Regularized Principal Component Analysis

J.O. Ramsay and B.W. Silverman

Presented by Long Yu

1. **Notation Introduction;**

2. **Motivation;**

3. **Estimating Smoothed Principal Components;**

4. **Finding the Regularized PCA in Practice;**
(a)Periodic Case; (b)Nonperiodic Case;

5. **Choosing the Smoothing Parameter by Cross-Validation;**

6. **An Example: The Pinch Force Data;**

7. **Alternative Way: Smoothing the Data Rather Than PCA;**

1. Notation Introduction:

(1)Multivariate Data: $\mathbf{x}$ and $\mathbf{y}$ are two vectors of length p.

- Euclidean Inner Product:

$$< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{p} x_i y_i$$

$$\begin{cases} symmetry: & < \mathbf{x}, \mathbf{y} >=< \mathbf{y}, \mathbf{x} > \\ positivity: & < \mathbf{x}, \mathbf{x} > \geq 0 \text{for all } \mathbf{x} \text{ with } < \mathbf{x}, \mathbf{x} >= 0 \\ & \text{iff } \mathbf{x} = 0 \\ bilinearity: & < a\mathbf{x} + b\mathbf{y}, \mathbf{z} >= a < \mathbf{x}, \mathbf{z} > +b < \mathbf{y}, \mathbf{z} > \\ & \text{for any real numbers } a \text{ and } b \end{cases}$$

Generalized Inner Product:

$$< \mathbf{x}, \mathbf{y} >= \mathbf{x}^{'} \mathbf{B} \mathbf{y}$$

where $\mathbf{B}$ is positive-definite symmetric matrix $\mathbf{B}$.

• PCA and Eigenanalysis:
$\mathbf{X}_{N\times p} = (\mathbf{x}_1, ..., \mathbf{x}_N)'$ where $\mathbf{x}_i$ is $p \times 1$ vector, $i = 1, ..., N$. The PCA problem is to find $\xi_{p\times 1}$:

$max_{<\xi,\xi>=1} < \xi, \mathbf{V}\xi >$ where $V = N^{-1}\mathbf{X}'\mathbf{X}$

$$\Longleftrightarrow$$

$\mathbf{V}\xi = \rho\xi$ subject to $< \xi, \xi >= 1$.

(2)Functional Data: $x(t)$ and $y(t)$ are two functions on $[0, T]$.
• Inner Product: $< x, y >= \int_0^T x(t)y(t)dt$

• Functional PCA:

Suppose $x_i, i = 1, ..., N$ are curves, define

$$\mathbf{V}\xi(s) = \int_0^T v(s,t)\xi(t)dt$$

where $v(s,t) = N^{-1}\sum_{i=1}^{N} x_i(s)x_i(t)$ and

$$< \xi, \mathbf{V}\xi >= \int_0^T \xi(s) \int_0^T v(s,t)\xi(t)dtds$$

.

The PCA problem is also to find $\xi(s)$

$max_{<\xi,\xi>=1} < \xi, \mathbf{V}\xi > \Longleftrightarrow \mathbf{V}\xi = \rho\xi$, $< \xi, \xi >= 1$

2. Motivation:

The Pinch Force Data.

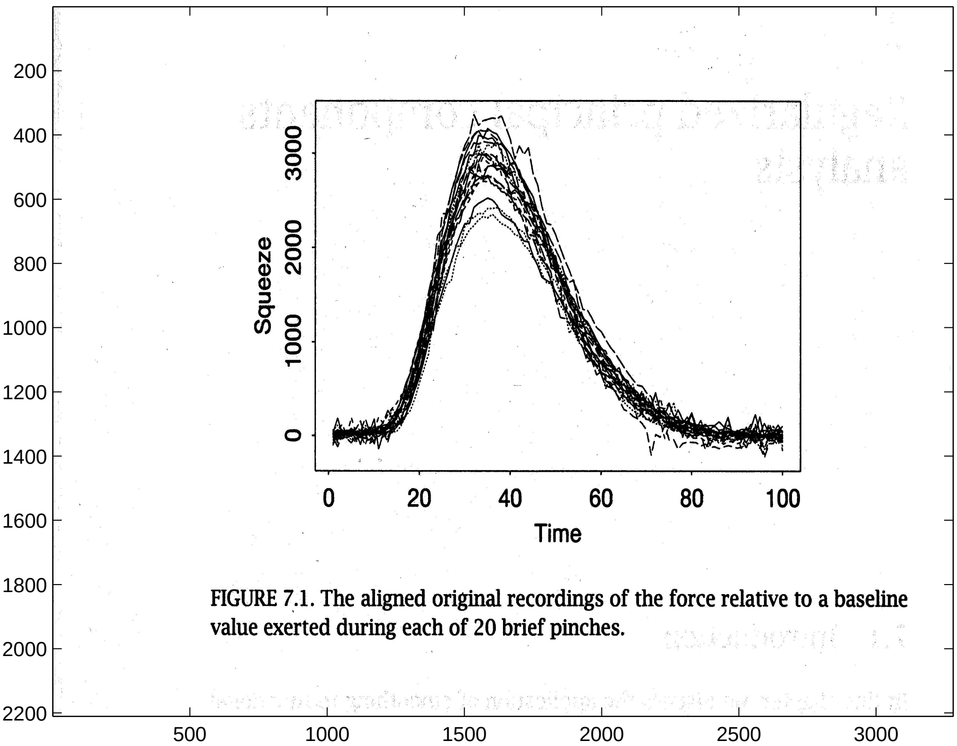

FIGURE 7.1. The aligned original recordings of the force relative to a baseline value exerted during each of 20 brief pinches.

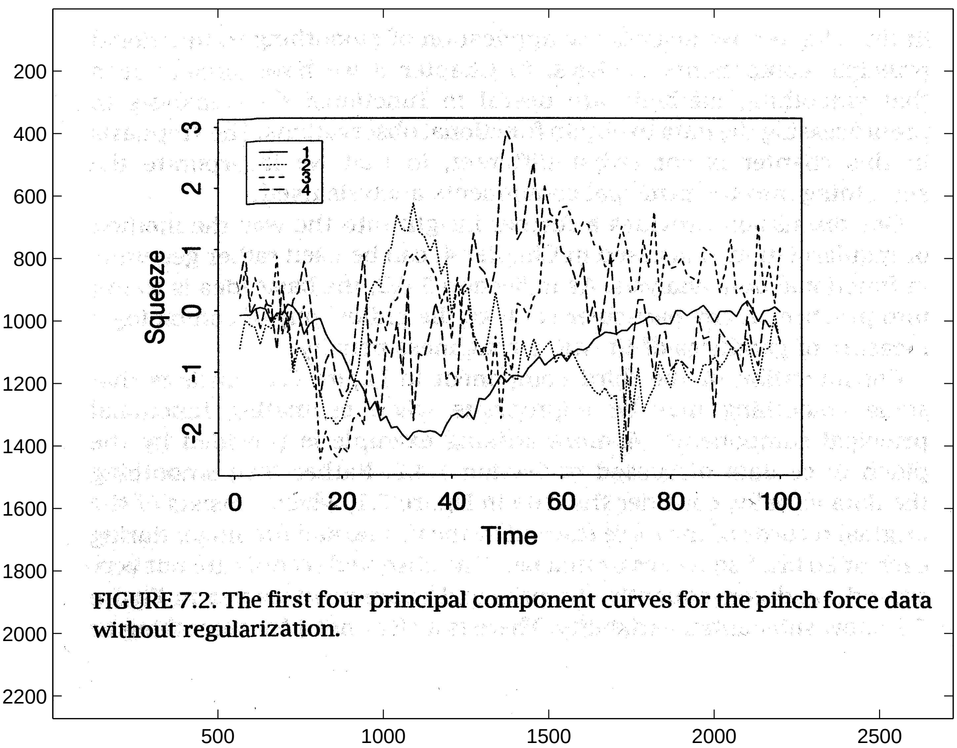

FIGURE 7.2. The first four principal component curves for the pinch force data without regularization.

Three Approaches:

(a)Incorporate the smoothing into the principal component analysis itself without smoothing the raw data.

(b)Smooth raw data and conduct PCA on the smoothed data.

(c)Combine the above two approaches.

Focus on the approach (a).

3. Estimating Smoothed Principal Components:

• Roughness Penalty:

$$PEN_2(\xi) = \|D^2\xi\|^2$$

where ξ is any principal component function and $D^2\xi$ is the second derivative of ξ.

• Regularity Conditions for S, the space of functions ξ:

(a)ξ has continuous derivative and square-integrable second derivative.

(b)relevant boundary conditions: periodicity of ξ and $D\xi$, and $\xi(0) = D\xi(0) = 0$. Additionally, (c)ξ has square-integrable derivatives up to degree four.

(d)ξ satisfies one of the following two conditions: (i)(natural) $D^2\xi$ and $D^3\xi$ are zero at the ends of the interval T. (ii)(periodic) $D^2\xi$ and $D^3\xi$ satisfy periodic boundary conditions on T.

A function is called *very smooth* on T if it satisfies either one of these two conditions.

• Alternative form of $PEN_2(\xi)$ if ξ is a very smooth:

$$PEN_2(\xi) = \|D^2\xi\|^2 =< \xi, D^4\xi >$$

Because, if x is any function in S and ξ is very smooth,

$$
\begin{aligned}
&< D^2x, D^2\xi > \\
&= \int_T D^2x(s)D^2\xi(s)ds \\
&= Dx(T)D^2\xi(T) - Dx(0)D^2\xi(0) - \int_T Dx(s)D^3\xi(s)ds \\
&= -\int_T Dx(s)D^3\xi(s)ds \\
&= x(T)D^3\xi(T) - x(0)D^3\xi(0) + \int_T x(s)D^4\xi(s)ds \\
&=< x, D^4\xi >
\end{aligned}
$$

So, $\|D^2\xi^2\|^2 =< \xi, D^4\xi >$.

- Penalized Sample Variance:

$$
\begin{aligned}
PCAPSV &= \frac{<\xi, \mathbf{V}\xi>}{\|\xi\|^2 + \lambda \times PEN_2(\xi)} \\
&= \frac{<\xi, \mathbf{V}\xi>}{<\xi, (I + \lambda D^4)\xi>}
\end{aligned}
$$

Recall Fact: If $\mathbf{A}$ is symmetric and $\mathbf{B}$ is positive-definite symmetric, then
maximizing the ratio $\frac{<\xi,\mathbf{A}\xi>}{<\xi,\mathbf{B}\xi>}$
$\Longleftrightarrow max < \xi, \mathbf{A}\xi >$ subject to $< \xi, \mathbf{B}\xi >= 1$
$\Longleftrightarrow$ Solve $\mathbf{A}\xi = \rho\mathbf{B}\xi$ subject to $< \xi, \mathbf{B}\xi >= 1$

Remark: The jth generalized eigenvector ξ_j solves problem:
$max < \xi, \mathbf{A}\xi >$ subject to $< \xi, \mathbf{B}\xi >= 1$ and $< \xi, \mathbf{B}\xi_i >= 0$ for $i < j$.
So maximizing $PCAPSV \Longleftrightarrow$ finding leading solution of

$$\mathbf{V}\xi = \rho(I + \lambda D^4)\xi \qquad (E1)$$

subject to $\|\xi\|^2 = 1$ and the modified form of orthogonality is $< \xi_j, \xi_k > +\lambda < D^2\xi_j, D^2\xi_k >= 0$ for $k = 1, ..., j - 1$.

4. Finding the Regularized PCA in Practice:

(a)The Periodic Case:

Let $\{\phi_j\}_{j=1}^k$ be the series of Fourier functions, for any x_i

$x_i(s) = \sum_{j=1}^k c_{ji}\phi_j(s) = c^{'}\phi(s)$ where

$c_{ji} =< x_i, \phi_j >$, $j = 1, ..., k$, $c_i = (c_{1i}, ..., c_{ki})^{'}$, and $\phi(s) = (\phi_1(s), ..., \phi_k(s))^{'}$,

for any principal component curve ξ,

$\xi(s) = y^{'}\phi(s)$ where $y = (y_1, ..., y_k)^{'}$.

Let $\mathbf{W}$ be the covariance matrix of vectors c_i and let $\mathbf{S}_{k\times k}$ be the diagonal matrix with entries $(S)_{jj} = (1 + \lambda\omega_j^4)^{-\frac{1}{2}}, j = 1, ..., k$ where $\omega_{2j-1} = \omega_{2j} = 2\pi j$.

$(E1)$ can be written as:

$$\mathbf{W}y = \rho \mathbf{S}^{-2} y$$

Because: $D^4 \phi_j = \omega_j^4 \phi_j$. Let
$\mathbf{C}_{N\times k} = (c_1, ..., c_N)^{'}$, $(x_1(s), ..., x_N(s))^{'} = \mathbf{C}\phi(s)$

$$\begin{aligned} v(s,t) &= N^{-1} \textstyle\sum_{i=1}^{N} x_i(s) x_i(t) = N^{-1} \phi(s)^{'} \mathbf{C}^{'} \mathbf{C} \phi(t) \\ &= \phi(s)^{'} \mathbf{W} \phi(t) \end{aligned}$$

$$\begin{aligned} \mathbf{V}\xi(s) &= \textstyle\int v(s,t)\xi(t) dt = \int \phi(s)^{'} \mathbf{W} \phi(t) \phi(t)^{'} y dt \\ &= \phi(s)^{'} \mathbf{W} \textstyle\int \phi(t)\phi(t)^{'} dt y = \phi(s)^{'} \mathbf{W} y \end{aligned}$$

$$\begin{aligned} \rho(I + \lambda D^4)\xi(s) &= \rho(\phi(s)^{'} y + \lambda (D^4 \phi(s))^{'} y) \\ &= \rho\phi(s)^{'} (y + \lambda \phi(s)^{'} (\omega_1^4 y_1, ..., \omega_k^4 y_k)^{'}) \\ &= \rho \phi(s)^{'} \mathbf{S}^{-2} y \end{aligned}$$

Since the equation holds for all s,

$$\mathbf{W}y = \rho\mathbf{S}^{-2}y$$

Remark: $(\mathbf{SWS})(\mathbf{S}^{-1}y) = \rho(\mathbf{S}^{-1}y)$ if u is an eigenvector of $\mathbf{SWS}$, then $y = \frac{\mathbf{S}u}{||\mathbf{S}u||}$

Here is the algorithm:

(i)Compute the coefficients c_i for the expansion of each sample function x_i in terms of basis ϕ.

(ii)Operate on these coefficients by the smoothing operator $\mathbf{S}$.

(iii)Carry out a standard PCA on the resulting smoothed coefficient vectors $\mathbf{S}c_i$.

(iv)Apply the smoothing operator $\mathbf{S}$ to the resulting eigenvectors u, and renormalize so that the resulting vectors y have unit norm.

(vi)Compute the principal component function ξ from $\xi(s) = y^{'}\phi(s)$.

(b)The nonperiodic case:

For any suitable basis $\{\phi_j\}_{j=1}^k$,
$x_i(s) = c_i^{'}\phi(s)$ and $\xi(s) = y^{'}\phi(s)$. Define

$$\mathbf{W} = N^{-1}\sum_{i=1}^{N} c_i c_i^{'}$$

$\mathbf{J}_{k\times k} = \int \phi\phi^{'}$ i.e. $(J)_{ij} = <\phi_i, \phi_j>$

$\mathbf{K}_{k\times k} = \int (D^2\phi)(D^2\phi)^{'}$ i.e. $(K)_{ij} = <D^2\phi_i, D^2\phi_j>$

The penalized sample variance is written as

$$PCAPSV = \frac{<\xi, \mathbf{V}\xi>}{\|\xi\|^2 + \lambda\|D^2\xi\|^2} = \frac{y^{'}\mathbf{W}y}{y^{'}\mathbf{J}y + \lambda y^{'}\mathbf{K}y} \text{(book)}$$
$$= \frac{y^{'}\mathbf{JWJ}y}{y^{'}\mathbf{J}y + \lambda y^{'}\mathbf{K}y} \text{(me)}$$

Because:

$$\begin{aligned}\mathbf{V}\xi(s) &= \int v(s,t)\xi(t)dt = \int \phi(s)^{'}\mathbf{W}\phi(t)\phi(t)^{'}ydt \\ &= \phi(s)^{'}\mathbf{W}\int \phi(t)\phi(t)^{'}dty = \phi(s)^{'}\mathbf{W}\mathbf{J}y\end{aligned}$$

$$\begin{aligned}< \xi, \mathbf{V}\xi > &= \int y^{'}\phi(s)\phi(s)^{'}\mathbf{W}\mathbf{J}yds \\ &= y^{'}\int \phi(s)\phi(s)^{'}ds\mathbf{W}\mathbf{J}y = y^{'}\mathbf{J}\mathbf{W}\mathbf{J}y\end{aligned}$$

So the eigenequation is given by

$$\mathbf{J}\mathbf{W}\mathbf{J}y = \rho(\mathbf{J} + \lambda\mathbf{K})y$$

Factorizating $\mathbf{J}+\lambda\mathbf{K} = \mathbf{L}\mathbf{L}^{'}$ and define $\mathbf{S} = \mathbf{L}^{-1}$. The equation is

$$(\mathbf{S}\mathbf{J}\mathbf{W}\mathbf{J}\mathbf{S}^{'})(\mathbf{L}^{'}y) = \rho\mathbf{L}^{'}y$$

Algorithm:
(i)Expand the observed data x_i with respect to the basis ϕ to obtain coefficient vectors c_i.

(ii)Solve $\mathbf{J}^{-1}\mathbf{L}d_i = c_i$ for each i to find the vectors $\mathbf{SJ}c_i = d_i$.

(iii)Carry out a stand PCA on the coefficient vectors d_i.

(iv)Apply the smoothing operator $\mathbf{S}^{'}$ to the resulting eigenvectors u by solving $\mathbf{L}^{'}y = u$ in each case, and renormalize so that the resulting vectors y have $y^{'}\mathbf{J}y = 1$.

(vi)Transform back to find the principal component function ξ using $\xi(s) = y^{'}\phi(s)$.

5. Choosing the Smoothing Parameter by Cross-Validation:

Let $\xi_1, ..., \xi_m$ be m functions, define the component of x orthogonal to the subspace spanned by $\xi_1, ..., \xi_m$ by

$$\zeta_m = x - \sum_{i=1}^{m} \sum_{j=1}^{m} (\mathbf{G}^{-1})_{ij} < \xi_i, x > \xi_j$$

where $\mathbf{G}$ is $m \times m$ matrix and $(G)_{ij} =< \xi_i, \xi_j >$.

The paradigm is the following:
(i)Subtract the overall mean from the observed data x_i.

(ii)For a given smoothing parameter λ, let $\xi_j^{[i]}(\lambda)$ be the estimate of ξ_j obtained from all the data except x_i.

(iii)Define $\zeta_m^{[i]}(\lambda)$ to be the component of x_i orthogonal to the subspace spanned by $\{\xi_j^{[i]}(\lambda) : j = 1, ..., m\}$

(iv)Combine $\zeta_m^{[i]}(\lambda)$ to obtain the cross-validation scores

$$CV_m(\lambda) = \sum_{i=1}^{N} \|\zeta_m^{[i]}(\lambda)\|^2$$

and hence

$$CV(\lambda) = \sum_{m=1}^{\infty} CV_m(\lambda)$$

In practice, we truncate the sum at $m = N - 1$.

(v)Minimize $CV(\lambda)$ to provide the choice of smoothing parameter.

6. An example: The Pinch Force Data:

Choose a grid of values of λ: $\lambda_0 = 0, \lambda_i = 1.5^{i-1}, i = 1, ..., 30$, and attain

$$\hat{\lambda} = argmin_{i=0,...,30} CV(\lambda_i) = 37$$

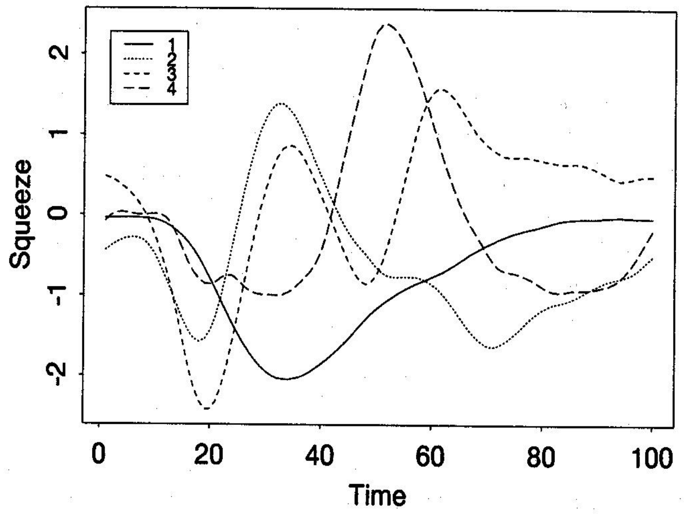

FIGURE 7.3. The first four smoothed principal components for the pinch force data, smoothed by the method of Section 7.3. The smoothing parameter is chosen by cross-validation.

7. Alternative Way: Smoothing the data rather than the PCA:

•

$$PENRSS = \|x - g\|^2 - \lambda \|D^2 g\|^2$$

over g in S. Expand x and g in terms of Fourier Series $\{\phi_j\}_{j=1}^k$, $x = c^{'}\phi$ and $g = d^{'}\phi$.
Then $PENRSS = \|c - d\|^2 + \lambda \sum_{j=1}^k \omega_j^4 d_j^2$.
So $d = \mathbf{S}^2 c$ where $\mathbf{S}_{k \times k}$ is a diagonal matrix $(S)_{jj} = (1 + \lambda\omega_j^4)^{-\frac{1}{2}}$.

• Use same $\hat{\lambda} = 37$ on the same pinch force data.

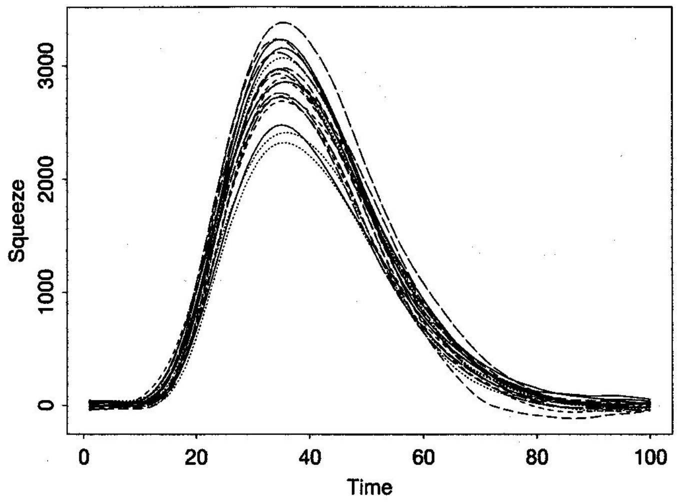

FIGURE 7.5. The pinch force data curves, smoothed by a roughness penalty method with the same smoothing parameter as used for the smoothed PCA, and with the baseline pressure subtracted.

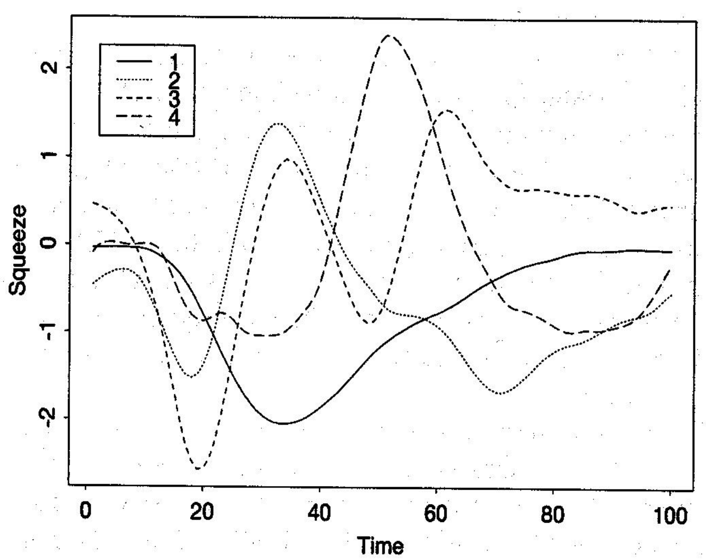

FIGURE 7.6. The first four principal component curves of the smoothed data as shown in Figure 7.5.