Barycentres de séries temporelles : une nouvelle approche basée sur la méthode de la signature

Journées MAS (Modélisation Aléatoire et Statistique) 2022 de la Société de Mathématiques Appliquées et Industrielles

Raphaël Mignot

Institut Elie Cartan de Lorraine
Université de Lorraine
raphael.mignot@univ-lorraine.fr

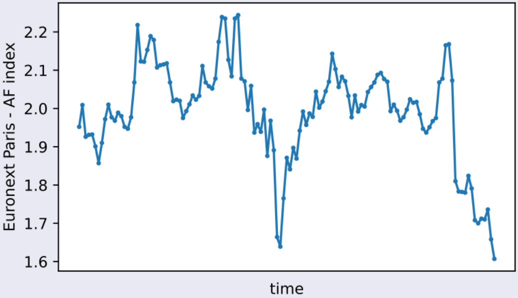
August 29, 2022, Rouen

Joint work with M. Clausel, K. Usevich, G. Oppenheim, L. Coutin, A. Lejay

Background

Background

Time series

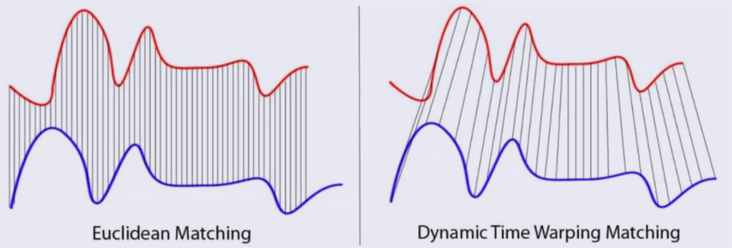


→ sequential data appears in many contexts!

Background

Time series

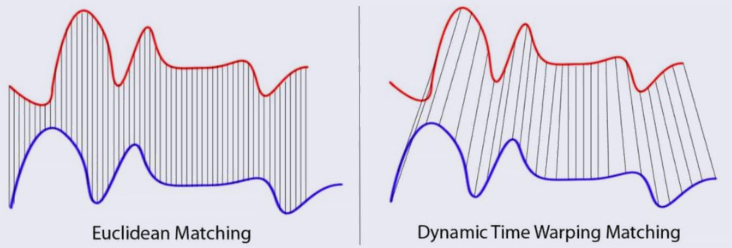
- Similarity measure between two time series ?



Background

Time series

- Similarity measure between two time series ?

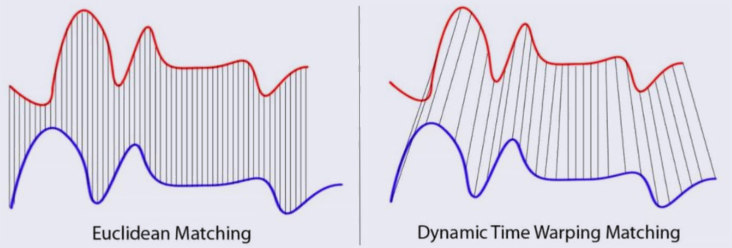


- Euclidean → not accurate for many tasks

Background

Time series

- Similarity measure between two time series ?



- Euclidean → not accurate for many tasks
- Dynamic Time Warping (DTW) → relevant, versatile

Background

Motivation

- Many useful methods in statistical learning rely on averaging

Background

Motivation

- Many useful methods in statistical learning rely on averaging
- **Problem:** how to extend those statistical learning procedure to the time series framework

Background

Motivation

- Many useful methods in statistical learning rely on averaging
- **Problem:** how to extend those statistical learning procedure to the time series framework

$\rightarrow$ How to average a set of time series?

Background

Motivation

- Many useful methods in statistical learning rely on averaging
- **Problem:** how to extend those statistical learning procedure to the time series framework

$\rightarrow$ How to average a set of time series?

What has been done?

- Pairwise averaging (with eg. euclidean distance)
- Averaging method using DTW[1]
- Averaging methods on Lie groups

[1] Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. Pattern recognition, 44(3), 678-693. [PKG11]

The signature transform

The signature transform

Definition

- Input: a continuous path $X : [0,1] \to \mathbb{R}^d$

The signature transform

Definition

- Input: a continuous path $X : [0,1] \to \mathbb{R}^d$
- Signature of order m:

$$S^{(m)}_{[0,1]}(X) \overset{\text{def}}{=} \int \cdots \int_{0<u_1<\cdots<u_m<1} dX_{u_1} \otimes \cdots \otimes dX_{u_m} \in (\mathbb{R}^d)^{\otimes m}$$

with $\otimes$ the tensor product.

The signature transform

Definition

- Input: a continuous path $X : [0,1] \to \mathbb{R}^d$
- Signature of order m:

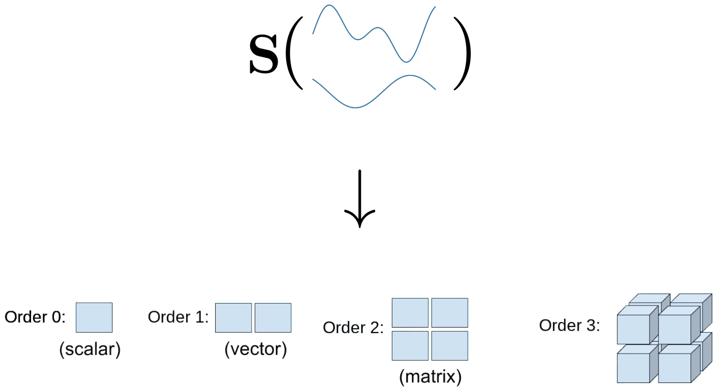
$$S^{(m)}_{[0,1]}(X) \stackrel{\text{def}}{=} \int \cdots \int_{0<u_1<\cdots<u_m<1} dX_{u_1} \otimes \cdots \otimes dX_{u_m} \in (\mathbb{R}^d)^{\otimes m}$$

  with $\otimes$ the tensor product.
- Signature of X is the infinite collection of signatures of all orders:

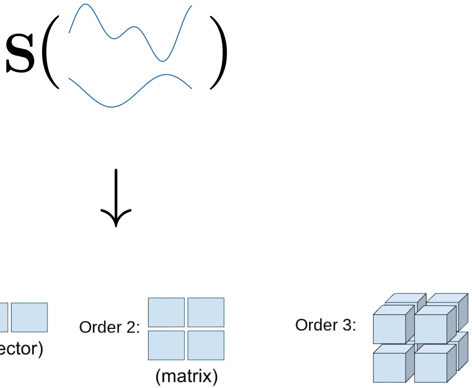
$$\mathbf{S}_{[0,1]}(X) = (1, S^{(1)}_{[0,1]}(X), S^{(2)}_{[0,1]}(X), S^{(3)}_{[0,1]}(X), \dots)$$

Chevyrev, I., & Kormilitzin, A. (2016). A primer on the signature method in machine learning. arXiv preprint arXiv:1603.03788. [CK16]

The signature transform

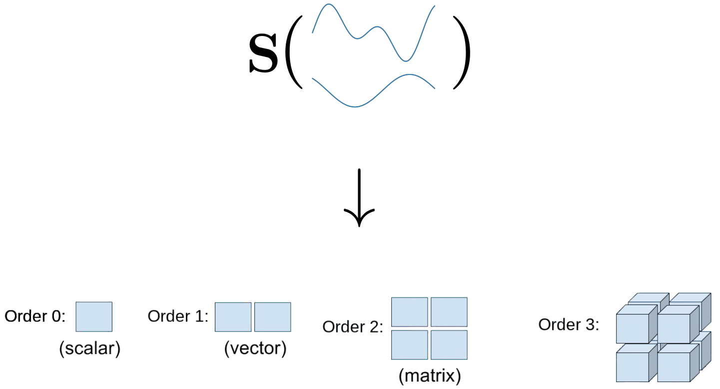


The signature transform



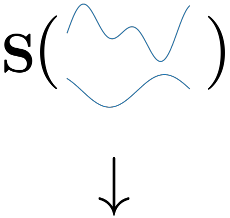
- Time series: $X \in \mathbb{R}^{2 \times 100}$

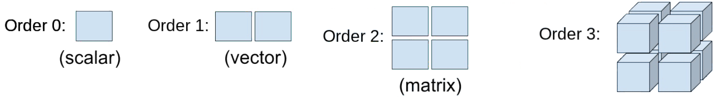
The signature transform



- Time series: $X \in \mathbb{R}^{2 \times 100}$
- Signature: $S^{(0)} \in \mathbb{R}^0$, $S^{(1)} \in \mathbb{R}^2$, $S^{(2)} \in \mathbb{R}^{2 \times 2}$, $S^{(3)} \in \mathbb{R}^{2 \times 2 \times 2}$, etc.

The signature transform

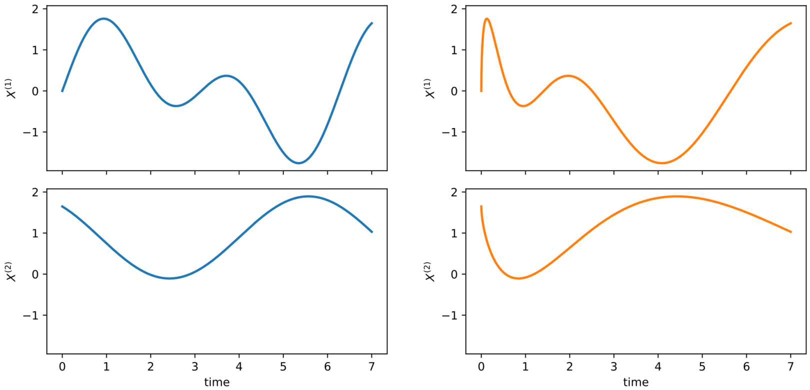




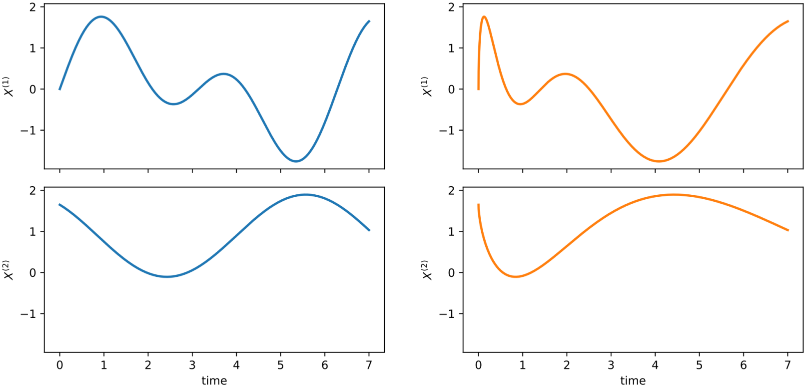
- Time series: $X \in \mathbb{R}^{2\times 100}$
- Signature: $S^{(0)} \in \mathbb{R}^0$, $S^{(1)} \in \mathbb{R}^2$, $S^{(2)} \in \mathbb{R}^{2\times 2}$, $S^{(3)} \in \mathbb{R}^{2\times 2\times 2}$, etc.

$$\mathbf{S}_{[0,1]}(X) = (1, S^{(1)}_{[0,1]}(X), S^{(2)}_{[0,1]}(X), S^{(3)}_{[0,1]}(X), \dots)$$

The signature transform



The signature transform



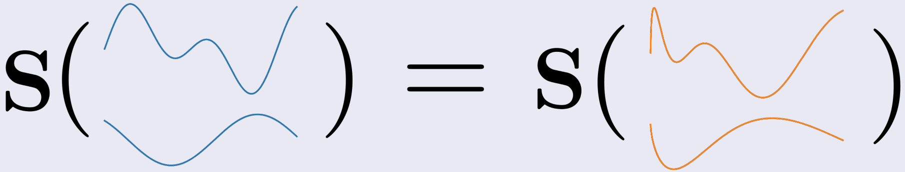
→ Same paths, but different time parametrizations!

The signature transform

Fundamental property #1 of the signature transform

- **Intrinsic characterization** of the path, ignoring translation and time reparametrization: let φ be a reparametrisation

$$\mathbf{S}_{[a,b]}(X_{\varphi(.)}) = \mathbf{S}_{[\varphi(a),\varphi(b)]}(X)$$



The signature transform

Fundamental property #2 of the signature transform

- The space of signatures is a non compact Lie group under $\otimes$ operation[1].

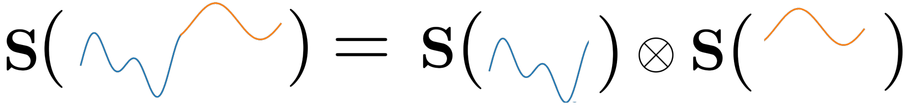
[1] $\otimes$ is an abuse of notation: it should be denoted $\boxtimes$ as it is different from the classical tensor product $\otimes$.

The signature transform

Fundamental property #2 of the signature transform

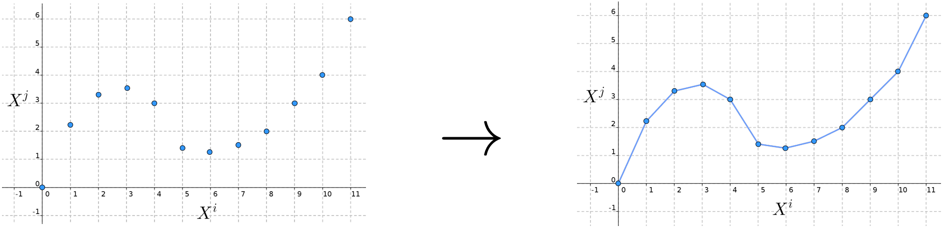
- The space of signatures is a non compact Lie group under $\otimes$ operation[1].
- This $\otimes$ operation is related to the concatenation of two paths through the so-called Chen relation: let $0 \leq u \leq 1$

$$\mathbf{S}_{[0,1]}(X \star Y) = \mathbf{S}_{[0,u]}(X) \otimes \mathbf{S}_{[u,1]}(Y)$$



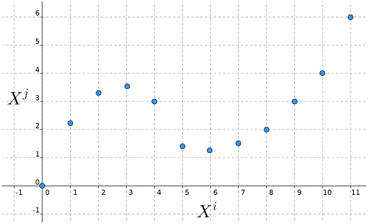
[1] $\otimes$ is an abuse of notation: it should be denoted $\boxtimes$ as it is different from the classical tensor product $\otimes$.

The signature transform

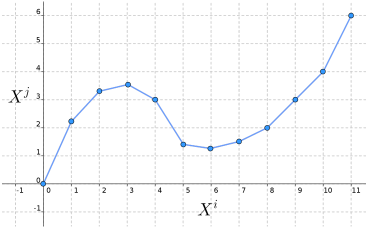


$\rightarrow$ linear interpolation

The signature transform



$\longrightarrow$



$\rightarrow$ linear interpolation

$\rightarrow$ why? easy to compute because of Chen relation

Averaging time series

Motivation

Why averaging using the signature?

- Invariance to time reparametrization (Fundamental prop. #1)

Motivation

Why averaging using the signature?

- Invariance to time reparametrization (Fundamental prop. #1)
- Suitable for multi-dimensional structure

Motivation

Why averaging using the signature?

- Invariance to time reparametrization (Fundamental prop. #1)
- Suitable for multi-dimensional structure
- deal with time series of different lengths
- deal with missing values

Motivation

Why averaging using the signature?

- Invariance to time reparametrization (Fundamental prop. #1)
- Suitable for multi-dimensional structure
- deal with time series of different lengths
- deal with missing values

$\rightarrow$ three averaging approaches designed

Motivation

Why averaging using the signature?

- Invariance to time reparametrization (Fundamental prop. #1)
- Suitable for multi-dimensional structure
- deal with time series of different lengths
- deal with missing values

$\rightarrow$ three averaging approaches designed

$\rightarrow$ all three are based on Fundamental prop. #2: signature $\in$ Lie group

Averaging the signature

Approach 1: the Log Euclidean barycenter

Let $x_1, \ldots, x_n$ be n multidimensional time series
Let $X_1, \ldots, X_n$ be their linear interpolations
Denote $\mathbb{X}_i = \mathbf{S}_{[0,1]}(X_i)$.

Averaging the signature

Approach 1: the Log Euclidean barycenter

Let $x_1, \ldots, x_n$ be n multidimensional time series
Let $X_1, \ldots, X_n$ be their linear interpolations
Denote $\mathbb{X}_i = \mathbf{S}_{[0,1]}(X_i)$.

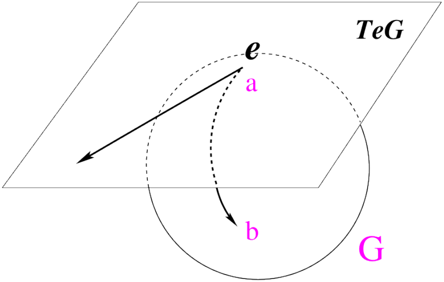


Figure: Lie Group G and its Lie algebra $T_e G$.

Averaging the signature

Approach 1: the Log Euclidean barycenter

Let $x_1, \ldots, x_n$ be n multidimensional time series
Let $X_1, \ldots, X_n$ be their linear interpolations
Denote $\mathbb{X}_i = \mathbf{S}_{[0,1]}(X_i)$.

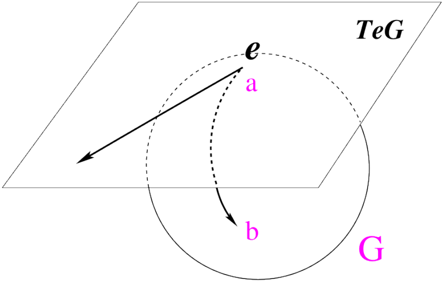


Figure: Lie Group G and its Lie algebra $T_e G$.

Definition (Approach 1)

The **Log Euclidean barycenter** is

$$\bar{\mathbb{X}}_{LE} = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log(\mathbb{X}_i)\right).$$

Averaging the signature

Approach 2: the group exponential barycenter

Averaging the signature

Approach 2: the group exponential barycenter

Definition (Approach 2)

The **group exponential barycenter** is defined iteratively as following:

- Step 0: initialize $\bar{\mathbb{X}}_{(0)}$.

Averaging the signature

Approach 2: the group exponential barycenter

Definition (Approach 2)

The **group exponential barycenter** is defined iteratively as following:

- Step 0: initialize $\bar{\mathbb{X}}_{(0)}$.
- Step k:

$$\bar{\mathbb{X}}_{(k+1)} = \bar{\mathbb{X}}_{(k)} \otimes \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log(\bar{\mathbb{X}}_{(k)}{}^{-1} \otimes \mathbb{X}_i)\right).$$

Averaging the signature

Approach 2: the group exponential barycenter

Definition (Approach 2)

The **group exponential barycenter** is defined iteratively as following:

- Step 0: initialize $\bar{\mathbb{X}}_{(0)}$.
- Step k:

$$\bar{\mathbb{X}}_{(k+1)} = \bar{\mathbb{X}}_{(k)} \otimes \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log(\bar{\mathbb{X}}_{(k)}{}^{-1} \otimes \mathbb{X}_i)\right).$$

→ Good news: under mild conditions, this algorithm converges!

Averaging the signature

Approach 2: the group exponential barycenter

Definition (Approach 2)

The **group exponential barycenter** is defined iteratively as following:

- Step 0: initialize $\bar{\mathbb{X}}_{(0)}$.
- Step k:

$$\bar{\mathbb{X}}_{(k+1)} = \bar{\mathbb{X}}_{(k)} \otimes \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log(\bar{\mathbb{X}}_{(k)}{}^{-1} \otimes \mathbb{X}_i)\right).$$

$\rightarrow$ Good news: under mild conditions, this algorithm converges!

Pennec & Arsigny (2013). Exponential barycenters of the canonical Cartan connection and invariant means on Lie groups. [PA13, Algorithm 1 & Corollary 5]

Averaging the signature

Approach 3: optimization on path space

$\rightarrow$ reconstruction of X given $\mathbf{S}(X)$: not easy!

Averaging the signature

Approach 3: optimization on path space

$\rightarrow$ reconstruction of X given $\mathbf{S}(X)$: not easy!

$x_1, \ldots, x_n$ be D-dimensional. $X_1, \ldots, X_n$ their linear interpolations.

Averaging the signature

Approach 3: optimization on path space

$\rightarrow$ reconstruction of X given $\mathbf{S}(X)$: not easy!

$x_1, \dots, x_n$ be D-dimensional. $X_1, \dots, X_n$ their linear interpolations.

Definition (Approach 3)

The barycenter is the path $\bar{X} \in \mathbb{R}^{D \times L}$ such that

$$\bar{X} = \arg\min_X \sum_{i=1}^{n} d(\mathbf{S}(X), \mathbf{S}(X_i))$$

Averaging the signature

Approach 3: optimization on path space

$\to$ reconstruction of X given $\mathbf{S}(X)$: not easy!

$x_1, \ldots, x_n$ be D-dimensional. $X_1, \ldots, X_n$ their linear interpolations.

Definition (Approach 3)

The barycenter is the path $\bar{X} \in \mathbb{R}^{D \times L}$ such that

$$\bar{X} = \arg\min_X \sum_{i=1}^{n} d(\mathbf{S}(X), \mathbf{S}(X_i))$$

- Distance d: euclidean, signature distance.
- Optimization: gradient descent.

Application: clustering time series

Application

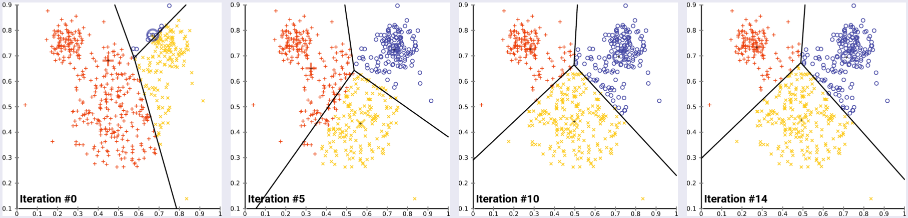
K-means clustering

- **Goal:** cluster n observations $(x_i)_{i=1,\dots,n}$ into K groups.

Application

K-means clustering

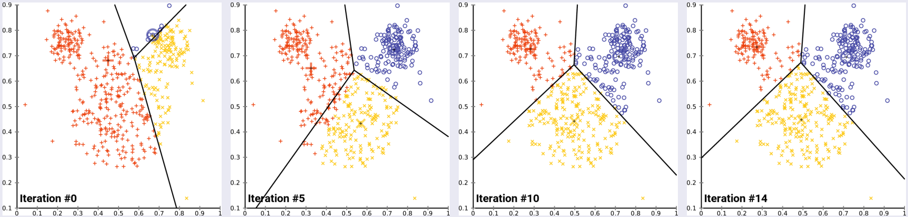
- **Goal:** cluster n observations $(x_i)_{i=1,\ldots,n}$ into K groups.



Application

K-means clustering

- **Goal:** cluster n observations $(x_i)_{i=1,\dots,n}$ into K groups.
- Two parameters to choose:
  1. similarity measure
  2. averaging method

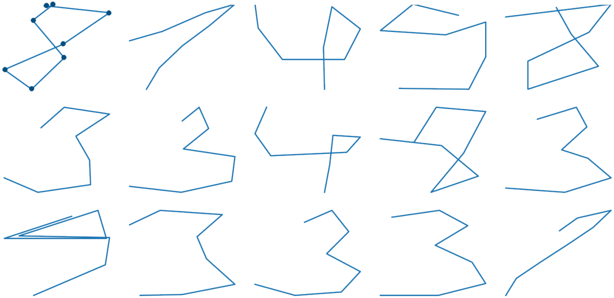


Application

Framework

- Dataset: PenDigits[1]
- 11000 bidimensional time series of length 8.

```
data.shape = (11000, 8, 2)
```



[1] Bagnall et al. 2018. The UEA multivariate time series classification archive.

Application

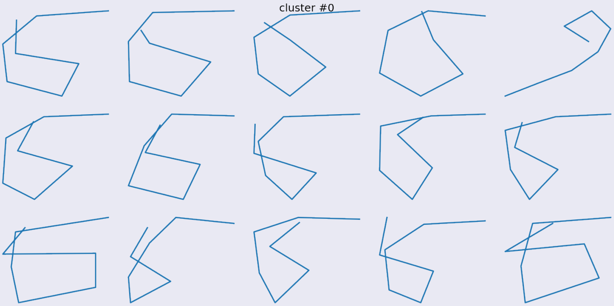


Figure: K-means' cluster #0 with averaging Approach 3.

Application: results

Table: K-means parameters used for benchmark. Pink indicates the use of the signature transform.

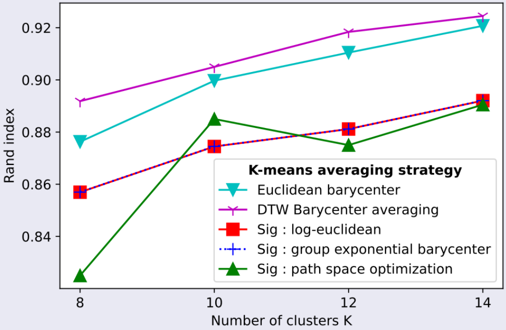
Averaging method	**Similarity measure**
Euclidean barycenter	euclidean
DTW Barycenter Averaging[1]	DTW
Approach 1: log-euclidean	euclidean
Approach 2: group exponential barycenter	euclidean
Approach 3: path space optimization	euclidean

[1] Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. Pattern recognition, 44(3), 678-693. [PKG11]

Application: results

Averaging method	Similarity measure
Euclidean barycenter	euclidean
DTW Barycenter Averaging	DTW
Approach 1: log-euclidean	euclidean
Approach 2: group exponential barycenter	euclidean
Approach 3: path space optimization	euclidean

Rand Index



Conclusion and next steps

Regarding the clustering of PenDigits

Performances **are not** state-of-the-art.

- Combine the signature with other clustering algorithms: e.g. Hierarchical clustering, DBSCAN.

Conclusion and next steps

Regarding the clustering of PenDigits

Performances **are not** state-of-the-art.

- Combine the signature with other clustering algorithms: e.g. Hierarchical clustering, DBSCAN.
- Use different datasets with various shapes: high-dimensional, long time series (reverse engineering).

Conclusion and next steps

Regarding the clustering of PenDigits

Performances **are not** state-of-the-art.

- Combine the signature with other clustering algorithms: e.g. Hierarchical clustering, DBSCAN.
- Use different datasets with various shapes: high-dimensional, long time series (reverse engineering).
- Try other similarity measures, more appropriate for the signature.

Conclusion and next steps

Regarding the clustering of PenDigits

Performances **are not** state-of-the-art.

- Combine the signature with other clustering algorithms: e.g. Hierarchical clustering, DBSCAN.
- Use different datasets with various shapes: high-dimensional, long time series (reverse engineering).
- Try other similarity measures, more appropriate for the signature.

Regarding the averaging approaches

- Extend other methods to the time series framework.
- How to represent a barycenter?

Thank you!

Bibliography

[CK16] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. *arXiv:1603.03788 [cs, stat]*, Mar 2016. arXiv: 1603.03788.

[PA13] Xavier Pennec and Vincent Arsigny. *Exponential Barycenters of the Canonical Cartan Connection and Invariant Means on Lie Groups*, page 123–166. Springer Berlin Heidelberg, 2013.

[PKG11] François Petitjean, Alain Ketterlin, and Pierre Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern Recognition*, 44(3):678–693, Mar 2011.