A Forward-Backward algorithm for geodesic PCA of histograms in the Wasserstein space

Jérémie Bigot

Institut de Mathématiques de Bordeaux
Université de Bordeaux

Stat’Learn Conference 2016 - Vannes
April 2016

Ongoing work with Elsa Cazelles, Nicolas Papadakis (Institut de Mathématiques de Bordeaux) and Marco Cuturi, Vivien Seguy (Graduate School of Informatics, Kyoto University)
Statistical analysis of (random) probability measures

Purposes of this talk:

- to propose an algorithmic approach to the notion of Geodesic Principal Component Analysis of probability measures on $\mathbb{R}$ with respect to the Wasserstein metric as proposed in


- (if time permits) to highlight similarities with existing algorithms for GPCA of (discrete) probability measures on $\mathbb{R}^d$ as proposed in

Motivations - Statistical analysis of histograms

Standard PCA in a Hilbert space

The Wasserstein space and its geometric properties

Geodesic PCA in the Wasserstein space

An algorithmic approach to GPCA
Statistical analysis of histograms

For a given name, an histogram represents the proportion of children born with that name per year in France between 1900 and 2013.

Source : INSEE
Statistical analysis of histograms

**Data available:** \( n = 1060 \) histograms of length 114 (number of years) such that

- a name has been given at least 1000 times over the years
- the highest mode of an histogram is between years 1920 and 1990.
Statistical analysis of histograms

**Question**: how to summarize this data set in an efficient way? What is the appropriate framework to define the notions of

- an average histogram?
- the main sources of variability (through PCA-like methods)?
1. Motivations - Statistical analysis of histograms

2. Standard PCA in a Hilbert space

3. The Wasserstein space and its geometric properties

4. Geodesic PCA in the Wasserstein space

5. An algorithmic approach to GPCA
Let $H$ be a separable Hilbert space $(H, \langle \cdot, \cdot \rangle, \| \cdot \|)$, and $x_1, \ldots, x_n$ be $n$ (random) vectors in $H$.

The (functional) **Principal Component Analysis** (PCA) of $x_1, \ldots, x_n \in H$ is carried out by diagonalizing the covariance operator $K : H \rightarrow H$ defined by

$$Kx = \frac{1}{n} \sum_{i=1}^{n} \langle x_i - \bar{x}_n, x \rangle (x_i - \bar{x}_n), \ x \in H,$$

where $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the **Euclidean mean** of $x_1, \ldots, x_n \in H$. 
Standard PCA in a separable Hilbert space

\[ Kx = \frac{1}{n} \sum_{i=1}^{n} \langle x_i - \bar{x}_n, x \rangle (x_i - \bar{x}_n), \quad x \in H, \]

**Main idea**: the eigenvectors of \( K \), associated to the largest eigenvalues, describe the principal modes of data variability around their Euclidean mean \( \bar{x}_n \).

The first and second principal modes of linear variation of the data are defined by the \( H \)-valued curves \( g^{(j)} : \mathbb{R} \to H, \ j = 1, 2 \) given by

\[ g^{(1)}_t = \bar{x}_n + tw_1 \quad \text{and} \quad g^{(2)}_t = \bar{x}_n + tw_2, \quad t \in \mathbb{R}, \]

where \( w_1 \in H \) (resp. \( w_2 \)) is the eigenvector associated to the largest eigenvalue \( \sigma_1 \geq 0 \) (resp. second largest \( \sigma_2 \)) of the covariance operator \( K \).
Standard PCA of histograms in $H = L^2(\mathbb{R})$

Data available: $n = 1060$ histograms $f_1, \ldots, f_n \in L^2(\mathbb{R})$.

Euclidean mean in $L^2(\mathbb{R})$: $\overline{f}_n = \frac{1}{n} \sum_{i=1}^{n} f_i$ is a pdf (probability density function)
Standard PCA of histograms in $H = L^2(\mathbb{R})$

Data available: $n = 1060$ histograms $f_1, \ldots, f_n \in L^2(\mathbb{R})$.

First mode of variation in $L^2(\mathbb{R})$

$$g_t^{(1)} = \bar{f}_n + tw_1 \text{ for } -0.15 \leq t \leq 0.12, \text{ where } w_1 \in L^2(\mathbb{R}).$$

Main issues: $g_t^{(1)}$ is not a pdf, and the $L^2$ metric only accounts for amplitude variation in the data.
Standard PCA of histograms in $H = L^2(\mathbb{R})$

Data available: $n = 1060$ histograms $f_1, \ldots, f_n \in L^2(\mathbb{R})$.

Second mode of variation in $L^2(\mathbb{R})$

$$g_t^{(2)} = \bar{f}_n + tw_2 \text{ for } -0.16 \leq t \leq 0.09,$$

where $w_2 \in L^2(\mathbb{R})$.

Main issues: $g_t^{(2)}$ is not a pdf, and the $L^2$ metric only accounts for amplitude variation in the data.
GPCA of histograms

1. Motivations - Statistical analysis of histograms

2. Standard PCA in a Hilbert space

3. The Wasserstein space and its geometric properties

4. Geodesic PCA in the Wasserstein space

5. An algorithmic approach to GPCA
Beyond the standard PCA of densities in $H = L^2(\mathbb{R})$

- **Drawbacks**: functional PCA of densities $f_1, \ldots, f_n$ in $L^2(\mathbb{R})$ is not always clearly interpretable.

- **An alternative**: to consider that the measures $\nu_1, \ldots, \nu_n$ with pdf $f_1, \ldots, f_n$ belong to the Wasserstein space $W_2$ of probability measures over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ endowed with the Wasserstein distance $d_{W_2}$ (associated to the quadratic cost).

- **Main issue**: the Wasserstein space $W_2$ is not a Hilbert space... but it is a geodesic space with a formal Riemannian structure.
The Wasserstein space $W_2(\Omega)$

- Let $\Omega \subset \mathbb{R}$ be some interval. $W_2(\Omega)$ is the set of probability measures over $(\Omega, \mathcal{B}(\Omega))$ with finite second order moment.

- If $F_\mu$ and $F_\nu$ are the cumulative distribution functions (cdf) of $\mu$ and $\nu$ in $W_2(\Omega)$, then

$$d_{W_2}^2(\mu, \nu) = \int_0^1 (F^-_\nu(y) - F^-_\mu(y))^2 dy = \int_\Omega (F^-_\nu \circ F^-_\mu(x) - x)^2 d\mu(x),$$

if $\mu \in W^{ac}_2(\Omega)$ (the subset of absolutely continuous measures), where $F^-_\nu$ (resp. $F^-_\mu$) is the quantile function of $\nu$ (resp. $\mu$).

- The optimal mapping between $\mu \in W^{ac}_2(\Omega)$ and $\nu$ is $T_* = F^-_\nu \circ F^-_\mu$ such that $\nu = T_* \# \mu$, which is the push-forward of $\mu$ onto $\nu$ via the mapping $T_*$ meaning that

$$\nu(A) = \mu(T_*^{-1}(A)) \text{ for any } A \in \mathcal{B}(\Omega).$$
The pseudo-Riemannian structure of $W_2(\Omega)$

Definitions [Ambrosio et al. (2004)] : for $\mu \in W_{ac}^2(\Omega)$

- the tangent space at $\mu$ is the Hilbert space $(L^2_\mu(\Omega), \langle \cdot, \cdot \rangle_\mu, \| \cdot \|_\mu)$ of real-valued, $\mu$-square-integrable functions on $\Omega$.

- the exponential map $\exp_\mu : L^2_\mu(\Omega) \rightarrow W_2(\Omega)$ and the logarithmic map $\log_\mu : W_2(\Omega) \rightarrow L^2_\mu(\Omega)$ are defined as

\[
\exp_\mu(\nu) = (\text{id} + \nu) \# \mu \quad \text{for} \quad \nu \in L^2_\mu(\Omega),
\]

and

\[
\log_\mu(\nu) = F_{\nu}^{-} \circ F_{\mu} - \text{id} \quad \text{for} \quad \nu \in W_2(\Omega).
\]
An isometric representation of $W_2(\Omega)$

Let $\mu \in W^{ac}_2(\Omega)$. Then, $\exp_\mu L^2_\mu(\Omega) \to W_2(\Omega)$ is an isometry when restricted to a specific subset of admissible functions $\nu$ in $L^2_\mu(\Omega)$.

**Definition**

The set of admissible functions is defined by

$$V_\mu(\Omega) := \log_\mu(W_2(\Omega)) = \{ \log_\mu(\nu); \ \nu \in W_2(\Omega) \} \subset L^2_\mu(\Omega).$$

**Proposition**

One has that $\exp_\mu (V_\mu(\Omega)) = W_2(\Omega)$. Moreover, the exponential map $\exp_\mu$ restricted to $V_\mu(\Omega)$ is an isometric homeomorphism, with inverse given by $\log_\mu$ i.e.

$$d^2_W(\nu_1, \nu_2) = \| \log_\mu(\nu_1) - \log_\mu(\nu_2) \|_\mu^2,$$

for any $\nu_1, \nu_2 \in W_2(\Omega)$. 
An isometric representation of $W_2(\Omega)$

**Proposition**

The set of admissible functions $V_\mu(\Omega) = \log_\mu(W_2(\Omega)) \subset L^2_\mu(\Omega)$ can be characterized as the set of functions $v \in L^2_\mu(\Omega)$ such that

$$T := \text{id} + v$$

is $\mu$-a.e. increasing and that $T(x) \in \Omega$, for all $x \in \Omega$.

**Proposition**

The set $V_\mu(\Omega)$ is not a linear space.

Nevertheless, $V_\mu(\Omega)$ is closed and convex in $L^2_\mu(\Omega)$. 
Let $\mu \in W_{2}^{ac}(\Omega)$ be a reference measure.

Then, the geodesics in $W_{2}(\Omega)$ are exactly the image under $\exp_{\mu}$ of straight lines in $V_{\mu}(\Omega)$.

**Proposition**

Let $\gamma : [0, 1] \to W_{2}(\Omega)$ be a curve. Let

$$v_0 := \log_{\mu}(\gamma(0)) \text{ and } v_1 := \log_{\mu}(\gamma(1)).$$

Then, $\gamma$ is a geodesic if and only if

$$\gamma(t) = \exp_{\mu}((1 - t)v_0 + tv_1),$$

for all $t \in [0, 1]$. 
1. Motivations - Statistical analysis of histograms
2. Standard PCA in a Hilbert space
3. The Wasserstein space and its geometric properties
4. Geodesic PCA in the Wasserstein space
5. An algorithmic approach to GPCA
Fréchet mean and principal geodesics in $W_2(\Omega)$

Main ingredients to define analogs of PCA in a geodesic space:

- a notion of averaging / barycenter
- a notion of principal directions of variability around this barycenter
Fréchet mean and principal geodesics in $W_2(\Omega)$

**Standard PCA** can also be formulated as the problem of finding a sequence of nested affine subspaces minimizing the norms of the projection residuals of the data.

Let $x_1, \ldots, x_n$ be $n$ vectors in a separable Hilbert space $(H, \langle \cdot, \cdot \rangle, \| \cdot \|)$. Then, the first eigenvector $w_1$ is a solution of

$$\min_{v \in H, \|v\|=1} \frac{1}{n} \sum_{i=1}^{n} d^2(x_i, S_v),$$

where $S_v = \{\bar{x}_n + tv, \ t \in \mathbb{R}\}$ is the geodesic (straight line in $H$) passing through $\bar{x}_n$ with direction $v \in H$ and $d(x, S) = \inf_{x' \in S} \|x - x'\|$ denotes the distance from $x \in H$ to a given subset $S \subset H$. 
An example of standard PCA in $H = \mathbb{R}^2$

Data $x_1, \ldots, x_n$ in $\mathbb{R}^2$ and $\bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i$ their Euclidean mean

First and second principal “geodesic sets” of linear variation:

$$S_{w_1} = \{ \bar{x}_n + tw_1, \ t \in [-a, a] \} \quad \text{and} \quad S_{w_2} = \{ \bar{x}_n + tw_2, \ t \in [-a, a] \}$$
Fréchet mean and principal geodesics in $W_2(\Omega)$

Let $\nu_1, \ldots, \nu_n \in W_2(\Omega)$ a set of $n$ probability measures.

**Definition**

An empirical Fréchet mean of $\nu_1, \ldots, \nu_n \in W_2(\Omega)$ is defined as an element of

\[
\arg \min_{\nu \in W_2(\Omega)} \frac{1}{n} \sum_{i=1}^{n} d_{W_2}^2(\nu_i, \nu).
\]

**Proposition**

*There exists a unique empirical Fréchet mean, denoted by $\bar{\nu}_n$. Furthermore,*

\[
\bar{F}_n = \frac{1}{n} \sum_{i=1}^{n} F_i^-
\]

where $\bar{F}_n$ the cdf of $\bar{\nu}_n$ and $F_1, \ldots, F_n$ are the cdf of $\nu_1, \ldots, \nu_n$ respectively.
Fréchet mean of histograms

Data available: \( n = 1060 \) histograms \( f_1, \ldots, f_n \in L^2(\mathbb{R}) \).

Euclidean mean in \( L^2(\mathbb{R}) \)
Fréchet mean of histograms

Data available: \( n = 1060 \) histograms \( \nu_1, \ldots, \nu_n \in W_2(\Omega) \).

pdf of the Fréchet mean \( \bar{\nu}_n \) in \( W_2(\Omega) \) with \( \Omega = [1900; 2013] \).
Fréchet mean and principal geodesics in $W_2(\Omega)$

Bigot, Gouet, Klein & Lopez (2015)

Let $\mu \in W_2^{ac}(\Omega)$ be a reference measure.

**Definition**

Let $G \subset W_2(\Omega)$ and $\nu \in W_2(\Omega)$. We define the distance between $\nu$ and $G$ by $d_{W_2}(\nu, G) = \inf_{\pi \in G} d_{W_2}(\nu, \pi)$.

Let $\nu_1, \ldots, \nu_n \in W_2^{ac}(\Omega)$ a set of $n$ probability measures.

**Definition**

The first principal direction of variation in $W_2(\Omega)$ of $\nu_1, \ldots, \nu_n$ is a geodesic such

$$
\gamma^{(1)} := \arg \min \left\{ \frac{1}{n} \sum_{i=1}^{n} d_{W_2}^2(\nu_i, \gamma) \mid \gamma \text{ is a geodesic passing through } \bar{\nu}_n \right\}.
$$
1. Motivations - Statistical analysis of histograms

2. Standard PCA in a Hilbert space

3. The Wasserstein space and its geometric properties

4. Geodesic PCA in the Wasserstein space

5. An algorithmic approach to GPCA
Notation:

- for \( u_1 \in L^2_{\bar{\nu}_n}(\Omega) \), \( \text{span}(u_1) \) denotes the subspace spanned by \( u_1 \)
- \( \Pi_{\text{span}(u_1)} v \) is the projection of \( v \in L^2_{\bar{\nu}_n}(\Omega) \) onto \( \text{span}(u_1) \)
- \( \Pi_{\text{span}(u_1) \cap V_{\bar{\nu}_n}(\Omega)} v \) is the projection of \( v \) onto the closed convex set \( \text{span}(u_1) \cap V_{\bar{\nu}_n}(\Omega) \)

Recall that \( V_{\bar{\nu}_n}(\Omega) = \log_{\bar{\nu}_n}(W_2(\Omega)) \) is the closed and convex set of functions \( v \in L^2_{\bar{\nu}_n}(\Omega) \) such that

\[
T := \text{id} + v
\]

is \( \bar{\nu}_n \)-a.e. increasing and that \( T(x) = x + v(x) \in \Omega \), for all \( x \in \Omega \).
Proposition

Let $\nu_1, \ldots, \nu_n \in W_2(\Omega)^{ac}$ a set of $n$ probability measures. Let $u_1^*$ be a minimizer of the following convex-constrained PCA problem:

$$
\frac{1}{n} \sum_{i=1}^{n} \left\| \log_{\bar{\nu}_n}(\nu_i) - \Pi_{\text{span}(u_1) \cap V_{\bar{\nu}_n}(\Omega)} \log_{\bar{\nu}_n}(\nu_i) \right\|_{\bar{\nu}_n}^2
$$

over $u_1 \in L^2_{\bar{\nu}_n}(\Omega)$. Then,

$$
\gamma_1^{(1)} := \exp_{\bar{\nu}_n}(\text{span}(u_1^*) \cap V_{\bar{\nu}_n}(\Omega)).
$$

is the first principal source of geodesic variation in the data, that is

$$
\gamma_1^{(1)} = \arg \min \left\{ \frac{1}{n} \sum_{i=1}^{n} d_{W_2}(\nu_i, \gamma) \mid \gamma \text{ is a geodesic passing through } \bar{\nu}_n \right\}
$$
PCA on logarithms  Bigot, Gouet, Klein & Lopez (2015)

**Question** : why not applying PCA in $L^2_{\bar{\nu}_n}(\Omega)$ to the log-data?

**Proposition**

Let $\nu_1, \ldots, \nu_n \in W^{ac}_2(\Omega)$ a set of $n$ probability measures. If $\tilde{u}_1 \in L^2_{\bar{\nu}_n}(\Omega)$ is the eigenvector associated to the largest eigenvalue of the covariance operator

$$Kv = \frac{1}{n} \sum_{i=1}^n \langle \omega_i - \bar{\omega}_n, v \rangle \bar{\nu}_n(\omega_i - \bar{\omega}_n), \ v \in L^2_{\bar{\nu}_n}(\Omega),$$

with $\omega_i = \log_{\bar{\nu}_n} \nu_i$, and if $\Pi_{\text{span}(\tilde{u}_1)} \omega_i \in V_{\bar{\nu}_n}, i = 1, \ldots, n$, then $\tilde{u}_1 = u_1^*$. 

**Interpretation** : GPCA in $W_2(\Omega)$ may be simply obtained from the standard PCA on logarithms... but not always!!

**Remark** : PCA on logarithms simply amounts to PCA on quantile functions with respect to the usual $L^2$ metric.

Verde, Irpino & Balzanella (2015) for symbolic data analysis + R package
A simple example of convex-constrained PCA

Simulations: PCA of \( n = 4 \) points \( \omega_i \) in \( \mathbb{R}^2 \) restricted to the convex set of constraints

\[
V = \left\{ x = (x^{(1)}, x^{(2)}) \in \mathbb{R}^2 ; \; x^{(1)} \geq -1 \right\}
\]

The red line \( \text{span}(\tilde{u}_1) \) is standard PCA (not constrained)

**Left**: \( \Pi_{\text{span}(\tilde{u}_1)} \omega_i \in V \) for all \( 1 \leq i \leq n \) and thus \( u^*_1 = \tilde{u}_1 \)

**Right**: Convex PCA on \( V \) leads to \( \text{span}(u^*_1) \neq \text{span}(\tilde{u}_1) \)
PCA on logarithms for GPCA in $W_{2}^{ac}(\Omega)$

Data available: $n = 1060$ histograms $\nu_1, \ldots, \nu_n \in W_{2}^{ac}(\Omega)$.

First mode of geodesic variation in $W_{2}^{ac}(\Omega)$ via log-PCA

$$\tilde{\gamma}_{t}^{(1)} = \exp_{\bar{\nu}_n}(t\tilde{u}_1) \text{ for } -47.71 \leq t \leq 26.1, \text{ where } \tilde{u}_1 \in L_{\bar{\nu}_n}^2(\Omega).$$
PCA on logarithms for GPCA in $W_{2}^{ac}(\Omega)$

Data available: $n = 1060$ histograms $\nu_1, \ldots, \nu_n \in W_{2}^{ac}(\Omega)$.

Second mode of geodesic variation in $W_{2}^{ac}(\Omega)$ via log-PCA

$\tilde{\gamma}_t^{(2)} = \exp_{\tilde{\nu}_n}(t\tilde{u}_2)$ for $-7.23 \leq t \leq 19.06$, where $\tilde{u}_2 \in L_{\tilde{\nu}_n}^2(\Omega)$. 
Does PCA on logarithms lead to exact GPCA?

- Log-PCA amounts to compute the eigenvector $\tilde{u}_1 \in L^2_{\bar{\nu}_n}(\Omega)$ with largest eigenvalue from the PCA in the Hilbert space $L^2_{\bar{\nu}_n}(\Omega)$ of the log-data $\omega_i = \log_{\bar{\nu}_n} \nu_i$ for $1 \leq i \leq n$.

- If the following condition holds

  $$\Pi_{\text{span}(\tilde{u}_1)} \omega_i \in V_{\bar{\nu}_n} \text{ for all } i = 1, \ldots, n,$$

then $\tilde{\gamma}^{(1)} = \exp_{\bar{\nu}_n}(\text{span}(\tilde{u}_1) \cap V_{\bar{\nu}_n}(\Omega))$ is the first principal source of geodesic variation in the data, that is

$$\gamma_*^{(1)} = \arg \min \left\{ \frac{1}{n} \sum_{i=1}^{n} d_{W_2}^2(\nu_i, \gamma) \mid \gamma \text{ is a geodesic passing through } \bar{\nu}_n \right\}$$
Does PCA on logarithms lead to exact GPCA?

- Log-PCA amounts to compute the eigenvector $\tilde{u}_1 \in L^2_{\bar{\nu}_n}(\Omega)$ with largest eigenvalue from the PCA in the Hilbert space $L^2_{\bar{\nu}_n}(\Omega)$ of the log-data $\omega_i = \log_{\bar{\nu}_n} \nu_i$ for $1 \leq i \leq n$.

- If, for all $i = 1, \ldots, n$, the following conditions hold
  
  \begin{align*}
  (a) & \quad x \mapsto \tilde{T}_i(x) \text{ is } \bar{\nu}_n\text{-a.e. increasing}, \\
  (b) & \quad \tilde{T}_i(x) \in \Omega,
  \end{align*}

  where

  \[
  \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}_n} \tilde{u}_1(x), \quad x \in \Omega,
  \]

  then $\tilde{\gamma}^{(1)} = \exp_{\bar{\nu}_n}(\text{span}(\tilde{u}_1) \cap V_{\bar{\nu}_n}(\Omega))$ is the first principal source of geodesic variation in the data.
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]

\[ i = 1 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp \tilde{\nu}_n (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]

\[ i = 21 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{i} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ i = 41 \]
Does PCA on logarithms lead to exact GPCA?

\[
\tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \\
\tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n
\]

\( i = 61 \)
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{\nu}_n} (\tilde{t}_i \tilde{u}_1) \quad \text{with} \quad \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ i = 81 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ i = 100 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{v}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{v}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{v}_n \]

\[ i = 300 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\bar{\nu}_n} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \]

\( i = 500 \)
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{\nu}} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \]

\[ i = 700 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]

\[ i = 900 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{\nu}_n}(\tilde{t}_i\tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ i = 1000 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]

\[ i = 1010 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{t_i} = \exp_{\bar{\nu}_n} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \]

\[ i = 1020 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ \tilde{\gamma}_{i}^{(1)} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ i = 1030 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ i = 1040 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]

\[ i = 1050 \]
Does PCA on logarithms lead to exact GPCA?

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp \tilde{\nu}_n (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]

\[ i = 1060 \]
An algorithmic approach for exact GPCA Work in progress...

**Exact GPCA** is the convex-constrained PCA problem:

\[
    u_1^* = \arg \min_{u \in L^2_{\nu_n}(\Omega)} \frac{1}{n} \sum_{i=1}^{n} \| \omega_i - \Pi_{\text{span}(u) \cap V_{\nu_n}(\Omega)} \omega_i \|_{\nu_n}^2 ; \text{ with } \|u\|_{\nu_n} = 1
\]  

**Proposition**

*Introducing the characteristic function of the convex set* \(V_{\nu_n}(\Omega)\) *as:*

\[
    \chi_{V_{\nu_n}(\Omega)}(v) = \begin{cases} 
    0 & \text{if } v \in V_{\nu_n}(\Omega) \\
    +\infty & \text{otherwise}
\end{cases}
\]

*the problem (1) is equivalent to*

\[
    u_1^* = \arg \min_{v \in L^2_{\nu_n}(\Omega)} \min_{t_0 \in [-1;1]} \left\{ \frac{1}{n} \sum_{i=1}^{n} \min_{t_i \in [-1;1]} \| \omega_i - (t_0 + t_i)v \|_{\nu_n}^2 ight. \\
    \left. + \chi_{V_{\nu_n}(\Omega)}((t_0 - 1)v) + \chi_{V_{\nu_n}(\Omega)}((t_0 + 1)v) \right\}
\]
In a discrete setting, for a given $t_0 \in [-1; 1]$, the problem of exact GPCA can be formulated as an optimisation problem of the form:

$$\min_{v \in \mathbb{R}^N} \min_{t \in \mathbb{R}^n} \left( \sum_{i=1}^{n} \sum_{j=1}^{N} \bar{f}_n(x_j) \left( w_i^j - (t_0 + t_i)v_j \right)^2 \right)$$

$$F(v,t) + \chi_D(v) + \chi_E(Kv) + \chi_{B^1_1}(t) \cdot G(v,t)$$

where $B^1_1$ is the $L^\infty$ ball of $\mathbb{R}^n$ with radius 1 dealing with the constraint $t_i \in [-1; 1]$.

In the numerical experiments, we took $t_0 = 0$ (as in Seguy and Cuturi (2015) for GPCA in $W_2(\mathbb{R}^d)$), but other values should be tried to find the best one!
An algorithmic approach for exact GPCA  Work in progress...

In a discrete setting, the problem of exact GPCA can be formulated as an optimisation problem of the form:

$$\min_{v \in \mathbb{R}^N} \min_{t \in \mathbb{R}^n} J(v, t) := F(v, t) + G(v, t)$$

**Remark:** $F$ is differentiable but non-convex in $(v, t)$ and $G$ is non-smooth and convex.

**Forward-Backward algorithm (numerical results in progress...)**

Denoting $X = (v, t) \in \mathbb{R}^{N+n}$, taking $\tau > 0$ and $X^{(0)} \in \mathbb{R}^{N+n}$, it reads:

$$X^{(\ell+1)} = \text{Prox}_{\tau G}(X^{(\ell)} - \tau \nabla F(X^{(\ell)})),$$

where

$$\text{Prox}_{\tau G}(\tilde{X}) = \arg \min_{X \in \mathbb{R}^{N+n}} \frac{1}{2\tau} \|X - \tilde{X}\|^2 + G(X),$$

with $\| \cdot \|$ the Euclidian norm.
Comparison between log-PCA and exact GPCA

Log-PCA : $x \mapsto \tilde{u}_1(x)$ versus Exact GPCA : $x \mapsto u^*_1(x)$
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}} \tilde{u}_1(x) \]

\[ T_i^*(x) = x + i^* u_1^*(x) \]

\[ \tilde{\gamma}^{(1)}_{i_i} = \exp_{\bar{\nu}}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}} \]

\[ \gamma^{(1)}_{i^*_i} = \exp_{\bar{\nu}}(i^*_i u_1^*(x)) \]

\( i = 500 \)
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]
\[ T^*_i(x) = x + i^*_i u^*_1(x) \]
\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\nu_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]
\[ \gamma^{(1)}_{t^*_i} = \exp_{\nu_n}(t^*_i u^*_1(x)) \]

\( i = 700 \)
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{V}_n} \tilde{u}_1(x) \]
\[ T_i^*(x) = x + t_i^* u_1^*(x) \]
\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{V}_n}(\tilde{t}_i \tilde{u}_1) \quad \text{with} \quad \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{V}_n} \]
\[ \gamma_{t_i^*}^{(1)} = \exp_{V_n}(t_i^* u_1^*(x)) \]

\[ i = 900 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}_n} \tilde{u}_1(x) \]
\[ T_i^*(x) = x + t_i^* u_1^*(x) \]
\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\bar{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}_n} \]
\[ \gamma_{t_i^*}^{(1)} = \exp_{\bar{\nu}_n}(t_i^* u_1^*(x)) \]

\[ i = 1000 \]
**Exact GPCA is different from PCA on logarithms!**

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \tilde{u}_1(x) \]

\[ T_i^*(x) = x + \tilde{t}_i^* \tilde{u}_1^*(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{\nu}}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \]

\[ \gamma_{t_i}^{(1)} = \exp_{\nu_n}(t_i^* u_1^*(x)) \]

\[ i = 1010 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu} \tilde{u}_1(x) \]

\[ T_i^*(x) = x + t_i^* u_1^*(x) \]

\[ \tilde{\gamma}^{(1)}_{t_i} = \exp_{\bar{\nu}}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu} \]

\[ \gamma^{(1)}_{t_i^*} = \exp_{\bar{\nu}}(t_i^* u_1^*(x)) \]

\[ i = 1020 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}} \tilde{u}_1(x) \]

\[ T_i^*(x) = x + t_i^* u_1^*(x) \]

\[ \tilde{\gamma}_i^{(1)} = \exp_{\bar{\nu}}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}} \]

\[ \gamma_{t_i^*}^{(1)} = \exp_{\bar{\nu}}(t_i^* u_1^*(x)) \]

\[ i = 1030 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ T^*_i(x) = x + t^*_i u^*_1(x) \]

\[ \tilde{\gamma}^{(1)}_{\tilde{t}_i} = \exp_{\tilde{\nu}_n} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ \gamma^{(1)}_{t^*_i} = \exp_{\nu_n} (t^*_i u^*_1(x)) \]

\( i = 1040 \)
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \tilde{u}_1(x) \]

\[ T^*_i(x) = x + i^*_i u^*_1(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\bar{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \]

\[ \gamma_{t^*_i}^{(1)} = \exp_{\bar{\nu}_n}(t^*_i u^*_1(x)) \]

\[ i = 1050 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{u}_1(x) \]
\[ T_i^*(x) = x + t_i^* u_1^*(x) \]
\[ \tilde{\gamma}_{t_i}^{(1)} = \exp_{\tilde{\nu}_n} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]
\[ \gamma_{t_i^*}^{(1)} = \exp_{\nu_n} (t_i^* u_1^*(x)) \]

\( i = 1060 \)
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \tilde{u}_1(x) \]

\[ T_i^*(x) = x + t_i^* u_1^*(x) \]

\[ \tilde{\gamma}_{t_i}^{(1)} = \exp_{\tilde{\nu}} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \]

\[ \gamma_{t_i^*}^{(1)} = \exp_{\nu} (t_i^* u_1^*(x)) \]

\[ i = 300 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \tilde{u}_1 (x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\bar{\nu}_n} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \bar{\nu}_n \]

\[ i = 100 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}_n} \tilde{u}_1(x) \]

\[ T^*_i(x) = x + i_i^* u_1^*(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\bar{\nu}_n} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\bar{\nu}_n} \]

\[ \gamma_{t^*_i}^{(1)} = \exp_{\bar{\nu}_n} (t^* u_1^*(x)) \]

\[ i = 81 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \tilde{u}_1(x) \]

\[ T^*_i(x) = x + t^*_i u^*_1(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}_n} \]

\[ \gamma_{t^*_i}^{(1)} = \exp_{\nu_n}(t^*_i u^*_1(x)) \]

\[ i = 61 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{u}_1(x) \]
\[ T_i^*(x) = x + t_i^* u_1^*(x) \]

\[ \tilde{\gamma}_{\tilde{t}_i}^{(1)} = \exp_{\tilde{\nu}_n} (\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]
\[ \gamma_{t_i^*}^{(1)} = \exp_{\nu_n} (t_i^* u_1^*(x)) \]

\[ i = 41 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \tilde{u}_1(x) \]
\[ T^*_i(x) = x + t^*_i u^*_1(x) \]
\[ \tilde{\gamma}^{(1)}_{t_i} = \exp_{\tilde{\nu}_n}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle \tilde{\nu}_n \]
\[ \gamma^{(1)}_{t^*_i} = \exp_{\tilde{\nu}_n}(t^*_i u^*_1(x)) \]

\[ i = 21 \]
Exact GPCA is different from PCA on logarithms!

\[ \tilde{T}_i(x) = x + \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \tilde{u}_1(x) \]
\[ T_i^*(x) = x + t_i^* u_1^*(x) \]
\[ \tilde{\gamma}_{t_i}^{(1)} = \exp_{\tilde{\nu}}(\tilde{t}_i \tilde{u}_1) \text{ with } \tilde{t}_i = \langle \omega_i, \tilde{u}_1 \rangle_{\tilde{\nu}} \]
\[ \gamma_{t_i}^{(1)} = \exp_{\nu}(t_i^* u_1^*(x)) \]
Perspectives

- Extend the algorithm for the computation of $k \geq 2$ principal geodesic directions of variation.

- Regularized version of GPCA to have smoother maps $T_i^*$. 

- Extension to histograms supported on $\mathbb{R}^d$ for $d \geq 2$. 