k as,

$$C_k(\hat{\gamma}) := \{X_1^k, X_1^k + X_2^k, \dots, \sum_{i=1}^{n} X_i^k\} \tag{17}$$

Remark

$C_k(\hat{\gamma})$ preserves the length of the sequence.

Zero augmentation

Definition (Zero augmentation)

Let $\hat{\gamma} := \{X_i\}_{i \in \{1,\dots,n\}}$ be a set of n points in $\mathbb{R}^d$. Define the **zero augmentation** as $A : \{X_1, X_2, \dots\} \mapsto \{0, X_1, X_2, \dots\}$.

Remark

The zero augmentation makes the signature sensitive to translations of the time series (if path is translated, the first order term clearly changes).

Cumulative lead-lag transform

We will show how we recover statistical properties of interest using the k-cumulative lead-lag transform Λ^k

Definition (Cumulative lead-lag transform)

Let $\hat{\gamma} := \{X_i\}_{i \in \{1,\dots,n\}}$ be a set of n points in $\mathbb{R}^d$. Define the k-**cumulative lead lag transformation** as

$$\Lambda^k(\hat{\gamma}) := (C_k \circ A)(\hat{\gamma}) \tag{18}$$

$$= \{0, X_1^k, X_1^k + X_2^k, \dots, \sum_{i=1}^{n-1} X_i^k\} \tag{19}$$

Recovering statistical properties of $\hat{\gamma}$ with cumulative lead-lag

DataSig

Lemma (Cumulative lead lag and statistical properties)

Let $\hat{\gamma} := \{X_i\}_{i\in\{1,\dots,n\}}$ be a set of n points in $\mathbb{R}^d$ and $\Lambda^k(\hat{\gamma}) := \{\tilde{X}_i^j\}_{i\in\{0,\dots,n\},j\in\{1,2\}}$ its 1-cumulative lead lag transformation. Then, ***the total increment $\Delta\Lambda^1(\hat{\gamma})$ and quadratic variation $QV(\Lambda^1(\hat{\gamma}))$ of $\Lambda^1(\hat{\gamma})$ characterise the mean and variance of $\hat{\gamma}$,***

$$\Delta\Lambda^1(\hat{\gamma})_j = \sum_{i=1}^{n} X_i, \quad j \in \{1,2\}, \tag{20}$$

$$QV(\Lambda^1(\hat{\gamma}))_j = \sum_{i=0}^{N-1} (\tilde{X}_{i+1}^j - \tilde{X}_i^j)^2 = \sum_{i=1}^{N} X_i^2, \quad j \in \{1,2\}. \tag{21}$$

Corollary (Cumulative lead-lag to recover mean and variance)

In particular,

$$Mean(X) = \frac{\Delta\Lambda^1(\hat{\gamma}))}{N} \tag{22}$$

$$Var(X) = \frac{1}{N}\left(QV(\Lambda^1(\hat{\gamma}) - \frac{1}{N}(\Delta\Lambda^1(\hat{\gamma})^2\right) \tag{23}$$

Outline

Signature as statistical properties

Let $\hat{\gamma} := \{X_i\}_{i\in\{1,\dots,n\}}$ be a set of n points in $\mathbb{R}^d$ and $\Lambda^1(\hat{\gamma}) := \{\tilde{X}_i^j\}_{i\in\{0,\dots,n\},j\in\{1,2\}}$ its 1-cumulative lead lag transformation. Denote the 2-truncated signature of the latter path as

$$S(\Lambda^1(\hat{\gamma})|_{L=2} = \left(1, S^{1;1}, S^{1;2}, S^{2;11}, S^{2;12}, S^{2;21}, S^{2;22}\right) \tag{24}$$

Lemma (Truncated signature of 1-cumulative lead lag)

The signature of the 1*-cumulative lead lag transformed path satisfies*

$$S^{1;1} = S^{1;2} = \sum_{i=1}^{N} X_i \tag{25}$$

$$S^{2;11} = S^{2;22} = \frac{1}{2}\left(\sum_{i=1}^{n} X_i^2\right) \tag{26}$$

$$S^{2;12} = \frac{1}{2}\left[\left(\sum_{i}^{n} X_i\right)^2 + \sum_{i}^{n} X_i^2\right] \tag{27}$$

$$S^{2;21} = \frac{1}{2}\left[\left(\sum_{i}^{n} X_i\right)^2 - \sum_{i}^{n} X_i^2\right] \tag{28}$$

Similar equations can be derived for the signature of the raw path $\hat{\gamma}$,

Lemma (Truncated signature of γ)

$$S^{1;1} = S^{1;2} = \sum_{i=1}^{N} X_{i+1} - X_i \tag{29}$$

$$S^{2;11} = S^{2;22} = \frac{1}{2}\left(\sum_{i=1}^{n}(X_{i+1} - X_i)\right)^2 \tag{30}$$

$$S^{2;12} = \frac{1}{2}\left[\left(\sum_{i}^{n}(X_{i+1} - X_i)\right)^2 + \sum_{i}^{n}(X_{i+1} - X_i)\right] \tag{31}$$

$$S^{2;21} = \frac{1}{2}\left[\left(\sum_{i}^{n}(X_{i+1} - X_i)\right)^2 + \sum_{i}^{n}(X_{i+1} - X_i)\right] \tag{32}$$

$$\tag{33}$$

Using the above equations, simple algebra yields the following relationships:

Corollary (Signature as statistical properties)

$$Mean(X) = \frac{1}{n} S^{1;1} \tag{34}$$

$$Var(X) = -\frac{n+1}{n^2} S^{2;1,2} + \frac{n-1}{n^2} S^{2;21} \tag{35}$$

Bibliography