Functional Principal Component Analysis

Advanced Statistics

Daniel López Montero

MSc Mathematics and Applications
Universidad Autónoma de Madrid

May 12, 2025

Index

- PCA
- PCA for Functional Data (Berrendero et al. 2011)
- Fourier PCA[1]
- Kernel Functional PCA[1]

[1] Contributions

Historical context

$$\mathcal{X} \longrightarrow \mathcal{Z}$$

PCA
(Karl Pearson, 1901)

$$\mathbb{R}^p \longrightarrow \mathbb{R}^k$$
$$\mathbf{x} \longmapsto \mathbf{U}^\top \mathbf{x}$$

Functional PCA
(Karhunen & Loève, 1947)

$$L^2(\mathbb{R}) \longrightarrow \mathbb{R}^N$$
$$X(t) \longmapsto \left\{\int X(t) e_j(t) dt\right\}_{j=1}^N$$

PCA for Functional Data
(Berrendero et. al, 2011)

$$L^2(\mathbb{R}, \mathbb{R}^p) \longrightarrow L^2(\mathbb{R}, \mathbb{R}^k)$$
$$\mathbf{x}(t) \longmapsto \mathbf{U}(t)^\top \mathbf{x}(t)$$

Problem Formulation

Every compression method has two parts, the projection into a lower-dimensional space and the reconstruction:

$$\Phi : \mathcal{X} \to \mathcal{Z} \qquad \text{(Transform)}$$

$$\Phi^* : \mathcal{Z} \to \mathcal{X} \qquad \text{(Inverse Transform)}$$

where $\mathcal{Z}$ is the *principal component space* or *latent space*. Let us denote by P the projection onto the subspace spanned by the principal components, i.e.,

$$P : \mathcal{X} \to \mathcal{X} \qquad P = \Phi^* \circ \Phi$$

Let X be a random variable on the space $\mathcal{X}$.

Maximize Variance

$$\max_{\Phi} \mathsf{Var}\left(\Phi(X)\right)$$

Minimize square error

$$\min_{\Phi} \mathbb{E}\|X - PX\|_{\mathcal{X}}^2$$

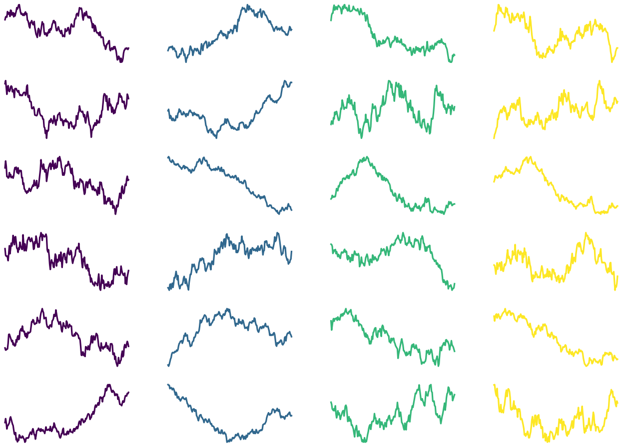
Problem Formulation

	Optimal value
PCA (Karl Pearson, 1901)	$\mathbf{\Sigma}\mathbf{u}_i = \lambda_i \mathbf{u}_i$
Functional PCA (Karhunen & Loève, 1947)	$\int K(t,s)e_i(t)ds = \lambda_i e_i(s)$
PCA for Functional Data (Berrendero et. al, 2011)	$\mathbf{\Sigma}(t)\mathbf{u}_i(t) = \lambda_i(t)\mathbf{u}_i(t)$

Correlated $n-$dimensional Brownian Motion

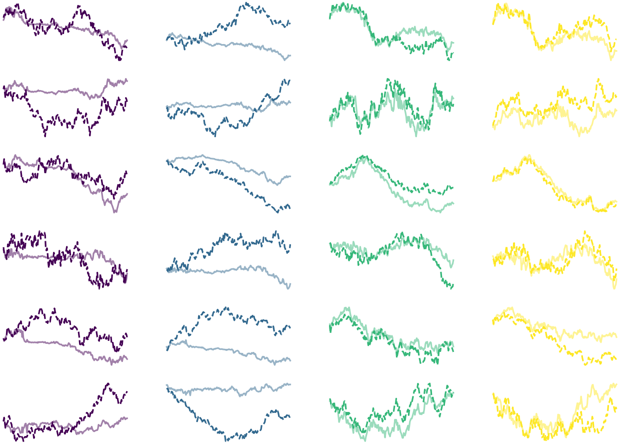
We generate several $4-$dimensional Brownian samples

$$X_{t+1} = X_t + \sqrt{\Delta t}\mathcal{N}(0, \Sigma)$$



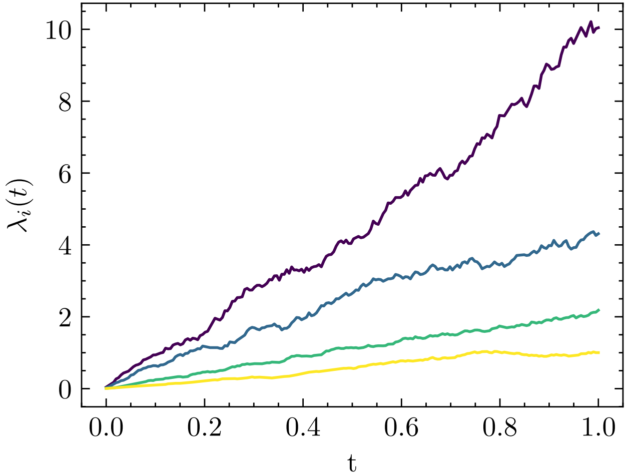
Correlated $n-$dimensional Brownian Motion

We test the PCA for Functional Data (Berrendero et al. 2011) for $k = 1$ and compare the reconstructed curves



Correlated $n-$dimensional Brownian Motion

We test the PCA for Functional Data (Berrendero et al. 2011) for $k = 1$ and compare the reconstructed curves



Anomaly Detection using PCA-like Methods

Definition

We define a sample $x \in \mathcal{X}$ as an ***anomaly*** if it *deviates significantly* from its projection, i.e.,

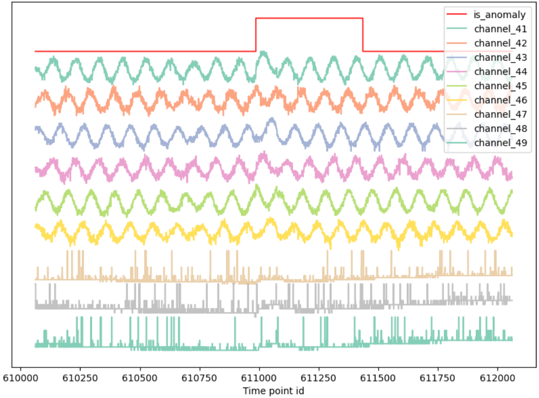
$$\|x - Px\|_{\mathcal{X}}^2 \geq \mu$$

where μ is a threshold that can be estimated or predefined.

If $X \sim \mathcal{N}(0, \Sigma)$, then $\|X - PX\|_2^2 \sim \chi_{n-k}^2$ where k is the dimension of the projection.

Anomaly Detection for Satellite Telemetry

- Continuous stream of data.
- Labelled anomalies
- 50 channels and 15 telecommands with over 50M timesteps

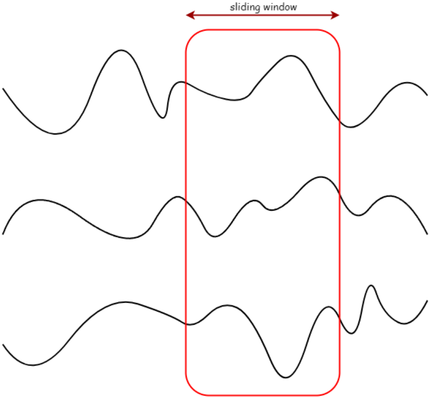


Anomaly Detection for Satellite Telemetry

Conventional FDA techniques are not compatible with continuous streams of data. We want to develop a method that is compatible with time shifts, i.e.,

$$\tau_s \circ P(X_t) = P \circ \tau_s(X_t) \qquad (\tau_s\text{-equivariant})$$

where $P : \mathcal{X} \to \mathcal{X}$ is the projection and $\tau_s(X_t) := X_{s-t}$.



Fourier PCA

The Fourier theory comes to the rescue.

$$\begin{array}{ccccc}
X(t) & & & & \widetilde{X}(t) \\
\Big\downarrow \mathcal{F} & & & & \Big\uparrow \mathcal{F}^{-1} \\
\widehat{X}(\xi) & \xrightarrow[\Phi]{\qquad\qquad} & \widehat{Z}(\xi) & \xrightarrow[\Phi^*]{\qquad\qquad} & \widetilde{\widehat{X}}(\xi)
\end{array}$$

Here, $\widetilde{X}$ and $\widetilde{\widehat{X}}$ denotes the reconstruction of X and $\widehat{X}$, respectively.

Fourier PCA

The Fourier theory comes to the rescue.

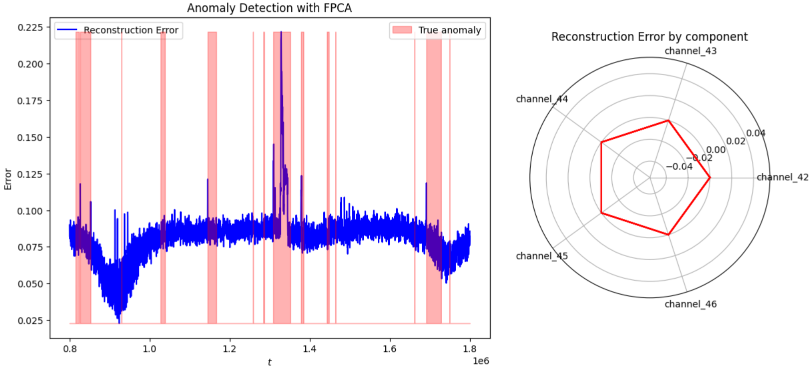
$$
\begin{tikzcd}
X(t) \arrow[r, "\Psi"'] \arrow[d, "\mathcal{F}"'] & Z(t) \arrow[r, "\Psi^*"'] & \widetilde{X}(t) \\
\widehat{X}(\xi) \arrow[r, "\Phi"'] & \widehat{Z}(\xi) \arrow[r, "\Phi^*"'] & \widetilde{\widehat{X}}(\xi) \arrow[u, "\mathcal{F}^{-1}"]
\end{tikzcd}
$$

If we define the projection $Q = \Psi^* \circ \Psi$. Then, it satisfies the τ_s-equivariant property and due to the Plancherel Theorem[2]:

$$\mathbb{E}\|\widehat{X}(\xi) - \widetilde{\widehat{X}}(\xi)\|_{L^2}^2 \approx \mathbb{E}\|X(t) - \widetilde{X}(t)\|_{L^2}^2$$

[2] Assuming that $\widetilde{\widehat{X}}(\xi) \approx \widehat{\widetilde{X}}(\xi)$.

Anomaly Detection for Satellite Telemetry



Kernel PCA

Key Ideas:

1. Maps to a higher dimension space $\varphi : \mathcal{X} \to \mathcal{H}$ where it performs PCA.
2. Maximizes the variance: $\max_{f \in \mathcal{H}} \mathsf{Var}\langle \varphi(x), f \rangle_{\mathcal{H}}$.
3. Non-linear and more expressive than PCA.
4. We need RKHS to prove there is an optimal solution (Representer Theorem).
5. Uses the RKHS generated by $K(x, x') = \langle \varphi(x), \varphi(x') \rangle_{\mathcal{H}}$.
6. φ does not need to be explicitly known.

Kernel PCA

There are several well-known reproducing kernels:

$$\begin{aligned}
\textbf{Exponential kernel}: &\quad K(x,x') := \exp(\langle x, x'\rangle_2),\\
\textbf{Gaussian RBF kernel}: &\quad K(x,x') := \exp(-\sigma^{-2}\|x - x'\|_2^2),\\
\textbf{Polynomial kernel}: &\quad K(x,x') := (\langle x, x'\rangle_2)^p,
\end{aligned}$$

Instead of working with $\mathcal{X} = \mathbb{R}^d$, we want to extend these definitions to spaces of functions, e.g., $L^2(\mathbb{R}), C(\mathbb{R})$, etc.

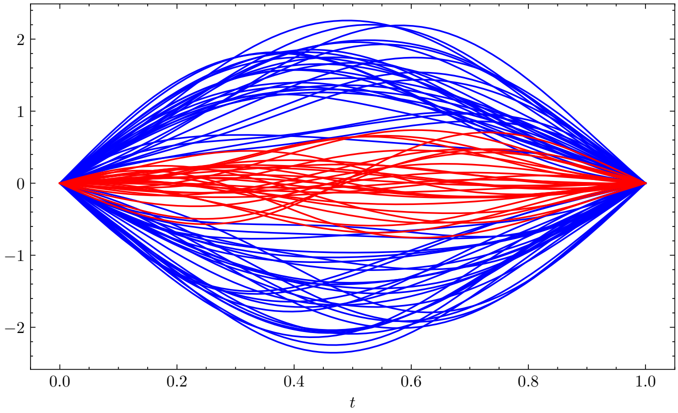
Kernel Functional PCA

We come up with these kernels and have proven that they are, indeed, kernels:

$$\begin{aligned}
\textbf{Exponential kernel}: \quad & K(x, x') := \exp(\langle x, x' \rangle_{L^2}), \\
\textbf{Gaussian RBF kernel}: \quad & K(x, x') := \exp(-\sigma^{-2} \|x - x'\|_{L^2}^2), \\
\textbf{Polynomial kernel}: \quad & K(x, x') := (\langle x, x' \rangle_{L^2})^p,
\end{aligned}$$

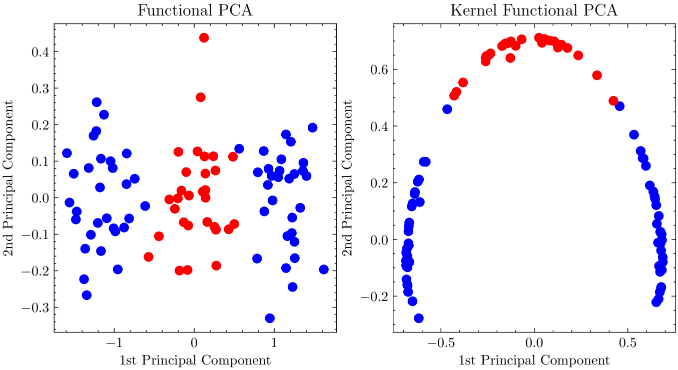
Kernel Functional PCA: Experiment

In this example, we have created two Gaussian processes, and it is evident that the two categories of points (red and blue) are not linearly separable.



FPCA vs Kernel FPCA

However, we can see that using the Kernel PCA with the Gaussian kernel, both categories can be separated.



Universality of Kernels

Definition

We say that a Kernel K is universal when $\mathcal{H}_K$ is dense in $C(\mathcal{X})$ or $L^p(\mathcal{X}, \mu)$.

- This property is useful because it affects the ability to learn and provide *arbitrarily accurate* decision functions for *all* distributions.
- For example, the Gaussian kernel and exponential kernel in $\mathbb{R}^d$ have been proven to be universal (Christmann et al. 2008).
- There is a sufficient condition that gives (Carmeli et al. 2008) and (Christmann et al. 2008) to prove that is universal. The function $S_x : L^p(\mathcal{X}, \mu) \to \mathcal{H}$ defined by

$$S_x g(x) := \int K(x, x')g(x')d\mu(x')$$

should be injective. In this case $\mathcal{H}$ is dense in $L^p(\mathcal{X}, \mu)$.

Conclusions

- PCA
- PCA for Functional Data
- Fourier PCA
- Kernel Functional PCA

References

Berrendero, J. R. et al. (Sept. 2011). "Principal components for multivariate functional data". In: *Computational Statistics & Data Analysis* 55.9, pp. 2619–2634. ISSN: 0167-9473. DOI: 10.1016/j.csda.2011.03.011. URL: https://www.sciencedirect.com/science/article/pii/S0167947311001022 (visited on 04/10/2025).

Carmeli, C. et al. (July 2008). *Vector valued reproducing kernel Hilbert spaces and universality*. arXiv:0807.1659 [math]. DOI: 10.48550/arXiv.0807.1659. URL: http://arxiv.org/abs/0807.1659 (visited on 05/11/2025).

Christmann, Andreas et al. (2008). *Support Vector Machines*. en. Information Science and Statistics. New York, NY: Springer New York. ISBN: 978-0-387-77241-7 978-0-387-77242-4. DOI: 10.1007/978-0-387-77242-4. URL: https://link.springer.com/10.1007/978-0-387-77242-4 (visited on 05/09/2025).