Principal geodesic analysis for time series encoded with signature features

Raphael Mignot, Marianne Clausel, Konstantin Usevich

Principal Geodesic Analysis for time series encoded with signature features

Raphael Mignot* Marianne Clausel* Konstantin Usevich†

January 13, 2024

We analyze multidimensional time series through the lens of their integrals of various moment orders, constituting their signature. The contribution of this article is to adapt the Principal Geodesic Analysis (PGA), the counterpart for manifolds of the Principal Component Analysis (PCA), to signature features which form a Lie group, by setting an appropriate connection structure. We show that, on both simulations and real data, our approach is more effective than the usual approximation which consists in projecting points of the manifold onto a tangent space and carrying a classical PCA.

Keywords— Geometric statistics, Principal geodesic analysis, Tensor algebra, Iterated integrals signature

Contents

*Université de Lorraine, CNRS, IECL, F-54000 Nancy, France
†Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

1. Introduction

In many scenarios, data is naturally recovered in the form of time series. The analysis of such stream of data has become key in various fields: health related recordings, environmental sciences, economic indicators, financial assets.

Principal Component Analysis. The Principal Component Analysis (PCA) is a ubiquitous method for dimension reduction and which provide useful visual insights. Given d-dimensional centered vector samples $x_1, \dots, x_N$, the goal is to find a sequence of vectors $v_1, \dots, v_K$, that we compute successively by solving

$$v_k = \underset{\substack{\|v\|=1 \\ v \perp v_1, \dots v_{k-1}}}{\arg\min} \sum_{i=1}^{N} d^2(x_i, \pi_v(x_i)) \tag{1}$$

where π_v is the orthogonal projection onto $\mathrm{span}(v)$ and where v_1 does not have an orthogonality requirement. The $(v_k)_k$ are called Principal Directions. We can compress the data by setting $K < d$ and projecting it on $\mathrm{span}(v_1, \dots, v_K)$. If the data belongs to a Euclidean space, Equation (1) have a closed form solution and PCA boils down to the diagonalization of the (symmetric) covariance matrix of $x = (x_1, \dots, x_N)$.

Extension to time series and to manifolds. PCA can be extended to time series in several ways. For instance, for sample index $i = 1, \dots, N$, let $\{x_i(t_1), \dots, x_i(t_T)\}$ be a d-dimensional time series of length T, that is for each i, x_i is a matrix of size $T \times d$. Then [Rao, 1958] and [Tucker, 1958] suggested to flatten each matrix into long vectors of size Td and consider the dataset as a (N, Td) matrix to inject into the usual PCA procedure. Further details on multivariate time series PCA can be found in [Ramsay and Silverman, 2005, Chapter 8].

Another strategy is to consider features of time series instead of the raw data. For instance, in [Cazelles et al., 2020], the authors compute the power spectral densities of the time series.

Those features might lie on a manifold. Thus the usual PCA cannot be applied as it is. An extension of the PCA for manifolds called Principal Geodesic Analysis (PGA) have been developed in [Fletcher et al., 2003]. The problem is defined similarly as in Equation (1), but π_v is now the projection on the geodesic starting from the origin with initial velocity v. Contrary to the Euclidean situation, this new optimization problem does not have a closed form solution in the general case. For instance, the projection π_v might require to be approximated. Because of this, most of the work involving numerical calculation of the PGA, including the original article, relies on an approximation of it: the tangent PGA. It consists in projecting the data onto the tangent space at the origin and performing a classical PCA. Also, note that the PCA is applied on centered data, thus a notion of barycenter must be defined and computed beforehand. In the end, PGA provides another way to extend the PCA to time series.

In this article, we study time series through the lens of their iterated integrals signatures, leveraging the benefits of such transformation and adapting the PGA for this innovative feature.

The signature method. The method consists in the analysis of the so-called iterated integrals signature of a time series. The signature has been originally developed for topological work [Chen, 1957] and later on has been used in the theory of rough paths [Lyons, 1998]. More recently, it has been used for time series analysis [Chevyrev and Kormilitzin, 2016] [Morrill et al., 2020]. In this context, it has obtained state-of-the-art performances in various applications such as handwriting recognition [Yang et al., 2015], medical condition detection [Morrill et al., 2019], human motion [Yang et al., 2022], oceanography [Sugiura and Hosoda, 2020], financial markets [Buehler et al., 2020]. The signature method has shown to be useful for multiple reasons. First, it is naturally suitable for multidimensional time series and can deal with a dataset composed of time series of various lengths T, which is standard in real life scenarios with a lot of missing values. Moreover, it is an intrinsic characterization of time series, ignoring time reparametrization and translations.

Contributions. Our contributions are the following:

- We define an extension of the PCA for the signature space, Proposition 5. Our approach relies on the unique properties inherent to this space, eliminating the need for the approximation involved in tangent PGA.
- We present an algorithm to numerically solve the resulting optimization problem, Algorithm 1, along with an implementation in Python.
- We perform experiments on synthetic and real-life data to illustrate the effectiveness of the method compared to classical approximation method (tangent PGA), Section 5.2.1 and Section 5.2.2.

Notations. Throughout the document, we use the following notations:

- X : multivariate time series
- x : signature of a time series
- N : Number of time series.
- d : Number of components (features) of each multivariate time series.
- T : Length of time series (number of timestamps).
- L : Truncation level of the signature feature.
- $G_{\leqslant L}$ and $\mathfrak{g}_{\leqslant L}$: Lie group of signatures truncated at level L and corresponding Lie algebra.

2. The signature space and its Lie group structure

We first explain how we can encode time series with signature features.

2.1. The signature feature for multivariate time series

We first define the signature of continuous processes.

Definition 1. *Let $X : [0,1] \to \mathbb{R}^d$ be a continuous function of bounded variations, that is $\|X\|_{\mathrm{TV}} < \infty$. The signature of level L of X is*

$$\mathbf{S}_{(L)}(X) := \int_{0 \leqslant t_1 \leqslant \cdots \leqslant t_T \leqslant 1} \cdots \int dX(t_1) \otimes \cdots \otimes dX(t_T) \tag{2}$$
$$= \int_{0 \leqslant t_1 \leqslant \cdots \leqslant t_T \leqslant 1} \cdots \int \dot{X}(t_1) \otimes \cdots \otimes \dot{X}(t_T) dt_1 \dots dt_T$$

where we have use the $\otimes$ notation for the tensor product: let v be an n-ways tensor and w an m-ways tensor, then for any multi-index $I := (i_1, \dots, i_n, i_{n+1}, \dots, i_{n+m})$, we have

$$(v \otimes w)_{i_1,\dots,i_n,i_{n+1},\dots,i_{n+m}} := v_{i_1,\dots,i_n} w_{i_{n+1},\dots,i_{n+m}}. \tag{3}$$

We call signature the infinite collection of signatures at all levels:

$$\mathbf{S}(X) = \{1, \mathbf{S}_{(1)}(X), \mathbf{S}_{(2)}(X), \dots\} \tag{4}$$

where 1 is a convention. In numerical experiments, the computation of the signature is done up to a fixed level L and we denote $\mathbf{S}_{\leqslant L}(X)$ the collection of the first L elements of $\mathbf{S}(X)$.

Example 2. *Let $X : [0,1] \to \mathbb{R}^d; t \mapsto at + b$ be a linear path, with $a, b \in \mathbb{R}^d$. Then*

$$\mathbf{S}_{(k)}(X) = \int_{0 \leqslant t_1 \leqslant \cdots \leqslant t_T \leqslant 1} \cdots \int a \otimes \cdots \otimes a dt_1 \dots dt_T \tag{5}$$
$$= \frac{1}{k!} a \otimes \cdots \otimes a. \tag{6}$$

To apply Definition 1 to a time series X, we need to change the discrete representation of X into a continuous one. This is done by considering the linear interpolation of X (or any other continuous interpolation). Thereafter, if X is a time series, $\mathbf{S}(X)$ denotes the signature of the linear interpolation of X.

Given a continuous function $X : [0,1] \to \mathbb{R}$, its truncated signature $\mathbf{S}_{\leqslant L}(X)$ is an element of a Lie group, that we denote $G_{\leqslant L}$. We now describe the structure of signature space $G_{\leqslant L}$ and need some background.

2.2. Background on Lie groups

A Lie group is a smooth manifold with a smooth group structure, that is the mappings $G \times G \to G, (g,h) \mapsto gh$ and $G \to G, g \mapsto g^{-1}$ are smooth. Examples of Lie groups are the general linear group $\mathrm{GL}_n(\mathbb{R})$, the orthogonal group $O_n(\mathbb{R})$ and the special orthogonal group $\mathrm{SO}_n(\mathbb{R})$. For any $g \in G$, we define the left translation $L_g : G \to G, h \mapsto gh$ and right translation $R_g : G \to G, h \mapsto hg$. The differential $dL_g : T_h G \to T_{gh} G, h \mapsto (dL_g)_h$ gives a natural identification of tangent spaces, where we have denoted $T_h G$ the tangent space at h of G.

For every Lie group G there is an associated Lie algebra $\mathfrak{g}$, a set that can be identified with $T_1 G$ the tangent space at identity.

We call adjoint representation of G the function $\mathrm{Ad} : G \to \mathrm{Aut}(\mathfrak{g}), g \mapsto \mathrm{Ad}_g$ defined such that $\mathrm{Ad}_g : T_1 G \to T_e G, v \mapsto (dL_g)_{g^{-1}} \circ (dR_{g^{-1}})_e(v)$. The derivative of Ad is called the adjoint representation of $\mathfrak{g}$, $\text{ad} : \mathfrak{g} \to \text{End}(\mathfrak{g}), v \mapsto \text{ad}_v := d(\text{Ad})_e(v)$. One can show that for any $v, w \in \mathfrak{g}$, we have $\text{ad}_v(w) = [v, w] := vw - wv$, where $[., .]$ is called the Lie bracket.

The group exponential is defined as follows: let $v \in T_1G$. There exists a unique curve $\gamma_v : \mathbb{R} \to G$ such that $\gamma_v(0) = e$ and $\dot{\gamma}_v(t) = (dL_{\gamma_v(t)})_e(v)$. The group exponential of v is $\exp(v) := \gamma_v(1)$. In finite dimension, the group exponential is a local diffeomorphism at 0. Its local inverse map is called the group logarithm and denoted log. Note that the Riemannian and group exponentials coincides if and only if G admits a Riemannian metric invariant to left and right translations. For instance, the signature space $G_{\leqslant L}$ does not admit such metric.

For a thorough introduction to manifolds see for instance [Tu, 2011] and for more details on Lie groups see [Duistermaat and Kolk, 2012].

2.3. The signature space G and its Lie group structure

In this section, we describe the Lie group structure of the signature space that we denote G in the sequel.We denote $g = (g_{(0)}, g_{(1)}, \ldots, g_{(L)})$ any element of $G_{\leqslant L}$, corresponding to a truncated signature. Note that $G_{\leqslant L}$ is embedded in the truncated tensor algebra of $\mathbb{R}^d$

$$T_{\leqslant L}(\mathbb{R}^d) := \bigoplus_{k=0}^{L} (\mathbb{R}^d)^{\otimes k} = \mathbb{R} \oplus \mathbb{R}^d \oplus (\mathbb{R}^d \otimes \mathbb{R}^d) \oplus \cdots \oplus (\mathbb{R}^d \otimes \cdots \otimes \mathbb{R}^d). \tag{7}$$

$(T_{\leqslant L}(\mathbb{R}^d), +, ., \otimes)$ is an associative algebra and $G_{\leqslant L}$ is called the nilpotent free Lie group of step L. An element $g \in T_{\leqslant L}$ is an element of $G_{\leqslant L}$ if and only if it satisfy a specific property called the shuffle property.

The (non commutative) group product between two elements $g, h \in G_{\leqslant L}$ is, for any $K = 1, \ldots, L$,

$$(gh)_{(K)} := \sum_{k=0}^{K} g_{(k)} \otimes h_{(K-k)}. \tag{8}$$

The neutral element of $G_{\leqslant L}$ is $1 := (1, 0, \ldots, 0)$ and the group inverse of any $g \in G_{\leqslant L}$ is

$$g^{-1} := \sum_{k=0}^{L} (-1)^k (g - 1)^k. \tag{9}$$

The group exponential $\exp : \mathfrak{g}_{\leqslant L} \to G_{\leqslant L}$ is

$$\exp(v) = \sum_{k=0}^{L} \frac{v^k}{k!}. \tag{10}$$

The group logarithm $\log : G_{\leqslant L} \to \mathfrak{g}_{\leqslant L}$ is defined everywhere on $G_{\leqslant L}$ and it verifies

$$\log(g) = \sum_{k=1}^{L} (-1)^{k+1} \frac{(g-1)^k}{k}. \tag{11}$$

Remark 3. *A natural question when it comes to the signature of a path X is: from the signature $S_{\leqslant L}(X)$, can we recover X, up to time reparametrization and tree-likeness? This task is actually difficult and at the moment there isn't a natural procedure for it.*

For a complete exposure to the signature topology and its properties see [Friz and Victoir, 2010, Chapter 7] and [Reutenauer, 1993].

3. Mean and PGA in Lie groups

In this section, we introduce a notion of barycenter (mean) for Lie groups which is a prerequisite to define the PGA. Then, we present the PGA for Lie groups and how we can adapt it to the signature space.

3.1. Mean

The definition of barycenter for Euclidean space $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ cannot be used for manifolds, since in many cases $\bar{x}$ might not belong to the manifold. Take for instance the space of $n \times n$ matrices with non zero determinant GL_n and $x_1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ and $x_2 = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$. We have $x_1, x_2 \in \mathrm{GL}_n$ but the Euclidean barycenter of x_1 and x_2 is $\bar{x} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ which has determinant zero, thus $\bar{x} \notin \mathrm{GL}_n$.

A generalization of the Euclidean barycenter to manifolds is the Fréchet mean: let $(\mathcal{M}, d)$ be a metric space. Given a set of points $x_1, \ldots, x_N \in \mathcal{M}$, the Fréchet mean is the point $\mu \in \mathcal{M}$ such that

$$\mu = \arg\min_{\mu} \sum_{i=1}^{N} d^2(\mu, x_i). \tag{12}$$

This definition can be used for Lie groups. Also, if $d(.,.)$ is a bi-invariant Riemannian metric, then μ is stable by group operations: left and right multiplication, inversion. Stability for the right multiplication means that μy is the Fréchet mean of $\{x_i y\}_{i=1,\ldots,N}$. However, if $d(.,.)$ is not bi-invariant, the stability of μ is not ensured. For such cases, the authors of [Pennec and Lorenzi, 2020] have defined a notion of barycenter on Lie groups called the group mean, see Definition 4.

Definition 4 ([Pennec and Lorenzi, 2020, Definition 11])**.** *Let G be a Lie group with globally defined logarithm and ν a probability measure on it. We say that $\mu \in G$ is a group mean of ν if*

$$0 = \int_G \log(\mu_\nu^{-1} x)\nu(dx).$$

In other words: empirically, for a set of signatures $\{x_1, \ldots, x_N\}$, we look for μ such that vectors v_i in the tangent space at the identity $T_1 G$ have mean zero, where $v_i := \log(\mu^{-1} x_i)$, see Figure 1.

For the Lie group of signatures, we do not have a bi-invariant metric, thus we rely on this definition of barycenter. In [Clausel et al., 2023], we show the existence and uniqueness of the group mean for signatures along with a method to compute numerically the group mean of a finite set of signatures.

3.2. PGA in Lie groups

3.2.1. General framework valid for all Lie groups

Principal Geodesic Analysis has first been introduced in [Fletcher et al., 2003]. The computation of PGA components involve an optimization problem, that is solved using a linear approximation. Our approach to solve the optimization problem associated to PGA is inspired from [Sommer et al., 2014]. The main idea is to not rely on the tangent PGA, which might lead to a too crude approximation of the geometry of the considered manifold.

Now, we introduce the Principal Geodesic Analysis in the specific context of Lie groups, as presented in [Said et al., 2007]. Let G denote the signature Lie group of dimension d. Let

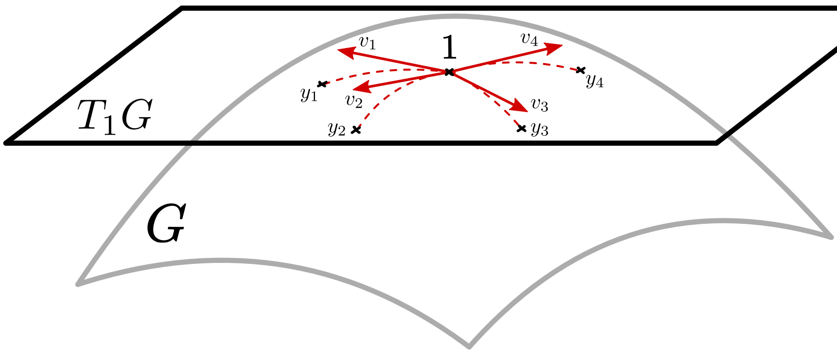


Figure 1: The group mean μ on a Lie group G is such that the sum of the v_i is zero where we have denoted $v_i := \log(y_i) = \log(\mu^{-1}x_i)$ vectors in the tangent space T_1G. Dotted lines are the geodesics on the Lie group G starting from the origin with initial velocity v_i.

$x_1, \ldots, x_N \in G$ be a dataset with group mean μ. Denote $y_i := \mu^{-1}x_i$ the centered data. The goal is to optimize the objective function $F : T_1G \to \mathbb{R}$

$$\underset{\|v\|=1}{\arg\max}\, F(v) \tag{13}$$

defined as

$$F(v) := \sum_{i=1}^{N} \underset{t\in\mathbb{R}}{\arg\min}\, d(y_i, \gamma_v(t))^2 \tag{14}$$

with $d(g, h) = \|\log(g^{-1}h)\|$ and γ_v the geodesic starting from e with velocity v. In other words, we want to minimize the distance between the data points y_i and their projections onto the geodesic γ_v or equivalently, we want to maximize the variance of data points projected onto a geodesic, see Figure 2. Then, we find a second principal geodesic by solving Equation (14) for $g_i^{(1)} := p_{i,1}^{-1}y_i$ instead of y_i, where $p_{i,1}$ is the projection of y_i on the geodesic γ_v. For the k-th geodesic, we solve Equation (14) for $g_i^{(k)} := p_{i,k}^{-1}g_i^{(k-1)}$. Note that we have the following reconstruction of the data points, for any k,

$$x_i = \mu p_{i,1}p_{i,2}\ldots p_{i,k}g_i^{(k)}. \tag{15}$$

As mentioned above, in [Sommer et al., 2014], the authors perform an exact optimization of the PGA on Lie groups that are also differentiable manifolds, in all generality considering as distance in Equation (14) the Riemannian distance. Therefore, they have to use the geodesics derived from their implicit equation (with Christoffel symbols) to perform their calculations. In our case, we shall provide an explicit formula taking benefit only of the Lie group structure and of an explicit closed form of the geodesic.

It has already be done in the work on Lie groups mentioned above [Said et al., 2007]. In this case the Lie group distance involved in Equation (14) is a bi-invariant Riemannian metric allowing to simplify the expression of Equation (14). Unfortunately, in our setting, we do not have a bi-invariant Riemmanian metric on the signature space. Since, the bi-invariance of any geometric characteristic is a natural requirement, we follow the same approach as in our definition of the signature mean and propose an alternative definition of PGA especially dedicated to the signature space.

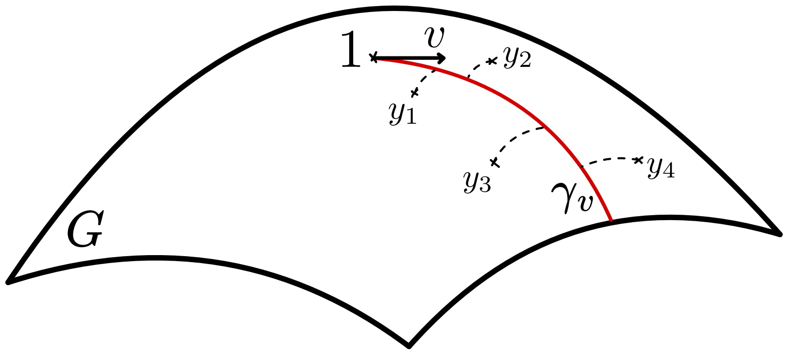


Figure 2: Illustration of Equation (14). To find the first principal direction, we search for an initial velocity v that minimizes the sum of the lengths of the dashed lines (distance to geodesic).

3.2.2. Our framework: an affine connection structure

A key ingredient of our approach to define PGA is the derivation of closed form expression of the geodesics, which simplify drastically the calculations. To this end, we need to choose a connection. Among the natural family of bi-invariant connections suggested in [Cartan, 1926], we choose the canonical symmetric Cartan-Schouten connection. The CCS connexion is the most natural one.

Indeed, as pointed out in [Pennec and Lorenzi, 2020], when there exists a bi-invariant metric on the Lie group, the canonical Cartan Schouten (CCS) connection is the Levi-Civita connection of that metric and when there is not, the CCS connection still exists.

We give now the principal characteristics of this connexion that we shall use in the sequel. The derived geodesics going through the identity with initial velocity v are of the form $\exp(tv)$ and the parallel transport along $\exp(tv)$ is, for any $v, w \in \mathfrak{g}_{\leqslant L}$,

$$\Pi_{1 \to \exp(v)} w = (dL_{\exp(v/2)})_{\exp(v/2)} (dR_{\exp(v/2)})_1 w. \tag{16}$$

Group exponential at $g \in G_{\leqslant L}$ is, for any $v \in T_g(G_{\leqslant L})$,

$$\exp_g(v) := g \exp((dL_{g^{-1}})_g v) \tag{17}$$

and group logarithm at $g \in G_{\leqslant L}$ is, for any $h \in G_{\leqslant L}$,

$$\log_g(h) = (dL_g)_1 \log(g^{-1} h). \tag{18}$$

See [Pennec and Lorenzi, 2020] for further details on the CCS connection.

4. Extension of PCA for signature

4.1. Estimation of Principal Geodesics in the signature space

In view of the previous section we see that the core of the PGA algorithm is solving the optimization problem defined in Equation (13). To solve Problem 13, we need to calculate the gradient of the objective function F.

Proposition 5. *Denote $f_i(v) := \arg\min_{t\in\mathbb{R}} h_i(v,t)$ with $h_i(v,t) := d(y_i, \gamma_v(t))^2$. For any given inner product $\langle\cdot,\cdot\rangle_1$ on $T_1(G_{\leqslant L})$, we have for any $v, u \in T_1(G_{\leqslant L})$,*

$$(\nabla F)_v u = -\sum_{i=1}^{N} \frac{1}{\partial_2^2 h(v, f_i(v))} D_u \partial_2 h(v, f_i(v)) \tag{19}$$

where

$$\partial_2^2 h(v,t) = \left\langle (d\log \circ dR_{ye^{-\frac{1}{2}tv}} \circ dL_{e^{-\frac{1}{2}tv}})(v), v \right\rangle_1 + \left\langle (d\log \circ dL_{e^{-\frac{1}{2}tv}ye^{-\frac{1}{2}tv}})(v), v \right\rangle_1 \tag{20}$$

and

$$\begin{aligned} D_u \partial_2 h(v,t) = &\left\langle (d\log) \circ (dR_{ye^{-1/2tv}}) \circ (d\exp)_{-1/2tv}(-\frac{1}{2}tu), v \right\rangle_1 \\ &+ \left\langle (d\log) \circ (dL_{e^{-1/2tv}y}) \circ (d\exp)_{-1/2tv}(-\frac{1}{2}tu), v \right\rangle_1 \\ &+ \left\langle \log(e^{-\frac{1}{2}tv}ye^{-\frac{1}{2}tv}), 1 \right\rangle_1. \end{aligned} \tag{21}$$

Before proving this proposition, we show how it is useful for our purpose. In order to solve Equation (13), we proceed as following. Initialize $v_k^{(0)}$ such that $\|v_k^{(0)}\| = 1$, then find $t_{k,0,i} := \arg\min_t h_i(v_k^{(0)}, t)$ for all $i = 1, \dots, N$. Now, repeat the following two steps until convergence:

i. **Find k-th Principal Direction step.** Denote $T_{k,l} := \{t_{k,l,1}, \dots, t_{k,l,N}\}$ and $F_{T_{k,l}}(v) := \sum_{i=1}^N d(y_i, \gamma_v(t_{k,l,i}))^2$ for any integers k, l. Perform one step of Gradient Descent:

$$v_k^{(l+1)} \leftarrow v_k^{(l)} - \alpha(\nabla F_{T_k})_{v_k^{(l)}}$$
$$v_k^{(l+1)} \leftarrow \frac{v_k^{(l+1)}}{\|v_k^{(l+1)}\|}$$

where index l denotes gradient descent steps.

ii. **Projection step.** For a fixed $v_k^{(l)}$, find $t_{k,l,i} := \arg\min_{t\in\mathbb{R}} h_i(v_k^{(l)}, t)$ for all $i = 1, \dots, N$: initialize $t^{(0)} \in \mathbb{R}$ and update the value of $t^{(j)}$ in the opposite direction of the gradient:

$$t^{(j+1)} \leftarrow t^{(j)} - \alpha \partial_2 h_i(v_k^{(l)}, t^{(j)}). \tag{22}$$

To sum up, we fix a direction $v \in T_1(G_{\leqslant L})$, project the data on the geodesic with velocity v, then fix a new direction, project again, etc. See Algorithm 1.

4.2. Proof of Proposition 5

4.2.1. Preliminary result

Before proving Proposition 5, we need a preliminary result.

Lemma 6. *Let $y \in G$. Assume that for each $v = (v_1, \dots, v_d) \in T_1G$ with $\|v\| = 1$, the following $f : T_1G \to \mathbb{R}$ exists and is unique:*

$$f(v) := \arg\min_{t\in\mathbb{R}} h(v,t)$$

Algorithm 1: PRINCIPAL GEODESIC ANALYSIS OF SIGNATURES

Input: A batch of N signatures $x_i := \mathbf{S}_{\leqslant L}(X_i)$
Number of geodesics k to keep
Output: $v_1, \ldots, v_K$ the first K geodesic directions

```
Compute group mean μ of input data {x_1, ..., x_N}
Center data: x_i ← μ^{-1} x_i
for k = 1, ..., K do
    Initialize v_k
    Repeat the following two steps until convergence:
    1. Update v_k^(l) : one step in the opposite direction of gradient, obtain v_k^(l+1)
    2. Project x_i on geodesic starting from identity with initial velocity v_k^(l+1)
return v_1, ..., v_K
```

where $h : T_1G \times \mathbb{R}, h(v,t) := d(y, \gamma_v(t))^2$. *Assume that* $\partial_2^2 h(v, f(v)) \neq 0$. *Then, we have*

$$\nabla f(v) = -\lambda^{-1} \begin{pmatrix} \frac{\partial}{\partial v_1}\partial_2 h(v, f(v)) & \ldots & \frac{\partial}{\partial v_d}\partial_2 h(v, f(v)) \end{pmatrix}^T \tag{23}$$

where $\lambda := \partial_2^2 h(v, f(v))$.

Proof of Lemma 6. We fix a $\tilde{v} \in T_1G$. Observe that by definition $\partial_2 h(\tilde{v}, f(\tilde{v})) = 0$. If $\partial_2^2 h(\tilde{v}, f(\tilde{v})) \neq 0$ then the implicit function theorem asserts the existence of a unique continuous mapping $\psi : \mathcal{V}_{\tilde{v}} \to \mathbb{R}$ with $\mathcal{V}_{\tilde{v}}$ a neighborhood of $\tilde{v}$, such that for any $v \in \mathcal{V}_{\tilde{v}}$ we have $\partial_2 h(v, \psi(v)) = 0$. By uniqueness of ψ, we have $\psi \equiv f|_{\mathcal{V}_{\tilde{v}}}$. Let us differentiate the previous equation. Using the chain rule, we have

$$\begin{pmatrix} \frac{\partial}{\partial v_1}\partial_2 h(v, f(v)) + \partial_2^2 h(v, f(v)) \frac{\partial}{\partial v_1} f(v) \\ \vdots \\ \frac{\partial}{\partial v_d}\partial_2 h(v, f(v)) + \partial_2^2 h(v, f(v)) \frac{\partial}{\partial v_d} f(v) \end{pmatrix} = 0 \tag{24}$$

and we obtain the result. $\square$

4.2.2. A first order derivative formula

We want apply Lemma 6 successively with $f_i(v) := \arg\min_{t \in \mathbb{R}} h_i(v,t)$ where $h_i(v,t) := d(\gamma_v(t), y_i)$. We first need to give a meaning to $\partial_2 h(v,t)$.

On a Riemannian manifold, we have

$$h(v,t) := d(\gamma_v(t), y)^2 = \|\mathrm{Log}_{e^{tv}} y\|^2 \tag{25}$$

where Log is the Riemannian logarithm and thus using Lemma 14,

$$\partial_2 h(v,t) = -2 \left\langle \mathrm{Log}_{e^{tv}}(y), \frac{d}{dt} e^{tv} \right\rangle. \tag{26}$$

By analogy with Equation (26) valid in the Riemannian case, we set as a definition on a Lie group equipped with a connection and an inner product $\langle ., . \rangle_1$ defined on the tangent space at the identity T_1G, that

$$\partial_2 h(v,t) := -2 \left\langle \log_{e^{tv}}(y), \frac{d}{dt} e^{tv} \right\rangle_{e^{tv}} \tag{27}$$

where $\langle .,. \rangle_g$ for any $g \in G$ is the parallel transport of $\langle .,. \rangle_1$, that is for any $g \in G$ and $u, v \in T_g G$,

$$\langle u, v \rangle_g := \langle \Pi_{g \to 1} u, \Pi_{g \to 1} v \rangle_1 \tag{28}$$

with $\Pi_{g \to 1}$ parallel transport of the connection.

We now show that the right hand side of Equation (27) can be reduced to a simpler form when the Lie group is $G_{\leqslant L}$.

Lemma 7. *We use Equation (27) as a definition for $\partial_2 h(v,t)$. Then, for the signature space $G_{\leqslant L}$ equipped with the CCS connection, we have for any $v \in T_1(G_{\leqslant L})$ and $t \in \mathbb{R}$,*

$$\partial_2 h(v,t) = -2 \left\langle \log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}), v \right\rangle_1 . \tag{29}$$

Proof. **Step 0.** On $G_{\leqslant L}$ equipped with the CCS connection, Equation (27) is equivalent to

$$\partial_2 h(v,t) := -2 \left\langle d(L_{e^{tv}})_1 \log(e^{-tv} y), d(L_{e^{tv}})_1 v \right\rangle_{e^{tv}} \tag{30}$$

where we have used that $\log_{e^{tv}}(y) = (dL_{e^{tv}})_e \log(e^{-tv} y)$, see Section 3.2.2, and $\frac{d}{dt} e^{tv} = (dL_{e^{tv}})_e v$.
Step 1: use parallel transport to give a meaning to inner product. Using the expression of the parallel transport of the CCS connection, see Equation (16), we have for any $g \in G_{\leqslant L}$, $\Pi_{g \to 1} : T_g(G_{\leqslant L}) \to T_1(G_{\leqslant L})$ with

$$\Pi_{g \to 1} = (dL_{g^{-1/2}})_{g^{1/2}} (dR_{g^{-1/2}})_g \tag{31}$$

where we have denoted $g^\alpha := \exp(\alpha \log g)$ for any real value α. Thus, we have

$$\begin{aligned} \partial_2 h(v,t) &= -2 \left\langle \Pi_{e^{tv} \to 1} d(L_{e^{tv}})_1 \log(e^{-tv} y), \Pi_{e^{tv} \to 1} d(L_{e^{tv}})_1 v \right\rangle_1 & (32) \\ &= -2 \left\langle d(L_{e^{\frac{1}{2}tv}}) d(R_{e^{-\frac{1}{2}tv}}) \log(e^{-tv} y), d(L_{e^{\frac{1}{2}tv}}) d(R_{e^{-\frac{1}{2}tv}}) v \right\rangle_1 & (33) \end{aligned}$$

where we have used the commutativity of dL and dR (Lemma 11) in the last equation and that $dL_g \circ dL_h = dL_{gh}$ (Lemma 12). In other words, from the definition of Ad,

$$\partial_2 h(v,t) = -2 \left\langle \mathrm{Ad}(e^{\frac{1}{2}tv}) \log(e^{-tv} y), \mathrm{Ad}(e^{\frac{1}{2}tv}) v \right\rangle_1 . \tag{34}$$

Step 2: simplify the formula using the fact we have explicit expressions in the case of the signature. Using Lemma 13 and that $g \log(h) g^{-1} = \log(g h g^{-1})$, we have

$$\begin{aligned} \mathrm{Ad}(e^{\frac{1}{2}tv}) \log(e^{-tv} y) &= e^{\frac{1}{2}tv} \log(e^{-tv} y) e^{-\frac{1}{2}tv} & (35) \\ &= \log(e^{\frac{1}{2}tv} e^{-tv} y e^{-\frac{1}{2}tv}) & (36) \\ &= \log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}). & (37) \end{aligned}$$

Using Lemma 13 and that $[v,v] = 0$, we have

$$\begin{aligned} \mathrm{Ad}(e^{\frac{1}{2}tv})(v) &= \exp(\mathrm{ad}_{\frac{1}{2}tv})(v) & (38) \\ &= \left(\mathrm{Id} + \mathrm{ad}_{\frac{1}{2}tv} + \frac{1}{2} \mathrm{ad}_{\frac{1}{2}tv} \circ \mathrm{ad}_{\frac{1}{2}tv} + \cdots + \frac{1}{L!} \mathrm{ad}_{\frac{1}{2}tv} \circ \cdots \circ \mathrm{ad}_{\frac{1}{2}tv} \right)(v) & (39) \\ &= v + \frac{1}{2} t[v,v] + \frac{1}{8} t^2 [v,[v,v]] + \cdots + \frac{1}{2^L L!} t^L [v, \ldots [v,v]] & (40) \\ &= v. & (41) \end{aligned}$$

Injecting Equations (37) and (41) into Equation (34) gives

$$\left\langle d(L_{e^{tv}})_1 \log(e^{-tv}y), d(L_{e^{tv}})_1 v\right\rangle_{e^{tv}} = \left\langle \log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}), v\right\rangle_1 \tag{42}$$

and thus

$$\partial_2 h(v,t) = -2\left\langle \log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}), v\right\rangle_1 . \tag{43}$$

□

We now have all the tools to prove the main result.

Proof of Proposition 5. Denote $f_i(v) := \arg\min_{t\in\mathbb{R}} h_i(v,t)$ with $h_i(v,t) := d(\gamma_v(t), y_i)^2$. That is,

$$\nabla F(v) = \sum_{i=1}^{N} \nabla f_i(v) \tag{44}$$

where F is our main objective function, given by Equation (14). To compute each $\nabla f_i(v)$, we shall use Lemma 6 which requires to differentiate $\partial_2 h$. We use the expression of $\partial_2 h$ as given in Lemma 7. Differentiating Equation (29) with respect to t gives:

$$\begin{aligned}\partial_2^2 h(v,t) = -2\Bigg(&-\frac{1}{2}\left\langle (d\log \circ dR_{ye^{-\frac{1}{2}tv}} \circ dL_{e^{-\frac{1}{2}tv}})(v), v\right\rangle_1 \\ &-\frac{1}{2}\left\langle (d\log \circ dL_{e^{-\frac{1}{2}tv}y} \circ dL_{e^{-\frac{1}{2}tv}})(v), v\right\rangle_1 \Bigg).\end{aligned} \tag{45}$$

That is

$$\partial_2^2 h(v,t) = \left\langle (d\log \circ dR_{ye^{-\frac{1}{2}tv}} \circ dL_{e^{-\frac{1}{2}tv}})(v), v\right\rangle_1 + \left\langle (d\log \circ dL_{e^{-\frac{1}{2}tv}ye^{-\frac{1}{2}tv}})(v), v\right\rangle_1 . \tag{46}$$

Differentiating Equation (29) with respect to v gives:

$$\begin{aligned}D_u \partial_2 h(v,t) = &\left\langle (d\log) \circ (dR_{ye^{-1/2tv}}) \circ (d\exp)_{-1/2tv}(-\frac{1}{2}tu), v\right\rangle_1 \\ &+ \left\langle (d\log) \circ (dL_{e^{-1/2tv}y}) \circ (d\exp)_{-1/2tv}(-\frac{1}{2}tu), v\right\rangle_1 \\ &+ \left\langle \log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}), 1\right\rangle_1 .\end{aligned} \tag{47}$$

□

5. Experiments

5.1. Practical implementation

5.1.1. Implementation details

Our PGA is implemented in Python, following the steps of Algorithm 1. The time series we use in our experiments are scaled to have a total variation norm equal to one. The computation of the signature up to level L is done with library `iisignature`[1]. We have implemented the basic group operations: product, inverse, exponential, logarithm. To center the data, we use the group mean algorithm presented in our previous work [Clausel et al., 2023]. To initialize the k-th principal direction v_k, we randomly draw one of the signature instance x^i and we compute its group logarithm. The update step is performed using the Adam optimization strategy [Kingma and Ba, 2014]. The projection step is done using `scipy` root finder function `fsolve()`. To compute ∇F as shown in Proposition 5, we need the differential of $\partial_2 h$. To that end, we have implemented $\partial_2 h$ as expressed in Equation (29) and then its derivative is computed using automatic differentiation via the library `autograd`[2].

For comparison purposes, we also perform the linear PGA (PCA in tangent space), that we denote tPGA. This is done by projecting the data into the tangent space using the group logarithm and then computing the usual (Euclidean) PCA. This last task is done using the scikit-learn implementation of the PCA: `sklearn.decomposition.PCA`.

After performing PGA or tPGA, we obtain the initial velocities $v_1, \dots, v_K$ of the K first principal geodesics (the principal directions). Denote $t_{k,i}$ the scalar value such that the projection of instance y_i onto the k-th principal geodesic is $\pi_{v_k}(y_i) = \exp(t_{k,i} v_k)$. For each principal direction, that is $k = 1, \dots, K$, we look at the standard deviation

$$\mathrm{std}(t_{k,.}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (t_{k,i})^2}. \tag{48}$$

5.1.2. Closed formula for low order cases

In this section, we show that given the truncation level of the signature, the computation time of Algorithm 1 can be greatly diminished, because of the nilpotency of $G_{\leqslant L}$.

Indeed, we show that Equation (29) can be reduced to a simpler form when the truncation level L of the signature is low, allowing us to reduce computational cost. Indeed, the log expression in Equation (29) can be computed using the following result.

Proposition 8. *Let a, b be element of a Lie algebra. Then*

$$\log(e^{a/2} e^b e^{a/2}) = a + b - \frac{1}{24}[a,[a,b]] + \frac{1}{12}[[a,b],b] + \frac{7}{5760}[a,[a,[a,[a,b]]]] + \dots \tag{49}$$

where $[.,.]$ denote the Lie bracket: $[u,v] = u \otimes v - v \otimes u$. This proposition is a special case of the Baker–Campbell–Hausdorff formula where Lie brackets containing an even number of terms vanish.

Using Proposition 8, we have in Equation (29)

$$\log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}) = -tv + u - \frac{1}{24} t^2 [v,[v,u]] - \frac{1}{12} t [[v,u],u] + \frac{7}{5760} t^4 [v,[v,[v,[v,u]]]] + \dots \tag{50}$$

where $u = \log y$.

Now, elements of $\mathfrak{g}_{\leqslant L}$ are nilpotent, that is for any $u \in \mathfrak{g}_{\leqslant L}$ we have $u^{\otimes K} = 0$ for any $K > L$. Thus, if $u, v \in \mathfrak{g}_{\leqslant L}$, then all Lie brackets with more than L terms vanish.

We now detail two specific cases to illustrate this.

[1] See https://github.com/bottler/iisignature
[2] See https://github.com/HIPS/autograd

Example 9. *If truncation level is $L = 2$, Equation (50) is truncated:*

$$\log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}) = -tv + u \tag{51}$$

that is, Equation (29) becomes

$$\partial_2 h(v,t) = 2\left\langle tv - \log y, v\right\rangle_1 = 2t\|v\|^2 - 2\left\langle \log y, v\right\rangle_1. \tag{52}$$

Example 10. *If truncation level is $L = 3$ or $L = 4$, Equation (50) is truncated and Equation (29) becomes*

$$\partial_2 h(v,t) = -2\left\langle -tv + u - \frac{1}{24}t^2[v,[v,u]] - \frac{1}{12}t[[v,u],u], v\right\rangle_1 \tag{53}$$

$$= -2\left\langle \log y, v\right\rangle_1 + t\left\langle 2v + \frac{1}{6}[[v,\log y],\log y], v\right\rangle_1 + \frac{1}{12}t^2\left\langle [v,[v,\log y]], v\right\rangle_1. \tag{54}$$

Thus,

$$\frac{d}{dt}\log(e^{-\frac{1}{2}tv} y e^{-\frac{1}{2}tv}) = -v - \frac{1}{12}[[v,u],u] - \frac{1}{12}t[v,[v,u]] \tag{55}$$

5.2. Numerical results

In this section, we show how our method helps us get better insights on set of time series compared to the tangent PGA. First, we take a glance at a simulated dataset from a random walk model. Then, we look at a real life dataset: population census in the United States over the 20th century.

5.2.1. Simulated data

Random walks with trend. We generate two clusters of time series in the following way: if X is a member of cluster $j \in \{1,2\}$, then

$$X(t) = \sum_{k=1}^{t} \varepsilon(k) \tag{56}$$

where $\varepsilon(k)$ are iid $\mathcal{N}(\mu_j, \mathrm{Id})$ and

$$\mu_j = \begin{cases} (-1,1) \text{ if } j = 1 \\ (1,1) \text{ if } j = 2 \end{cases}. \tag{57}$$

Note that this model is not stationary. We simulate 20 bivariate time series of length 6 using this toy model (10 in each cluster). Those are shown in Figure 3. We perform PGA using the signatures up to level $L = 5$. Also, we perform a tangent PGA for comparison purposes. Cumulated dispersion of the projections onto geodesics (see Equation (48)) are shown in Figure 4 and projections onto the first two principal geodesics in Figure 5.

We can see that as the number of components increases, the PGA explain more variance than the tangent PGA, which seems to stabilize after PG3. Also, observe that for both methods, projections onto the first two principal geodesics can be linearly separated in a way that we recover the two true clusters. Using a simple clustering strategy such as the Nearest Neighbors, only one instance might be assigned to the wrong cluster.

Overall, on this dataset the PGA seems to be more relevant than its linear counterpart since it explains more variance and both methods separate the data similarly.

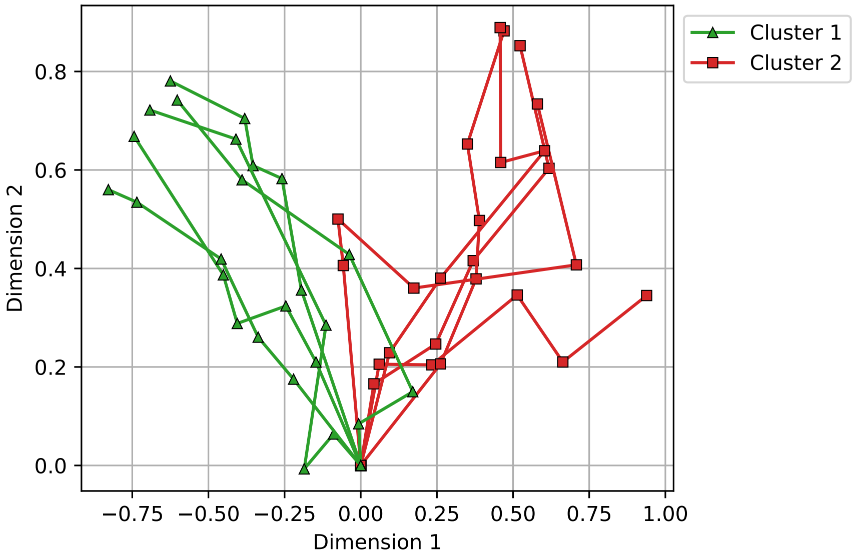


Figure 3: Simulated data (20 bivariate time series).

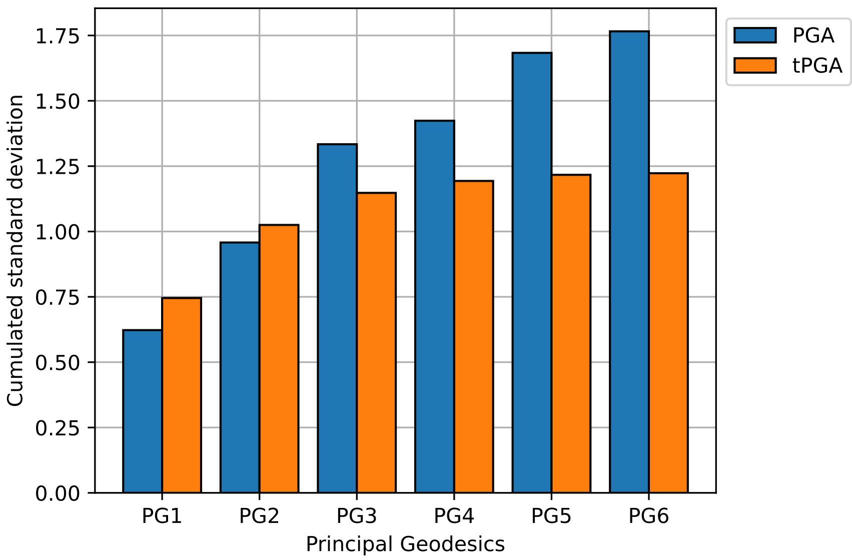
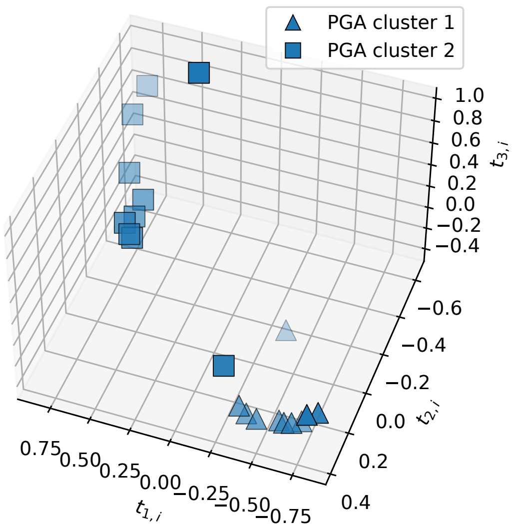
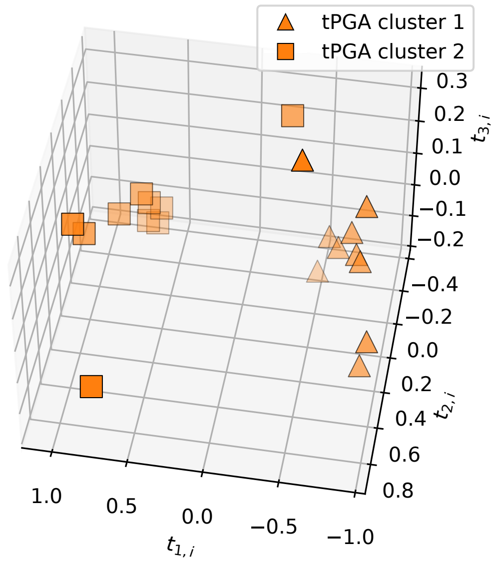


Figure 4: PGA and tPGA results on the simulated data.



(a) PGA



(b) tPGA

Figure 5: Projections on first three principal geodesics of the simulated data.

5.2.2. Real data

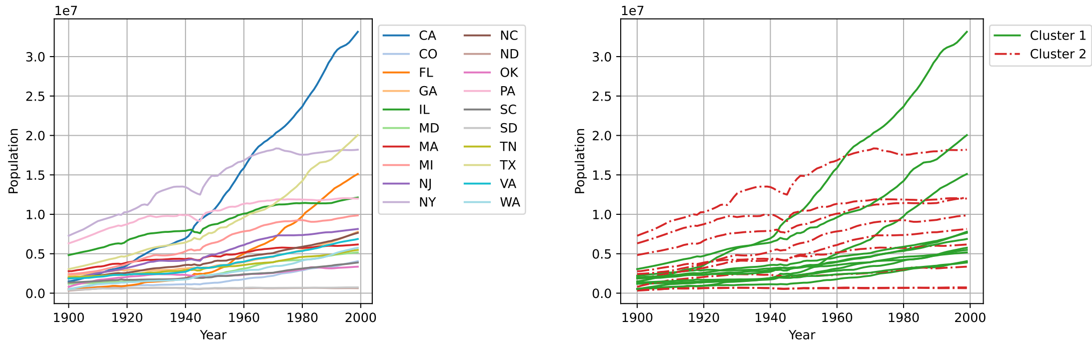
Population dataset. The population dataset[3] is a collection of time series representing the population from 1900-1999 in 20 states of the USA. Following the analysis of [Kalpakis et al., 2001], two clusters can be observed and will be used as the reference clustering: states with an exponential growth of population over time and states with a stabilizing growth. The goal is to retrieve those two clusters in an unsupervised fashion. Curves are shown on Figure 6a and the two clusters in Figure 6b. We compute the signature up to order 5 of the curves and perform PGA. Cumulated dispersion of the projections onto geodesics (see Equation (48)) are shown in Figure 7. We can see that the tangent PGA captures more information on the first two principal geodesics but then stays at the same level of explained variance. On the other hand, the PGA increases for each new principal geodesic added. If we choose to keep the first six principal geodesics, more than a third more information is captured with the PGA compared to the tangent PGA.

Projections onto the first two Principal Geodesics are shown in Figure 8. We can see in Figures 8a and 8b that the two set of curves can be linearly separated on the first geodesic plan, either with the tangent PGA or with the PGA.

6. Conclusion

We have proposed an extension of the PCA for signature features of time series by means of an adaptation of the PGA. We have provided theoretical tools along with a numerical implementation to apply this new method. Also, we have shown through experiments, both on simulated and real data, that our approach provide better results than the usual approximation.

[3] Available at `https://www.census.gov/data.html`.



(a) Raw data.

(b) The two reference clusters.

Figure 6: Population dataset. 20 curves sampled once a year on the period 1900 – 1999.

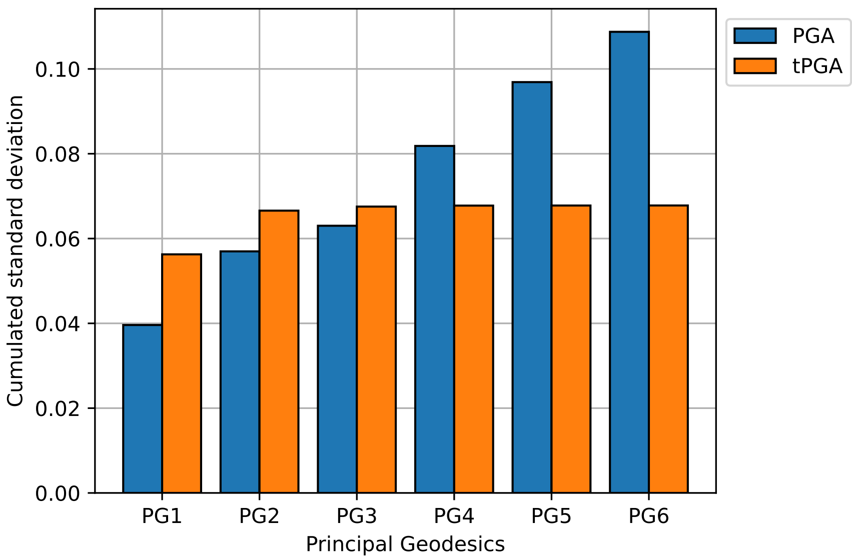
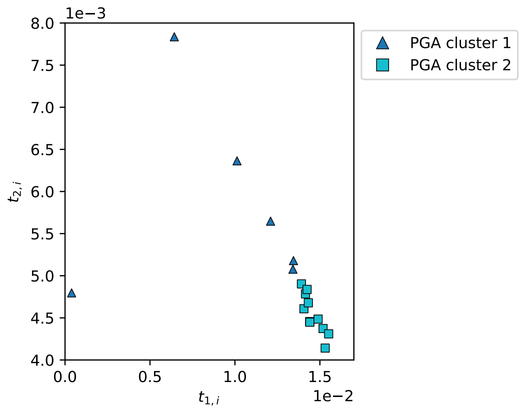
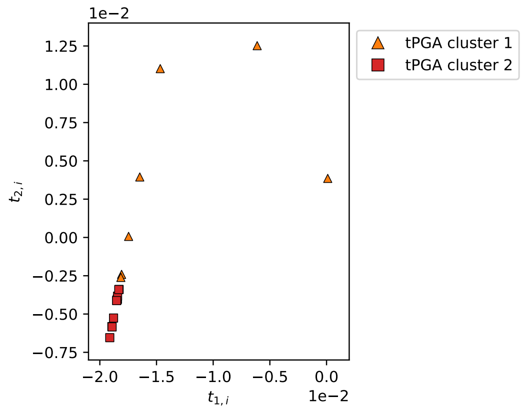


Figure 7: PGA and tPGA results on the Population dataset.



(a) PGA.



(b) Tangent PGA

Figure 8: PGA results on the population dataset. Zoom in on the projection on the first two principal geodesics.

A. Technical Lemmas

Lemma 11. *For any $g, h \in G$ and $v \in T_1 G$, we have*

$$(dL_g)_h \circ (dR_h)_e = (dR_h)_g \circ (dL_g)_e. \tag{58}$$

Proof. Let $g, h, p \in G$. Using the associativity rule of the group, we have

$$(L_g \circ R_h)(p) = g(ph) = (gp)h = (R_h \circ L_g)(p) \tag{59}$$

thus L_g and R_h commute for any g, h. Differentiating Equation (59), with $p = e$ and $v \in T_1 G$, we have

$$((dL_g)_h \circ (dR_h)_e)(v) = ((dR_h)_g \circ (dL_g)_e)(v) \tag{60}$$

thus dL_g and dR_h commute for any g, h. Note that this stays true for any $p \in G$. □

Lemma 12. *For any $g, h \in G$, we have*

$$(dL_{gh})_e = (dL_g)_h \circ (dL_h)_e. \tag{61}$$

Proof. This identity comes from the differentiation of $L_{gh} = L_g \circ L_h$. □

Lemma 13. *Let $g = e^u$ be an element of the signature space $G_{\leqslant L}$ and v an element of the signature Lie algebra $\mathfrak{g}_{\leqslant L}$. Then*

$$\mathrm{Ad}(g)(v) = gvg^{-1} = e^{ad_u} v. \tag{62}$$

Proof. See [Reutenauer, 1993, Theorem 3.2]. □

Lemma 14. *Let x, y be two points on a Riemannian manifold $\mathcal{M}$ and denote* Log *the Riemannian logarithm. Then, we have*

$$\nabla_x \|\mathrm{Log}_x(y)\|^2 = -2\mathrm{Log}_x(y). \tag{63}$$

Proof of Lemma 14. Denote $\Gamma : (-\varepsilon, \varepsilon) \times [0,1] \to \mathcal{M}; \Gamma(s,t) = \gamma_s(t)$ the geodesic variation of γ_s geodesic with start point $\gamma_s(0) = x(s)$, fixed end point $\gamma_s(1) = y$ and velocity $\dot{\gamma}_s(0) = \mathrm{Log}_{x(s)}(y)$. We define the length of any curve γ as

$$\mathrm{Length}(\gamma) := \int_0^1 \|\dot{\gamma}(t)\|_{\mathcal{M}} dt. \tag{64}$$

Denote V the variational field of Γ, i.e. $V(t) := \partial_s|_{s=0} \Gamma(s,t)$ and $c := \|\dot{\gamma}\|_{\mathcal{M}}$. The first variation formula of the length is[4]

$$\left.\frac{d}{ds}\right|_{s=0} \mathrm{Length}(\Gamma_s) = -\frac{1}{c}\int_0^1 \langle V, D_t\dot{\gamma}\rangle\, dt + \frac{1}{c}\langle V(1), \dot{\gamma}(1)\rangle - \frac{1}{c}\langle V(0), \dot{\gamma}(0)\rangle\,. \tag{65}$$

Since γ is a geodesic we have $D_t\dot{\gamma} = 0$. Also, $\partial_s \gamma_s(1) = \partial_s y = 0$ i.e. $V(1) = 0$. Thus,

$$\left.\frac{d}{ds}\right|_{s=0} \mathrm{Length}(\Gamma_s) = -\frac{1}{c}\langle V(0), \dot{\gamma}(0)\rangle \tag{66}$$

$$= -\frac{1}{c}\langle \partial_s|_{s=0}\, \gamma_s(0), \dot{\gamma}(0)\rangle \tag{67}$$

$$= -\frac{1}{c}\left\langle \left.\frac{d}{ds}\right|_{s=0} x(s), \mathrm{Log}_x(y)\right\rangle. \tag{68}$$

Remark that since γ_s is a geodesic, it has constant speed and we have from the definition $\mathrm{Length}(\gamma) = \|\dot{\gamma}_s\| = \|\mathrm{Log}_{x(s)} y\|$. Thus

$$\left.\frac{d}{ds}\right|_{s=0} \|\mathrm{Log}_{x(s)} y\|^2 = -2\left\langle \left.\frac{d}{ds}\right|_{s=0} x(s), \mathrm{Log}_x(y)\right\rangle \tag{69}$$

where we can identify $\nabla_x \|\mathrm{Log}_x y\|^2 = -2\mathrm{Log}_x(y)$. □

Lemma 15. *Let $x, y \in \mathcal{M}$. We have*

$$D_x(\mathrm{Log}_x)(y) = -\left[D(\mathrm{Exp}_x)(\mathrm{Log}_x(y))\right]^{-1} D_x(\mathrm{Exp}_x)(\mathrm{Log}_x(y)). \tag{70}$$

Proof. By definition,

$$\mathrm{Exp}_x(\mathrm{Log}_x(y)) = y. \tag{71}$$

Differentiating with respect to x gives

$$D_x(\mathrm{Exp}_x)(\mathrm{Log}_x(y))) + D(\mathrm{Exp}_x)(\mathrm{Log}_x(y)).D_x(\mathrm{Log}_x)(y) = 0 \tag{72}$$

and rearranging the terms gives the result. □

References

[Buehler et al., 2020] Buehler, H., Horvath, B., Lyons, T., Arribas, I. P., and Wood, B. (2020). A data-driven market simulator for small data environments. *arXiv preprint arXiv:2006.14498.* 1

[Cartan, 1926] Cartan, É. (1926). On the geometry of the group-manifold of simple and semi-groups. *Proc. Akad. Wetensch., Amsterdam*, 29:803–815. 3.2.2

[4] See for instance [Lee, 2018, Theorem 6.3].

[Cazelles et al., 2020] Cazelles, E., Robert, A., and Tobar, F. (2020). The Wasserstein–Fourier distance for stationary time series. *IEEE Transactions on Signal Processing*, 69:709–721. 1

[Chen, 1957] Chen, K. (1957). Integration of paths, geometric invariants and a generalized Baker–Hausdorff formula. *Annals of Mathematics*, pages 163–178. 1

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

[Clausel et al., 2023] Clausel, M., Diehl, J., Mignot, R., Schmitz, L., Sugiura, N., and Usevich, K. (2023). The barycenter in free nilpotent Lie groups and its application to iterated-integrals signatures. *arXiv preprint arXiv:2305.18996*. 3.1, 5.1.1

[Duistermaat and Kolk, 2012] Duistermaat, J. J. and Kolk, J. A. (2012). *Lie groups*. Springer Science & Business Media. 2.2

[Fletcher et al., 2003] Fletcher, P., Lu, C., and Joshi, S. (2003). Statistics of shape via principal geodesic analysis on Lie groups. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I. 1, 3.2.1

[Friz and Victoir, 2010] Friz, P. K. and Victoir, N. B. (2010). *Multidimensional stochastic processes as rough paths: theory and applications*, volume 120. Cambridge University Press. 2.3

[Kalpakis et al., 2001] Kalpakis, K., Gada, D., and Puttagunta, V. (2001). Distance measures for effective clustering of ARIMA time-series. In *Proceedings 2001 IEEE international conference on data mining*, pages 273–280. IEEE. 5.2.2

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. 5.1.1

[Lee, 2018] Lee, J. M. (2018). *Introduction to Riemannian manifolds*, volume 2. Springer. 4

[Lyons, 1998] Lyons, T. (1998). Differential equations driven by rough signals. *Revista Matemática Iberoamericana*, 14(2):215–310. 1

[Morrill et al., 2020] Morrill, J., Fermanian, A., Kidger, P., and Lyons, T. (2020). A generalised signature method for multivariate time series feature extraction. *arXiv preprint arXiv:2006.00873*. 1

[Morrill et al., 2019] Morrill, J., Kormilitzin, A., Nevado-Holgado, A., Swaminathan, S., Howison, S., and Lyons, T. (2019). The signature-based model for early detection of sepsis from electronic health records in the intensive care unit. In *2019 Computing in Cardiology (CinC)*, pages Page–1. IEEE. 1

[Pennec and Lorenzi, 2020] Pennec, X. and Lorenzi, M. (2020). Beyond Riemannian geometry: The affine connection setting for transformation groups. In *Riemannian Geometric Statistics in Medical Image Analysis*, pages 169–229. Elsevier. 3.1, 4, 3.2.2, 3.2.2

[Ramsay and Silverman, 2005] Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis (447 p.)*. Springer series in statistics. Springer, New York, 2nd ed edition. 1

[Rao, 1958] Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14(1):1–17. 1

[Reutenauer, 1993] Reutenauer, C. (1993). *Free Lie algebras*. Oxford University press. 2.3, A

[Said et al., 2007] Said, S., Courty, N., Le Bihan, N., and Sangwine, S. J. (2007). Exact principal geodesic analysis for data on SO(3). In *2007 15th European Signal Processing Conference*, pages 1701–1705. IEEE. 3.2.1, 3.2.1

[Sommer et al., 2014] Sommer, S., Lauze, F., and Nielsen, M. (2014). Optimization over geodesics for exact principal geodesic analysis. *Advances in Computational Mathematics*, 40:283–313. 3.2.1, 3.2.1

[Sugiura and Hosoda, 2020] Sugiura, N. and Hosoda, S. (2020). Machine learning technique using the signature method for automated quality control of Argo profiles. *Earth and Space Science*, 7(9):e2019EA001019. 1

[Tu, 2011] Tu, L. W. (2011). *An Introduction to Manifolds*. Springer New York. 2.2

[Tucker, 1958] Tucker, L. R. (1958). Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23(1):19–23. 1

[Yang et al., 2015] Yang, W., Jin, L., and Liu, M. (2015). Chinese character-level writer identification using path signature feature, DropStroke and deep CNN. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, pages 546–550. IEEE. 1

[Yang et al., 2022] Yang, W., Lyons, T., Ni, H., Schmid, C., and Jin, L. (2022). Developing the path signature methodology and its application to landmark-based human action recognition. In *Stochastic Analysis, Filtering, and Stochastic Optimization: A Commemorative Volume to Honor Mark HA Davis's Contributions*, pages 431–464. Springer. 1