Lecture 3: Functional Principal Component Analysis

Elements of Functional Data Analysis

March 2022

Functional Principal Components allow one to find an optimal orthonormal basis in which to represent the data. As a consequence, we can reduce the dimension of infinitely dimensional functional data to a small finite dimension in an optimal way. We then continue to describe functional versions of canonical correlation analysis (CCA), one of the most important tools of multivariate statistical analysis. Following Leurgans et al. [1993] we show that generalisation to infinite dimensions is non-trivial. We shall be following the exposition of Horváth and Kokoszka [2012] and Kokoszka and Reimherr [2017].

1 The Karhunen Loeve expansion through optimisation

Let's recover the KL expansion for an $L^2[0,1]$-valued random variable taking an optimisation approach. To this end we shall seek an orthonormal system of M -elements which best approximates a mean zero square integrable random variable X in $L^2([0,1])$. To this end, consider a random variable X on $L^2([0,1])$ and let $u_1, u_2, \ldots, u_M$ be any orthonormal system on $L^2([0,1])$. Define the loss function:

$$S(u_1, \ldots, u_M) = \mathbb{E}\left\| X - \sum_{i=1}^{M} \langle X, u_k \rangle u_k \right\|^2, \tag{1}$$

over all orthonormal systems $\{u_1, \ldots, u_M\}$. The orthonormality of the functions implies that

$$S(u_1, \ldots, u_M) = \mathbb{E}\|X\|^2 - \sum_{i=1}^{M} \mathbb{E}\langle X, u_i \rangle^2.$$

But the last term can be rewritten as

$$\mathbb{E}\langle X, e_k \rangle^2 = \mathbb{E}\langle u_k, X \otimes X u_k \rangle = \langle \mathcal{K} u_k, u_k \rangle,$$

minimising $S(u_1, \ldots, u_M)$ is equivalent to maximising $\sum_{i=1}^{M} \langle \mathcal{K} u_i, u_i \rangle$. Suppose now that the eigenvalues of the operator $\mathcal{K}$ are ordered $\lambda_1 > \lambda_2 > \lambda_3 > \ldots$ and distinct. From Section 1.5 of Lecture 2, the largest possible value of $\langle u_1, \mathcal{K} u_1 \rangle$ is λ_1 the largest eigenvalue of $\mathcal{K}$ and is attained when u_1 is the associated eigenfunction. Since u_2 is orthogonal to u_1, the largest value of $\langle \mathcal{K} u_2, u_2 \rangle$ is given by λ_2 and u_2 is associated eigenfunction. Continuing inductively, we recover the following result.

Theorem 1.1. *Suppose X is a square integrable mean zero random variable on $L^2([0,1])$ its eigenvalues are ordered and distinct. Then the expected error (1) is minimised when $u_1, u_2, \ldots = e_1, e_2, \ldots$, which are the eigenfunctions of $\mathcal{K}$ associated with $\lambda_1, \lambda_2, \ldots$.*

The associated expansion is then

$$X = \mathbb{E}[X] + \sum_{j=1}^{\infty} \xi_j e_j,$$

where $\xi_j = \langle X - \mu, e_j \rangle$. This is the Karhunen-Loeve expansion as detailed in Theorem 2.8 of Lecture 2. The functions $e_1, e_2, \ldots$ are the *functional principal components* of the random variable X. The last theorem motivatives their derivation in terms of an optimal representation for a given random element X. In practice, we do not have access to X, but rather we have access to independent realisations $X_1, \ldots, X_N$. One must estimate the eigenvalues and eigenfunctions of the inferred covariance operator with much care. This is the main focus of this section.

Immediately we encounter an identifiability conditions, even though the eigenfunctions are normalised so that $\|e_j\|$ but this does not determine the sign of e_j. If $\widehat{e}_j$ is an estimate from the data, we can only hope that $\widehat{c}_j \widehat{e}_j$ is close to v_j where

$$\widehat{c}_j = \text{sign}(\langle \widehat{e}_j, e_j \rangle).$$

For clarity of discussion, let's assume that we're always taking into account the correct sign of the eigenfunction, so that we can omit writing it out every time.

Remark 1.1. Note that $\widehat{c}_j$ cannot be computed from $X_1, \ldots, X_N$, so it must be ensured that statistics we want to work with do not depend on the $\widehat{c}_j$.

In the reminder of this discussion suppose the eigenvalues of the covariance operator are decreasing and distinct $\lambda_1 > \lambda_2 \ldots > \lambda_M > \ldots$. We define the *empirical functional principal compoenents* (EFPC) by

$$\int \widehat{\gamma}(t,s)\widehat{e}_j(s)\, ds = \widehat{\lambda}_j \widehat{e}_j, \quad j = 1, \ldots, M.$$

The system $\{\widehat{e}_1, \ldots, \widehat{e}_M\}$ are estimates of the functional principal components, with $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_M$ being estimates of the associated eigenvalues. The following theorem summarises the consistency and rate of contraction of the estimators to the true values.

Theorem 1.2. *Suppose $\mathbb{E}\|X\|^4 < \infty$ and that the eigenvalues are ordered and distinct then for each $1 \leq j \leq M$,*

$$\limsup_{N \to \infty} N\mathbb{E}\|\widehat{e}_j - e_j\|^2 < \infty, \quad \limsup_{N \to \infty} N\mathbb{E}\|\widehat{\lambda}_j - \lambda_j\|^2 < \infty$$

Note that consistency of the empirical functional principal components requires that their direction is adjusted with the constants $\widehat{c}_j$ as indicated above. Note that if the assumptions in the above theoerm do not hold, then estimates $\widehat{e}_j$ may be very different to e_j, see Johnstone and Lu (2009) for more explicit examples. However, the independence example can be relaxed while still obtaining the same conclusion, for example if the X_i follow a stationary functional time series model such that the dependence between the observations decays sufficiently fast with time separation, see Section 16.2 of Horváth and Kokoszka [2012].

One can also show that, under the above conditions the eigenvalue estimators $\widehat{\lambda}_i$ and $\widehat{e}_i$ are asymptotically normal.

Theorem 1.3. *Let $X_1, \ldots, X_N$ be IID realisations of a random variable X on H with $\mathbb{E}\|X\|^4 < \infty$, and assume the first M eigenvalues are distinct. Then,*

$$\sqrt{N}(\widehat{\lambda}_j - \lambda_j) \xrightarrow{D} \mathcal{N}(0, \langle \Gamma, e_j \otimes e_j \otimes e_j \otimes e_j \rangle), \tag{2}$$

ere is the order 4 tensor given by $\Gamma = \mathbb{E}\left[(\langle X_1 \otimes X_1) \otimes (X_1 \otimes X_1)\right] - \mathbb{E}[X_1 \otimes X_1] \otimes \mathbb{E}[X_1 \otimes X_1]$. *Similarly,*

$$\sqrt{N}(\widehat{e}_j - e_j) \xrightarrow{D} \mathcal{N}(0, C_j), \tag{3}$$

Note that the covariance in (2) is given by

$$\langle \Gamma, e_j \otimes e_j \otimes e_j \rangle = \mathbb{E}[\langle e_j, X\rangle\langle e_j, X\rangle\langle e_j, X\rangle\langle e_j, X\rangle] - \langle \mathcal{K}e_k \otimes e_k\rangle\langle \mathcal{K}e_k \otimes e_k\rangle = 2\lambda_j^2,$$

while the covariance in (3) is given by

$$C_j = \sum_{k\neq l}\sum_{l\neq j}(\lambda_j - \lambda_k)^{-1}(\lambda_j - \lambda_l)^{-1}\langle \Gamma, e_k \otimes e_j \otimes e_l \otimes e_j\rangle\langle e_k \otimes e_l\rangle = \sum_{k\neq j}\frac{\lambda_k\lambda_j}{(\lambda_j - \lambda_k)^2}e_k \otimes e_k.$$

Suppose we observe functions $X_1, \ldots, X_N$. Fix an integer $M < N$. We think of M as being much smaller than M. Once the functional principal components have been computed, we can then accurately represent X_i as

$$\mathbf{x}_i = [\langle X_i, e_1\rangle, \ldots, \langle X_i, e_M\rangle]^\top.$$

This is the same fundamental idea as using spline or Fourier basis expansions; the difference is that the e_k are determined by the observed functions. In most applications, it is important to determine a value of p such that the actual data can be replaced by the approximation. From Exercise 2.1 of Lecture 2 we have that

$$\mathbb{E}\|X\|^2 = \sum_{j=1}^{\infty}\mathbb{E}\langle X, e_j\rangle^2 = \sum_{j=1}^{\infty}\langle \mathcal{K}e_j, e_j\rangle = \sum_{j=1}^{\infty}\lambda_j.$$

If the covariance operator $\mathcal{K}$ is trace class then the sum of $\lambda_j's$ is a finite quantity. We can then consider how much "variance is explained" by a given principal component e_j by calculating

$$\frac{\lambda_j}{\sum_{k=1}^{\infty}\lambda_k}.$$

This gives us some heuristic by which we can determine the index at which to cut-off the basis expansion, e.g. choose M such that the first M principal components account for 85% of the total variance, i.e.

$$\frac{\sum_{i=1}^{M}\lambda_i}{\sum_{j=1}^{\infty}\lambda_j}.$$

If we use the empirical estimation of the eigenvalues $\widehat{\lambda}_j$, we recover the so-called *CPV method* (Cumulative Percentage of Total Variance),

$$\mathrm{CPV}(M) = \frac{\sum_{k=1}^{M}\widehat{\lambda}_k}{\sum_{k=1}^{N}\widehat{\lambda}_k}.$$

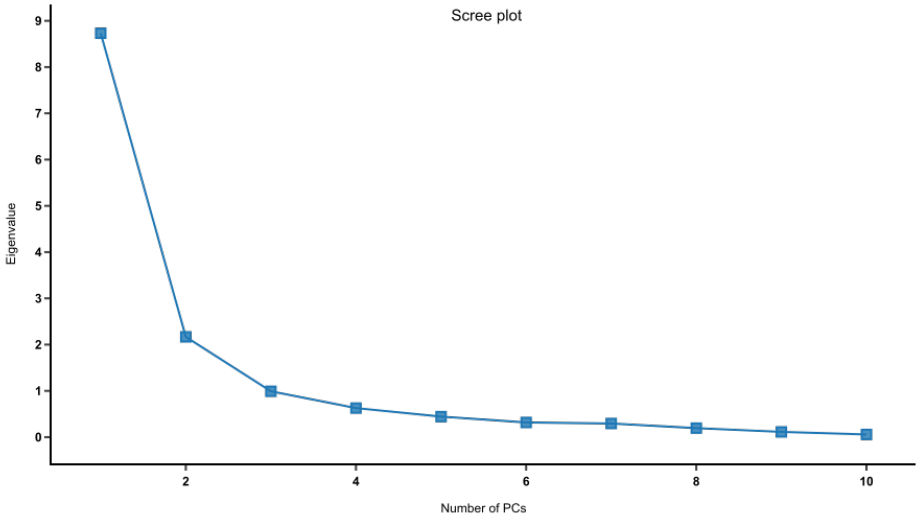

Figure 1: An example of a scree plot for a functional PCA problem. The selection of the cut-off based on scree plots is somewhat subjective, but the above plot suggests it is the first 4 components which account for most ofthe variance.

We choose M for which $CPV(M)$ exceeds a desired level, e.g. 85% is the recommended value for most applications. Another apprroach, similar to the normal multivariate PCA one can resort to *scree plots*. This is a graphical method proposed, in the context of factor analysis, in Cattell [1966]. To apply it, one plots the successive eigenvalues λ_j against j, as shown in Figure 1.

1.1 Computation of functional principal components

The `pca.fd` function in the `fda` library computes the empirical principal component values $\widehat{e}_j$ and corresponding eigenvalues $\widehat{\lambda}_j$ and the *scores* $\langle X_i - \overline{X}_N, \widehat{e}_j \rangle$.

```
pca<-pca.fd(data.fd, nharm = 3, centerfns = TRUE)
```

The argument `nharm` specifies the number M of the principal components (also called harmonics) to be estimated. The flag `centerfns = TRUE` means that the principal components and the scores are computed for the centered functions. Once the object pca has been created $\widehat{e}_1, \ldots, \widehat{e}_M$ are stored in `pca$harmonics`. The scores $\langle X_i - \overline{X}_N, \widehat{e}_j \rangle$ are stored in `pca$scores`.

2 Canonical Correlation Analysis

Functional Canonical Correlation Analysis (FCA) is a tool to quantify correlations between pairs of observed random curves for which a sample is available. Functional or curve data are increasingly common, and the analysis of correlations between functional data has applications in biology , medical sciences (see Leurgans et al. [1993]), ecology, and the environmental sciences (compare

Service et al. [1998])). In Hannan [1961] the authors studied functional canonical analysis for stationary Gaussian processes. A sample version of functional canonical correlation with spline smoothing was proposed in Leurgans et al. [1993], which forms the basis for the following discussion.

Let us first define functional canonical components. We shall assume that we have two separable Hilbert spaces H_1 and H_2. We lose nothing from the discussion if we assume that $H_1 = L^2([0,1])$ and $H_2 = L^2([0,1])$. We consider square integrable random functions $X_1 = X$ and $X_2 = Y$ on H_1 and H_2, respectively, both with mean zero. The canonical components are determined solely by the covariance structure and do not depend on the means. Thus, we define the covariance functions

$$c_{ij}(s,t) = \mathbb{E}[X_i(t)X_j(s)], \quad i,j = 1,2.$$

Next, we define the operators

$$C_{ij} : H_1 \to H_2,$$

defined by

$$(C_{ij}x)(t) = \int c_{ij}(t,s)x(s)\,ds = \mathbb{E}[\langle X_i, x\rangle X_j(t)].$$

The operators $C_{1,1}$ and $C_{2,2}$ are covariance operators, so they are symmetric, positive definite and Hilbert Schmidt. The cross-covariance operator $C_{1,2}$ is also Hilbert Schmidt since

$$\int\int c_{1,2}^2(t,s)\,ds\,dt \leq \mathbb{E}\|X_1\|^2\mathbb{E}\|X_2\|^2.$$

We define the k^{th} canonical correlation k and the associated weight functions a_k and b_k, if they exist, by

$$\rho_k = \mathrm{Cov}(\langle a_k, X_1\rangle, \langle b_k, X_2\rangle) = \sup\left(\mathrm{Cov}(\langle a, X_1\rangle, \langle b, X_2\rangle) \,:\, a \in H_1, b \in H_2\right),$$

subject to the constraints:

$$\mathrm{Var}[\langle a, X_1\rangle] = \mathrm{Var}[\langle b, X_2\rangle] = 1,$$

and also such that, for $A_k = \langle a_j, X_1\rangle$ and $B_k = \langle b_k, X_2\rangle$, the random variables $A_1, A_2, \ldots, B_1, B_2, \ldots$ are all mutually pairwise uncorrelated . We can rewrite this as

$$\rho_k = \langle a_k, C_{12}b_k\rangle = \sup(\langle a, C_{12}b\rangle \,: a \in H_1, b \in H_2),$$

subject to

$$\langle a, C_{11}a\rangle = 1, \text{ and } \langle b, C_{11}b\rangle = 1.$$

Let's revise how one would compute CCA in finite dimensions. To this end, we must solve the constrained optimisation problem where we introduce two lagrange multipliers:

$$\mathcal{L} = \langle a, C_{12}b\rangle - \frac{\rho_1}{2}(\langle a, C_{11}a\rangle - 1) - \frac{\rho_2}{2}(\langle b, C_{22}b\rangle - 1).$$

Taking the derivative with respect to a and b respectively we obtain the relations:

$$C_{12}b - \rho_1 C_{11}a = 0, \quad \text{and } C_{21}a - \rho_2 C_{22}b = 0.$$

Solving for ρ_i we obtain:

$$\langle a, C_{12}b\rangle = \rho_1, \text{ and } \langle b, C_{21}a\rangle = \rho_2,$$

so that $\rho_1 = \rho_2$. Then $a = \frac{C_{11}^{-1} C_{12} b}{\rho}$. Substituting a and b in the above equations, we obtain

$$\rho^2 f = C_{22}^{-1/2} C_{21} C_{11}^{-1} C_{12} C_{22}^{-1/2} f,$$

where $b = C_{22}^{-1/2} f$. We can see that this is a standard eigenvalue problem where $\lambda = \rho^2$ and $A = C_{22}^{-1/2} C_{21} C_{11}^{-1} C_{12} C_{22}^{-1/2}$. We can solve a similar eigenvalue problem for a writing $a = C_{11}^{-1/2} e$ in a similar fashion. We can rewrite the matrix A as $A = R^* R$, where

$$R = C_{11}^{-1/2} C_{12} C_{22}^{-1/2}. \tag{4}$$

It follows that the eigenvalues $\rho_1, \ldots,$ are non-negative.

2.1 Existence of Functional Canonical Components

As discovered by Leurgans et al. [1993], see also Chapter 11 of Ramsay and Silverman [1997], direct application of CCA to functional data is not a meaningful approach because it is possible to find functions a and b such that the sample correlation between $\langle a, X \rangle$ and $\langle b, Y \rangle$ can be made arbitrarily close to 1. The issue basically arises if the correlations between the X and Y variables, is too strong, in an appropriate sense. We first present sufficient conditions under which we can guarantee that standard CCA is well-defined for Hilbert space-valued data.

For an infinite dimensional generalisation to hold we must define an appropriate generalisation of the matrix (4). First, we must have some notion of the square root of a covariance operator $\mathcal{K}$. Recall the decomposition

$$\mathcal{K}(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, e_j \rangle e_j, \quad x \in L^2,$$

in which the λ_j are non-negative, the e_j form a basis and $\mathcal{K} e_j = \lambda_j e_j$. An operator S is called a square root of $\mathcal{K}$ if $SS = \mathcal{K}$. Every positive definite operator has a unique positive-definite square root defined by

$$\mathcal{K}^{1/2} x = \sum_{j=1}^{\infty} \lambda_j^{1/2} \langle x, e_j \rangle e_j, \quad x \in L^2.$$

Direct verification shows that $\mathcal{K}^{1/2}$ is symmetric and positive-definite. The image $\mathrm{Im}(\mathcal{K}^{1/2})$ is the set

$$\mathrm{Im}(\mathcal{K}^{1/2}) = \{ y \in L^2 : \sum_{i=1}^{\infty} \lambda_j^{-1} \langle y, e_j \rangle^2 < \infty \}.$$

We note that this defines a linear subpsace of L^2, but it does not span the whole of L^2. For example $y = \sum_{k=1}^{\infty} \lambda_k^{1/2} e_k$. Since $\sum_{i=1}^{\infty} \lambda_k < \infty$, $y \in L^2$. However, y clearly doesn't lie in $\mathrm{Im}(C_{11}^{1/2})$.

It is clear that further conditions are required for one to be able to construct a generalisation of (4) on an infinite dimensional Hilbert space. The issue arises due to the covariance operators C_{ii}, which arer positive-definite Hilbert-Schmidt operators (and so act as a smoothing operator). If $C_{11}^{1/2}$ has expansion given by

$$C_{11}^{1/2} x = \sum_{i=1}^{\infty} \lambda_i^{1/2} \langle x, e_i \rangle e_i,$$

then we might formally write the inverse as

$$C_{11}^{-1/2} y = \sum_{i=1}^{\infty} \lambda_i^{-1/2} \langle y, e_i \rangle e_i,$$

It is clear however, that $C_{11}^{-1/2} y$ will not lie in H_1 in general, for example, taking $y = \sum_{k=1}^{\infty} \lambda^{1/2} e_i$, then $C_{11}^{-1/2} y$ clearly cannot lie in H_1. It follows that $C_{11}^{1/2}$ will only be invertible over a subspace of H_1, which is given by

$$\{y \in H_1 \,:\, C_{11}^{-1/2} y \in H_1\} = \{y \in H_1 \,:\, \sum_{i=1}^{\infty} \lambda_i^{-1} |\langle y, e_i \rangle|^2 < \infty\} = \mathrm{Im}(C_{11}^{1/2}).$$

We call such an operator *densely defined* on H_1. For a matrix R similar to (4) to be well defined over H_1 we require that the operator C_{12} maps H_2 into the domain of $C_{11}^{-1/2}$, so that $C_{11}^{-1/2} C_{12}$ is well defined, similarly for $C_{22}^{-1/2}$ and C_{21}. The conditions we require are that

$$C_{12}(H_2) \subset \mathrm{Im}(C_{11}^{1/2}) \text{ and } C_{21}(H_1) \subset \mathrm{Im}(C_{22}^{1/2}). \tag{5}$$

We can immediately see how these relationships could be "broken". For example, the "smoother" C_{11} and C_{22} are as operators on H_i, the smaller (forrmally speaking), the images of $C_{ii}^{1/2}$ will be, making this condition harder to hold. Similarly, if C_{12} is "too rough" as an operator between H_1 and H_2. How do we make these conditions precise? To establish a convenient sufficient condition for (5), consider the expansions

$$X = \sum_i \xi_i e_i, \quad Y = \sum_i \zeta_i f_i,$$

where $\xi_i = \langle X, e_i \rangle$, $\zeta_i = \langle Y, f_i \rangle$, and e_i and f_i are the eigenfunctions of the covariance operator C_{11} and C_{22} of X and Y respectively, i.e.

$$C_{11} e_i = \lambda_i e_i, \quad C_{22} f_i = \gamma_i f_i,$$

where $\lambda_i > 0$ and $\gamma_i > 0$. Define the correlation coefficients

$$r_{ji} = \frac{\mathbb{E}[\xi_i \zeta_j]}{\sqrt{\mathbb{E}\xi_i^2 \mathbb{E}\zeta_j^2}} = \frac{\mathbb{E}[\xi_i \zeta_j]}{\lambda_i^{1/2} \gamma_j^{1/2}}.$$

Proposition 2.1. *Under the above conditions,* (5) *will hold if* $\sum_{i,j=1}^{\infty} r_{ji}^2 < \infty$.

Under this condition we can define the cross correlation operator

$$R = C_{11}^{-1/2} C_{12} C_{22}^{-1/2} : \mathrm{Im}(C_{22}^{1/2}) \to H_1.$$

One can show that R and its adjoint $R^* : H_1 \to \mathrm{Im}(C_{22}^{1/2})$. Under the conditions of Proposition 2.1 both R and R^* are Hilbert-Schmidt operators, moreover we can write

$$R e_j = \sum_{k=1}^{\infty} r_{jk} f_k, \quad R^* f_i = \sum_{k=1}^{\infty} r_{ki} e_k.$$

It follows that the operator $M = R^*R : \mathrm{Im}(C_{22}^{1/2}) \to \mathrm{Im}(C_{22}^{1/2})$ is symmetric and positive-definite, and Hilbert Schmidt, being the composition of two Hilbert Schmidt operators. By the Hilbert-Schmidt theorem, the operator M admits a decomposition

$$My = \sum_{k=1}^{\infty} \rho_k^2 \langle y, f_k \rangle f_k,$$

This is not the end of the story. As in the finite dimensional case, we need to map the eigenfunctions e_i and f_i to canonical components a_i and b_i. To ensure the existence of these canonical components, we need to further strengthen condition (5). The reason is that we need to relate e_k and f_k to weight functions a_k and b_k in H_1 and H_2 respectively. To do so, we need $e_k \in \mathrm{Im}(C_{11}^{1/2})(H_1)$ so that $a_k = C_{11}^{-1/2} e_k$ lies in H_1. In terms of r_{ij} we have the following condition.

Proposition 2.2. *Under the above conditions of Proposition 2.1 and the further conditions*

$$\sum_{i,j=1}^{\infty} \lambda_i^{-1} r_{ji}^2 < \infty \text{ and } \sum_{i,j=1}^{\infty} \gamma_i^{-1} r_{ji}^2 < \infty.$$

Then

$$R = C_{11}^{-1/2} C_{12} C_{22}^{-1/2} : Im(C_{22}^{1/2})(H_2) \to Im(C_{11}^{1/2})(H_1),$$
$$S = C_{22}^{-1/2} C_{21} C_{11}^{-1/2} : Im(C_{11}^{1/2})(H_1) \to Im(C_{22}^{1/2})(H_2),$$

and $S = R^*$, $R = S^*$. The associated operators $M = R^*R$ and $M' = S^*S$ have the same eigenvalues ρ_k^2, all eigenvalues are positive and the normalised eigenfunctions e_k of M are related to those of M' by

$$e_k = \rho_k^{-1} R f_k.$$

Moreover, we can define the weight functions

$$a_k = C_{11}^{-1/2} e_k \in H_1 \text{ and } b_k = C_{22}^{-1/2} f_k \in H_2.$$

The following theorem shows that the functional canonical components are defined by ρ_k, a_k and b_k.

Theorem 2.3. *If the assumptions of Proposition 2.2 hold then*

1. $\langle a_k, C_{12} b_k \rangle = \rho_k$, $\langle a_k, C_{11} a_k \rangle = 1$, $\langle b_k, C_{22} b_k \rangle = 1$.
2. $\langle a, C_{12} b \rangle \leq \rho_k \sqrt{\langle a, C_{11} a \rangle \langle b, C_{22} b \rangle}$ *for all* $a \in H_1$, $b \in H_2$ *such that* $\langle a, X \rangle$, $\langle b, Y \rangle$ *are uncorrelated with* $\langle a_j, X \rangle$, $\langle b_j, Y \rangle$ *for* $j < k$.
3. *If* $j \neq k$ *the pairs* $\langle a_j, X \rangle$, $\langle b_j, Y \rangle$ *and* $(\langle a_k, X \rangle, \langle b_k, Y \rangle)$ *are uncorrelated.*

2.2 Sample functional canonical components

Suppose we observe a sample of pairs of functions

$$(X_1, Y_1), \ldots, (X_N, Y_N),$$

and we would like to obtain functional canonical components. We would thus like to maximize the sample correlation between the vectors

$$\widehat{A} = (\langle a, X_1 \rangle, \ldots, \langle a, X_N \rangle)^\top$$

and

$$\widehat{B} = (\langle b, Y_1 \rangle, \ldots, \langle b, Y_N \rangle)^\top$$

The main message of the previous section is that FCA is only well-defined in functional data settings subject to additional assumptions on the covariance operators which are underlying the data. In the general sample setting we wouldn't have access to C_{11}, C_{22} and C_{12} making it hard to ascertain the above sufficient conditions. An alternative approach would be to introduce a *regularisation* of the operators C_{11} and C_{22}, by replacing C_{ii} with $C_{ii,\alpha} = C_{ii} + \alpha I$, where $I : H_i \to H_i$ is an identity operator for $i = 1, 2$. We now note that $C_{11,\alpha}$ has the expansion

$$C_{11,\alpha}^{1/2} x = \sum_{i=1}^{\infty} (\lambda_i + \alpha)^{1/2} \langle x, e_i \rangle e_i,$$

so that

$$C_{11,\alpha}^{-1/2} x = \sum_{i=1}^{\infty} (\lambda_i + \alpha)^{-1/2} \langle x, e_i \rangle e_i,$$

which implies that $\|C_{11,\alpha}^{-1/2} x\| \leq \frac{1}{\alpha^{1/2}} \|x\|$, and similarly for $C_{22,\alpha}$. This means that $C_{ii,\alpha}^{-1/2}$ are now bounded operators on H_i, and in particular, $\mathrm{Im}(C_{ii,\alpha}^{1/2}) = H_i$. It follows that we can define a regularised operator R_α which is an infinite dimensional generalisation of (4) as follows

$$R_\alpha = C_{11,\alpha}^{-1/2} C_{12} C_{22,\alpha}^{-1/2} : H_2 \to H_1.$$

One then solves the associated eigenvalue problem for $M_\alpha = R_\alpha^* R_\alpha : H_2 \to H_2$, obtaining orthonormal bases $\{e_1, e_2, \ldots, \}$ for H_1 and $\{f_1, f_2, \ldots\}$. For each e_k and f_k we can then readily recover the regularised canonical components $a_k = C_{11,\alpha}^{-1/2} e_k$ and $b_k = C_{22,\alpha}^{-1/2} f_k$. One can see (see associated R code), that increasing α discourages the choice of canonical components a, b, which are highly irregular. Typically, this can be chosen subjectively in an exploratorry process to extract informative weight functions a and b, however it can also be chosen automatically via cross validation. We also note that it is trivial to generalise the above discussion to consider separate regularisation parameters α_1 and α_2 for the operators C_{11} and C_{22} respectively.

2.3 Computation of canonical principal components

The `cca.fd` function in the `fda` package carrries out a functional canonical correlation analysis with regularization or roughness penalties on the estimated canonical variables. It is called as follows:

```
cca.smoothed <- cca.fd(fd.s.1, fd.s.2, ncan=3, ccafdPar1, ccafdPar2),
```

where `fd.s.1` and `fd.s.2` are the curves/functional objects associated with the X and Y variables rerspectively. The `ncan` parameter specifies the number of canonical variables and weight functions to be computed. The regularisation parameters are specified through `ccafdPar1` and `ccafdPar2`. The resulting object `cca.smoothed` contains the function weights in `cca.smoothed$weight1` and `cca.smoothed$weight2`. See the associated R example for a more explicit example.

References

Raymond B Cattell. The scree test for the number of factors. *Multivariate behavioral research*, 1 (2):245–276, 1966.

Edward J Hannan. The general theory of canonical correlation and its relation to functional analysis. *Journal of the Australian Mathematical Society*, 2(2):229–242, 1961.

Lajos Horváth and Piotr Kokoszka. *Inference for functional data with applications*, volume 200. Springer Science & Business Media, 2012.

Piotr Kokoszka and Matthew Reimherr. *Introduction to functional data analysis*. CRC Press, 2017.

Sue E Leurgans, Rana A Moyeed, and Bernard W Silverman. Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(3): 725–740, 1993.

James Ramsay and Bernard W Silverman. Functional data analysis (springer series in statistics). 1997.

SK Service, JA Rice, and FP Chavez. Relationship between physical and biological variables during the upwelling period in monterey bay, ca. *Deep Sea Research Part II: Topical Studies in Oceanography*, 45(8-9):1669–1685, 1998.