MODELLISTICA E CALCOLO SCIENTIFICO
MODELING AND SCIENTIFIC COMPUTING

MOX–Report No. 35/2012

Distances and Inference for Covariance Functions

Pigoli, D.; Aston, J.A.D.; Dryden, I.L.; Secchi, P.

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Distances and Inference for Covariance Functions

Davide Pigoli[♯], John A. D. Aston[♭], Ian L. Dryden[†], Piercesare Secchi[♯]

[♯] MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
Milano, Italy
davide.pigoli@mail.polimi.it, piercesare.secchi@polimi.it

[♭] CRiSM, Dept of Statistics
University of Warwick
Coventry, UK
j.a.d.aston@warwick.ac.uk

[†] Department of Statistics
University of South Carolina
Columbia, SC 29208, USA
dryden@mailbox.sc.edu

Keywords: Distance Metrics; Functional Data Analysis; Procrustes Analysis; Shape Analysis.

Abstract

A framework is developed for inference concerning the covariance operator of a functional random process, where the covariance operator itself is an object of interest for the statistical analysis. Distances for comparing positive definite covariance matrices are either extended or shown to be inapplicable for functional data. In particular, an infinite dimensional analogue of the Procrustes size and shape distance is developed. The convergence of the finite dimensional approximations to the infinite dimensional distance metrics is also shown. To perform inference, a Fréchet estimator for the average covariance function is introduced, and a permutation procedure to test the equality of the covariance operator between two groups is then considered.

The proposed techniques are applied to two problems where inference concerning the covariance is of interest. Firstly, in data arising from a study into cerebral aneurysms, it is of interest to determine whether two groups of data can be combined when comparing with a third group. For this to be done, it is necessary to assess whether the covariance structures of the two groups are the same or different. Secondly, in a philological study of cross-linguistic dependence, the use of covariance operators has been suggested as a way to incorporate quantitative phonetic information. It is shown that distances between languages derived from phonetic covariance functions can provide insight into relationships between the Romance languages.

1 Introduction

Data sets are increasingly becoming available that are best described as being functional. In recent years, research in this field has provided many statistical techniques to deal with these kinds of data (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006). However, this work has mainly focused on mean functions and using associated bases to provide insight into these means, and little attention has been paid to the explicit analysis of the covariance operator. In many applied problems, the covariance operator is either directly or indirectly intrinsically interesting in its own right. This paper is primarily focused on providing inference for the covariance operator of a functional random process.

While little studied, there is some recent work that has examined testing the equality of covariance structures from two groups of functional curves by defining a test statistic through the Karhunen-Loéve expansions of the two covariance structures (Panateros et al., 2010; Fremdt et al., 2012). The techniques we propose here will take a somewhat different view, in that, the underlying objects of interest in our studies are the covariance operators themselves, not necessarily the underlying curves, and as such we will approach the problem through the definition of functional distance metrics for covariance operators. While it will be seen that some finite dimensional distances for positive definite covariance matrices (Dryden et al., 2009) naturally lend themselves to functional analogues, others do not have natural extensions. However, it will be shown that the extended metrics can be intuitively understood as providing methods for testing different properties of the underlying covariance structure.

Analysis of the covariance operator arises in many applied contexts, two of which will be detailed in Section 5. Firstly, in data associated with brain aneurysms, to gain power, several patient populations are routinely considered similar enough to treat as one population. We will explore their covariance structures, and assess through our distance based permutation tests whether they can indeed be so combined. Secondly, in the linguistic analysis of human speech, the overall mean structure of the data produced is often not of interest, but rather the variations that can be found within the language. Here we will show that different languages can be compared and even predicted through functional distances, allowing, for the first time, a quantitative analysis of comparative philological relations based on speech recordings rather than discrete textual analysis.

2 Some remarks on operators on $L^2(\Omega)$

In this section we focus on properties and definitions that will be useful below. More details and proofs can be found, e.g., in Zhu (2007).

Definition 2.1 *Let B_1 be the closed ball of unitary radius in $L^2(\Omega)$, consisting of all $f \in L^2(\Omega)$ such that $||f||_{L^2(\Omega)} \leq 1$, where $L^2(\Omega)$ is the Hilbert space of square-*

integrable functions on $\Omega \subseteq \mathbb{R}$. A bounded linear operator $T : L^2(\Omega) \to L^2(\Omega)$ is compact if $T(B_1)$ is compact in the norm of $L^2(\Omega)$. A bounded linear operator T is self-adjoint if $T = T^*$

An important property of a compact operator on $L^2(\Omega)$ is the existence of a canonical decomposition. This decomposition implies that two orthonormal bases $\{u_k\}, \{v_k\}$ for $L^2(\Omega)$ exist so that

$$Tf = \sum_k \sigma_k \langle f, v_k \rangle u_k,$$

or, equivalently,

$$Tv_k = \sigma_k u_k,$$

where $\langle ., . \rangle$ indicates the scalar product in $L^2(\Omega)$. The sequence $\{\sigma_k\} \in \mathbb{R}$ is called the sequence of singular values for T. If the operator is self-adjoint, there exists a basis $\{v_k\}$ such that

$$Tf = \sum_k \lambda_k \langle f, v_k \rangle v_k,$$

or, equivalently,

$$Tv_k = \lambda_k v_k$$

and $\{\lambda_k\} \in \mathbb{R}$ is called the sequence of eigenvalues for T.

A compact operator T is said to be *trace class* if

$$\text{trace}(T) := \sum_k \langle Te_k, e_k \rangle < +\infty$$

for an orthonormal basis $\{e_k\}$. It can be seen that the definition is independent of the choice of the basis. We indicate with $S(L^2(\Omega))$ the space of the trace class operators on $L^2(\Omega)$.

A compact operator T is said to be Hilbert-Schmidt if its Hilbert-Schmidt norm is bounded, i.e.

$$||T||_{HS}^2 = \text{trace}(T^*T) < +\infty.$$

This is a generalisation of the Frobenius norm for finite-dimensional matrices.

These properties are crucial in the context of the statistical analysis of functional data. Indeed, let $\mathbf{f}$ be a random function which takes values in $L^2(\Omega)$, $\Omega \subseteq \mathbb{R}$, such that $\mathbb{E}[||\mathbf{f}||_{L^2(\Omega)}^2] < +\infty$. The covariance operator is

$$C_{\mathbf{f}} g(t) = \int_\Omega c_{\mathbf{f}}(s,t) g(s) \mathrm{d}s,$$

for $g \in L^2(\Omega)$ and

$$c_{\mathbf{f}}(s,t) = \text{cov}(\mathbf{f}(s), \mathbf{f}(t)) = \mathbb{E}\left[\left(\mathbf{f}(s) - \mathbb{E}\left[\mathbf{f}(s)\right]\right)\left(\mathbf{f}(t) - \mathbb{E}\left[\mathbf{f}(t)\right]\right)\right].$$

Then, $C_{\mathbf{f}}$ is a trace class, self-adjoint, compact operator on $L^2(\Omega)$ with non negative eigenvalues (see, e.g., Bosq, 2000, Section 1.5).

Finally, we recall the definition of unitary operator on $L^2(\Omega)$, which will be needed for the definition of Procrustes distance in the functional setting.

Definition 2.2 *A bounded linear operator R on $L^2(\Omega)$ is said to be unitary if*

$$||Rf||_{L^2(\Omega)} = ||f||_{L^2(\Omega)} \quad \forall f \in L^2(\Omega).$$

We indicate with $O(L^2(\Omega))$ the space of unitary operators on $L^2(\Omega)$.

We take now advantage of these tools from functional analysis to introduce possible distances that can be used to measure the difference between two covariance operators.

3 Distances between covariance operators

In this section we propose several distances that can be used to compare the covariance operators of two random functions taking values in $L^2(\Omega)$. These are a generalisation to the functional setting of metrics that have been proved useful for the case of positive semi-definite matrices (Dryden et al., 2009). However, not all matrix based distances are extendable in the functional case.

Two popular metrics for finite dimensional covariance matrix analysis are the log Euclidean metric and the affine invariant Riemannian metric. While both would appear to be natural candidates for generalisation to covariance operators, in both cases, this is not straightforward due to the natural trace class structure of the covariance operator. The trace class property implies that the (descending) ordered eigenvalues λ_i are summable, i.e.

$$\sum_i \lambda_i < \infty \Rightarrow \lambda_i \to 0 \text{ as } i \to \infty$$

The log Euclidean distance for two positive definite matrices, M_1 and M_2, is defined as

$$d_{\log}(M_1, M_2) = || \log(M_1) - \log(M_2)||.$$

with the $\log()$ indicating the matrix logarithm. This is not well defined for trace class operators as this quantity tends to infinity. The affine invariant Riemannian metric for positive definite matrices is defined as

$$d_{\text{Riem}}(M_1, M_2) = || \log(M_1^{-\frac{1}{2}} M_2 M_1^{-\frac{1}{2}})||$$

which requires consideration of the inverse. For a compact operator, even when it is positive definite, the inverse is not well defined (see, e.g., Zhu, 2007, Section 1.3).

Even though in applications only finite dimensional representations are available, these are usually not full rank (i.e. they have zero eigenvalues), meaning that these metrics have to be computed on subspaces which should be carefully chosen to avoid instability in the computation of the distance (coming from small eigenvalues) while taking into account all the significant information. Moreover, since the distance between infinite dimensional operators is not well defined, it is not clear how to interpret the asymptotic behaviour of the distance computed between finite dimensional representations. This could be an issue when the dimensionality of the problem is high and different choices are possible for the reduced space.

We thus resort to some alternative distances which are well defined for self-adjoint trace class operators with nonnegative eigenvalues.

3.1 Distance between kernels in $L^2(\Omega \times \Omega)$

Distances between covariance operators can be naturally defined using the distance between their kernels in $L^2(\Omega \times \Omega)$. Let S_1 and S_2 be two covariance operators and

$$S_i f(t) = \int_\Omega s_i(s,t) f(s) \mathrm{d}s, \quad \forall \ f \in L^2(\Omega).$$

Then, we can define the distance

$$d_L(S_1, S_2) = ||s_1 - s_2||_{L^2(\Omega\times\Omega)} = \sqrt{\int_\Omega \int_\Omega (s_1(x,y) - s_2(x,y))^2 \mathrm{d}x\mathrm{d}y}.$$

This distance is correctly defined, since it inherits all the properties of the distance in the Hilbert space $L^2(\Omega\times\Omega)$. However, it does not exploit in any way the particular structure of covariance operators and therefore it need not to be useful for highlighting significant differences between covariance structures. In addition, as will be seen later, it is not constrained to always provide estimates within the space of covariance operators.

3.2 Spectral distance

A second possibility is to regard the covariance operator as an element of $\mathfrak{L}(L^2(\Omega))$, the space of the linear bounded operators on $L^2(\Omega)$. It follows that the distance between S_1 and S_2 can be defined as the operator norm of the difference. We recall that the norm of a self-adjoint bounded linear operator on $L^2(\Omega)$ is defined as

$$||T||_{\mathfrak{L}(L^2(\Omega))} = \sup_{v \in L^2(\Omega)} \frac{|\langle Tv, v\rangle|}{||v||^2_{L^2(\Omega)}}$$

and for a covariance operator it coincides with the absolute value of the first (i.e. largest) eigenvalue. Thus,

$$d_{\mathfrak{L}}(S_1, S_2) = ||S_1 - S_2||_{\mathfrak{L}(L^2(\Omega))} = |\widetilde{\lambda}_1|$$

where $\widetilde{\lambda}_1$ is the first eigenvalue of the operator $S_1 - S_2$. The distance $d_{\mathfrak{L}}(.,.)$ generalises the matrix spectral norm which is often used in the finite dimensional case (see, e.g., El Karoui, 2008). This distance takes into account the spectral structure of the covariance operators, but it appears restrictive in that it focuses only on the behaviour on the first mode of variation.

3.3 Square root operator distance

Since covariance operators are trace class, we can generalise the square root matrix distance (see Dryden et al., 2009). Indeed, S being a self-adjoint trace class operator, there exists a Hilbert-Schmidt self adjoint operator $(S)^{\frac{1}{2}}$ defined as

$$(S)^{\frac{1}{2}} f = \sum_k \lambda_k^{\frac{1}{2}} \langle f, v_k \rangle v_k, \tag{1}$$

where λ_k are eigenvalues and v_k eigenfunctions of S. We can therefore define the square root distance between two covariance operators S_1 and S_2 as

$$d_R(S_1, S_2) = ||(S_1)^{\frac{1}{2}} - (S_2)^{\frac{1}{2}}||_{HS}.$$

Inspiration for this kind of distance comes from the log-Euclidean distance for positive definite matrices. There, the logarithmic transformation allows to map the non Euclidean space in a linear space. As mentioned above, a logarithmic map for covariance operators is not available. Thus, we choose a different transformation, namely, the square root transformation. This has been shown to behave in a similar way in the finite dimensional setting (see Dryden et al., 2009) but it is also well defined for trace class operators.

Any power greater than $1/2$ would be a possible candidate distance, but for general trace class operators, the square root operator is the smallest power that can be defined while still ensuring finite distances, meaning that it is the closest available to the log-Euclidean distance. In addition, it can be interpreted as a distance which takes into account the full eigenstructure of the covariance operator, both eigenfunctions and eigenvalues.

3.4 Procrustes size-and-shapes distance

The square root operator distance looks at the distance between the square root operators $(S_1)^{\frac{1}{2}}$ and $(S_2)^{\frac{1}{2}}$ in the space of Hilbert-Schmidt operators. However, this is only a particular choice of a broad family of distances, which are based on the mapping of the two operators S_1 and S_2 from the space of covariance operators to the space of Hilbert-Schmidt operators. We can consider in general a transformation $S_i \to L_i$, so that $S_i = L_i L_i^*$ and define the distance as the Hilbert-Schmidt norm of $L_1 - L_2$. Considering this more general framework, it is easy to see that any of this transformation is defined up to a unitary operators R:

$$(L_i R)(L_i R)^* = L_i R R^* L_i^* = L_i L_i^* = S_i.$$

To avoid the arbitrariness of the transformation, it is meaningful to use a Procrustes approach which looks for the unitary operator R which best matches the two operators L_1 and L_2, however they are defined.

In Dryden et al. (2009), a Procrustes size-and-shape distance is proposed to compare two positive definite matrices. Our aim is to generalise this distance to the case of covariance operators on $L^2(\Omega)$. Let S_1 and S_2 be two covariance operators on L^2. We define the square of the Procrustes distance between S_1 and S_2 as

$$d_P(S_1, S_2)^2 = \inf_{R \in O(L^2(\Omega))} ||L_1 - L_2 R||^2_{HS} = \inf_{R \in O(L^2(\Omega))} \text{trace}((L_1 - L_2 R)^*(L_1 - L_2 R)),$$

where $||.||_{HS}$ indicates the Hilbert-Schmidt norm on $L^2(\Omega)$ and L_i are such that $S_i = L_i L_i^*$ for $i = 1, 2$.

As mentioned above, the decomposition $S_i = L_i L_i^*$ can be seen as a general form of transformation, mapping S_i to a space where a linear metric is appropriate. In particular, a good choice could be the square root transformation.

Proposition 3.1 *Let σ_k be the singular values of the compact operator $L_2^* L_1$. Then*

$$d_P(S_1, S_2)^2 = ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\sum_{k=1}^{+\infty} \sigma_k$$

Proof. Note that

$$\begin{aligned} d_P(S_1,S_2)^2 &= \inf_{R \in O(L^2(\Omega))} \text{trace}((L_1 - L_2 R)^*(L_1 - L_2 R)) \\ &= \inf_{R \in O(L^2(\Omega))} \{\text{trace}(L_1^* L_1) + \text{trace}(L_2^* L_2) - 2\text{trace}(R^* L_2^* L_1)\} \\ &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2 \sup_{R \in O(L^2(\Omega))} \text{trace}(R^* L_2^* L_1). \end{aligned}$$

We therefore look for the unitary operator R which maximises $\text{trace}(R^* L_2^* L_1)$. Exploiting the definition of the trace operator and the singular value decomposition for the compact operator $L_2^* L_1$ (which is trace class - see Bosq (2000, Section 1.5)),

$$L_2^* L_1 v_k = u_k \text{ for } k = 1, \dots, +\infty,$$

we obtain

$$\text{trace}(R^* L_2^* L_1) = \sum_{k=1}^{+\infty} \langle R^* L_2^* L_1 e_k, e_k \rangle = \sum_{k=1}^{+\infty} \langle R^* L_2^* L_1 v_k, v_k \rangle =$$

$$= \sum_{k=1}^{+\infty} \sigma_k \langle R^* u_k, v_k \rangle \le \sum_{k=1}^{+\infty} \sigma_k ||R^* u_k||_{L^2(\Omega)} ||v_k||_{L^2(\Omega)} = \sum_{k=1}^{+\infty} \sigma_k ||u_k||_{L^2(\Omega)} ||v_k||_{L^2(\Omega)} = \sum_{k=1}^{+\infty} \sigma_k.$$

Thus, the maximum is reached for the operator $\widetilde{R}$ such that

$$\widetilde{R}^* u_k = v_k \ \ \forall k = 1, \dots, +\infty,$$

or, equivalently,

$$\widetilde{R} v_k = u_k \ \ \forall k = 1, \dots, +\infty.$$

Substituting this optimal transformation in the definition of the distance,

$$\begin{aligned} d_P(S_1,S_2)^2 &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\text{trace}(\widetilde{R}^* L_2^* L_1) \\ &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\sum_{k=1}^{+\infty} \langle \widetilde{R}^* L_2^* L_1 e_k, e_k \rangle \\ &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\sum_{k=1}^{+\infty} \langle \widetilde{R}^* L_2^* L_1 v_k, v_k \rangle \\ &= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\sum_{k=1}^{+\infty} \sigma_k \underbrace{\langle \widetilde{R}^* u_k, v_k \rangle}_{=1}. \quad \square \end{aligned}$$

□

The Procrustes distance takes into account the arbitrariness in the definition of the map $S_i \to L_i$. It is worth noticing that the unitary transformation allows the operator L_i to become non self-adjoint. Thus, this extends the analysis in the Hilbert-Schmidt space to go beyond symmetric operators applied to a function inducing covariances to explanations that include operators such triangular operators with interpretations such as causal flow directions in time.

3.5 Finite dimensional approximation

In practical applications, we observe only a finite dimensional representation of the operators of interest. Therefore, ideally we would require square root distance and Procrustes size-and-shape distance between two finite dimensional representations to be a good approximation of the distance between the infinite dimensional operators. We show this fact for the more general case of Procrustes distance, with the square root distance being a special case where $L_i = (S_i)^{\frac{1}{2}}$ and R is constrained to be the identity operator.

Let $\{e_k\}_{k=1}^{+\infty}$ be a basis for $L^2(\Omega)$, $V_p = \text{span}\{e_1, \ldots, e_p\}$ and S_i^p be the restriction of S_i on V_p, i.e.

$$S_i^p g = \sum_{k=1}^{p} \langle g, e_k \rangle S_i e_k \quad \forall \; g \in V_p.$$

In practical situations, V_p will be the subspace which contains the finite dimensional representation of the functional data. Let us assume that, for $p \to +\infty$, $L_i^p \to L_i$ with respect to the Hilbert-Schmidt norm, where $S_i^p = L_i^p L_i^{p*}$. This is not restrictive, since we can choose for instance $L_i = (S_i)^{\frac{1}{2}}$, but every choice which guarantees this convergence is suitable. Then, the distance between the two restricted operators is

$$d_P(S_1^p, S_2^p)^2 = ||L_1^p||_{HS}^2 + ||L_2^p||_{HS}^2 - 2 \sum_{k=1}^{p} \langle \widetilde{R}^p L_2^{p*} L_1^p e_k, e_k \rangle.$$

Since $V_p \subset L^2(\Omega)$, we can choose a subset $v_1^p, \ldots, v_p^p$, $v_k^p \in \{v_k\}_{k=1}^{+\infty}$ which is an orthonormal basis for V_p. However, they need not be the first p elements of the basis coming from the canonical decomposition of $L_2^* L_1$. This happens because the space V_p depends only on the original basis $\{e_k\}_{k=1}^{p}$ and it does not depend on the covariance structure of the data. Since the subspaces V_p are nested, we can define a permutation $s : \mathbb{N} \to \mathbb{N}$, so that $\{v_{s(1)}, \ldots, v_{s(p)}\}$ provides a basis for V_p, for every p. Since the trace of an operator does not depend on the choice of the basis, we obtain

$$d_P(S_1^p, S_2^p)^2 = ||L_1^p||_{HS}^2 + ||L_2^p||_{HS}^2 - 2 \sum_{k=1}^{p} \langle \widetilde{R}^p L_2^{p*} L_1^p v_{s(k)}, v_{s(k)} \rangle$$

$$= ||L_1^p||_{HS}^2 + ||L_2^p||_{HS}^2 - 2 \sum_{k=1}^{p} \sigma_{s(k)},$$

where $\{\sigma_{s(k)}\}_{k=1}^{p}$ are singular values for $L_2^* L_1$. This comes from the fact that the action of the operator $L_2^{p*} L_1^p$ should be equal to the action of the operator $L_2^* L_1$ on every element belonging to the subspace V_p and $v_{s(k)} \in V_p$ for $k = 1, \ldots, p$. Finally, as $L_2^* L_1$ is trace class, the series of its singular values is absolutely convergent and therefore also unconditionally convergent (convergent under any permutation). Thus,

$$\lim_{p \to +\infty} d_P(S_1^p, S_2^p)^2 = ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2 \sum_{k=1}^{+\infty} \sigma_{s(k)}$$

$$= ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2 \sum_{k=1}^{+\infty} \sigma_k = d_P(S_1, S_2)^2.$$

4 Statistical inference for covariance operators

In this section we illustrate two inferential problems that can be addressed using the distances introduced in Section 3. First, we consider the estimation of a common covariance operator. Then, we propose a procedure to test the equality of the covariance operator between two groups.

4.1 Fréchet averaging with Square root and Procrustes distances

In many applications averaging among covariance operators of different groups is needed. Let $S_1, \ldots, S_g$ be the covariance operators, sampled from the same distribution on the space of covariance operators. Then, a possible estimator of the mean covariance operator Σ is

$$\widehat{\Sigma} = \frac{1}{n_1 + \cdots + n_g}(n_1 S_1 + \cdots + n_g S_g).$$

where weights n_i, $i = 1, \ldots, g$ are the numbers of observations from which the covariance operator S_i has been obtained. This formula arises from the minimisation of square Euclidean deviations, weighted with the number of observations. If we choose a different distance to compare covariance operators, it is more coherent to average covariance operators with respect to the chosen distance.

The Fréchet mean of a random element S, with probability distribution μ on the space of covariance operators, can be defined as $\Sigma = \operatorname{arginf}_P \int d(S,P)^2 \mu(dS)$. If a sample S_i, $i = 1, \ldots, g$ from μ is available, a least square estimator for Σ can be defined using the weighted sample Fréchet mean:

$$\widehat{\Sigma} = \arg\inf_S \sum_{i=1}^{g} n_i d(S, S_i)^2,$$

The actual computation of the sample Fréchet mean $\widehat{\Sigma}$ depends on the choice of the distance $d(.,.)$. In general, it requires the solution of a high dimensional minimisation problem but some distances admit an analytic solution while for others efficient minimisation algorithms are available. Note that $\widehat{\Sigma}$ may not be unique for positively curved spaces, although it is unique for suitably concentrated data (see Kendall, 1990; Le, 2001).

Proposition 4.1 *For the square root distance d_S,*

$$\widehat{\Sigma} = \arg\min_S \sum_{i=1}^{g} n_i d_S(S, S_i)^2 = (\frac{1}{G}\sum_{i=1}^{g} n_i (S_i)^{\frac{1}{2}})^2. \tag{2}$$

where $G = n_1 + \cdots + n_g$.

Proof. We prove that in general

$$\arg\min_L \sum_{i=1}^{g} n_i ||L - L_i||_{HS}^2 = \frac{1}{G}\sum_{i=1}^{g} n_i L_i,$$

which gives the desired result for the particular case of $L_i = (S_i)^{\frac{1}{2}}$ and $L = (S)^{\frac{1}{2}}$. We have

$$\begin{aligned}
\arg\min_L \sum_{i=1}^{g} n_i ||L - L_i||_{HS}^2 &= \arg\min_L \sum_{i=1}^{g} n_i \mathrm{trace}\big((L - L_i)^*(L - L_i)\big) \\
&= \arg\min_L \sum_{i=1}^{g} n_i \{||L||_{HS}^2 + ||L_i||_{HS}^2 - 2\mathrm{trace}(L^* L_i)\} \\
&= \arg\min_L \sum_{i=1}^{g} n_i \{||L||_{HS}^2 - 2\mathrm{trace}(L^* L_i)\} \\
&= \arg\min_L \Big[G||L||_{HS}^2 - 2\mathrm{trace}\Big(L^* \sum_{i=1}^{g} n_i L_i\Big)\Big],
\end{aligned}$$

exploiting the linearity of the trace operator. It can be noticed that the second term is a scalar product in the operator space and therefore it is minimum when L is proportional to $\sum_{i=1}^{g} n_i L_i$. We thus obtain a minimisation problem in $\alpha = ||L||_{HS}$:

$$\begin{aligned}
&\arg\min_L G||L||_{HS}^2 - 2\mathrm{trace}\Big(L^* \sum_{i=1}^{g} n_i L_i\Big) \\
&\quad = \arg\min_\alpha G\alpha^2 - 2\mathrm{trace}\Big(\frac{\alpha}{||\sum_{i=1}^{g} n_i L_i||_{HS}} \Big(\sum_{i=1}^{g} n_i L_i\Big)^* \Big(\sum_{i=1}^{g} n_i L_i\Big)\Big) \\
&\quad = \arg\min_\alpha G\alpha^2 - 2\frac{\alpha}{||\sum_{i=1}^{g} n_i L_i||_{HS}} \underbrace{\mathrm{trace}\Big(\Big(\sum_{i=1}^{g} n_i L_i\Big)^* \Big(\sum_{i=1}^{g} n_i L_i\Big)\Big)}_{=||\sum_{i=1}^{g} n_i L_i||_{HS}^2} \\
&\quad = \arg\min_\alpha G\alpha^2 - 2\alpha||\sum_{i=1}^{g} n_i L_i||_{HS}
\end{aligned}$$

and the minimum is reached for $\alpha = \frac{1}{G}||\sum_{i=1}^{g} n_i L_i||_{HS}$. Therefore

$$L = \frac{\alpha}{||\sum_{i=1}^{g} n_i L_i||_{HS}} \sum_{i=1}^{g} n_i L_i = \frac{1}{G}\sum_{i=1}^{g} n_i L_i. \quad \square$$

□

For the Procrustes size-and-shape distance an analytic solution is not available. However, the Procrustes mean can be obtained by an adaptation of the algorithm proposed in Gower (1975). It is an iterative method that alternates a registration step and an averaging step.

1.Initialization The algorithm is initialized with $\widehat{\Sigma}^{(0)} = L^{(0)}L^{(0)*}$, where $L^{(0)} = \frac{1}{G}\sum_{i=1}^{g} n_i L_i$, L_i so that $S_i = L_i L_i^*$ and $G = n_1 + \cdots + n_g$.

2.Registration step For all the groups $i = 1, \ldots, g$, $L_i^{(k)} = L_i^{(k-1)} R_i$, where R_i is the unitary operator which minimises the Hilbert Schmidt norm of $L^{(k-1)} - L_i^{(k-1)} R_i$.

3.Averaging step The new Procrustes mean is computed: $\widehat{\Sigma}^{(k)} = L^{(k)}L^{(k)*}$, where $L^{(k)} = \frac{1}{G}\sum_{i=1}^{g} n_i L_i^{(k)}$, since this minimises $\sum_{i=1}^{g} n_i ||L - L_i^{(k)}||_{HS}^2$ as shown in the proof of Proposition 4.1.

Steps 2 and 3 are iterated until convergence, i.e. when the Hilbert-Schmidt norm of the difference between $L^{(k)}$ and $L^{(k-1)}$ is below a chosen tolerance. In practice, the algorithm will give a local minimum, often called Karcher mean (Karcher, 1977), in few iterations, if it is initialised with the estimate provided by (2).

The algorithm above is adapted from one of a number of variants of the Procrustes algorithm, all of which have been shown in the finite dimensional setting to have similar convergence properties (see Groisser, 2005). It is conjectured that analogous convergence properties are also true in the infinite dimensional setting (in particular that the finite dimensional algorithm converges to the correct infinite dimensional limit), but the geometric arguments using in the finite dimensional proof by Groisser (2005) are not immediately available for the infinite dimensional setting and we leave this for future work.

We also compared our version of the algorithm with the one proposed by Dryden and Mardia (1998), where each operator L_i is aligned to the average obtained from all the other operators, namely $\frac{1}{G-n_i}\sum_{j\neq i}^{g} n_j L_j^{(k)}$. However, this algorithm, in the examples below, provided the same result as the one above, while also having very similar convergence speed and computational burden.

4.2 A permutation test for two - sample comparison of the covariance structure

In this section we show an example of how the proposed distances can be used in an inferential procedure. We would like to use the distance between two sample covariance operators to carry out inference on the difference between the true covariance operators. However, the complicated expression of the available distances makes it difficult to elicit their distributions, even when random curves are generated from a known parametric model. Thus, we propose to resort to a non parametric approach, namely permutation tests.

Permutation tests are non parametric tests which rely on the fact that, if there is no difference among experimental groups, the labelling of the observations is completely arbitrary. Therefore, the null hypothesis that the labels are arbitrary is tested by comparing the test statistic with its permutation distribution, i.e. the value of the test statistics for all possible permutations of labels. In practice, only a subset of permutations, chosen at random, is used to assess the distribution. A sufficient condition to apply this permutation procedure is exchangeability: under the null hypothesis, curves can be assigned indifferently to any group.

Let us consider two samples of random curves. Curves in the first sample $f_1^1(t), \ldots, f_{n_1}^1(t) \in L^2(\Omega)$ are realisations of a random process with mean $\mu(t)$ and covariance operator Σ_1. Curves in the second sample $f_1^2(t), \ldots, f_{n_2}^2(t) \in L^2(\Omega)$ are realisations of a random process with mean $\mu(t)$ and covariance operator Σ_2. We would like to test the hypothesis

$$\mathrm{H}_0 : \Sigma_1 = \Sigma_2 \text{ vs } \mathrm{H}_1 : \Sigma_1 \neq \Sigma_2.$$

We reformulate the test using distances between covariance operators,

$$\mathrm{H}_0 : d(\Sigma_1, \Sigma_2) = 0 \text{ vs } \mathrm{H}_1 : d(\Sigma_1, \Sigma_2) \neq 0.$$

Let S_1 and S_2 be the sample covariance operators of the two groups. We use $d(S_1, S_2)$ as a test statistic, since large values of $d(S_1, S_2)$ are evidence against the null hypothesis.

For this formulation of the permutation test, equality of mean functions is essential. However, if the two groups have different (and unknown) means, an approximated permutation test can be performed, having first centred the curves using their sample means. This is a common strategy for testing scaling parameters, such as variance, for univariate real random variables (see, e.g., Good, 2005, Section 3.7.2). The test obtained is approximate in the sense that the nominal level of the test is exact only asymptotically for $n_1, n_2 \to +\infty$. This happens because the observations are only asymptotically exchangeable, due to the fact that $\hat{\mu}_i(t) = \frac{1}{n_i} \sum_{k=1}^{n_i} f_k^i(t) \to \mu_i(t)$ and therefore centred observations asymptotically do not depend on the original labels.

Now we have all the ingredients to set up a permutation test. We consider M random permutation of the labels $\{1, 2\}$ on the sample curves and we compute $d(S_1^{(m)}, S_2^{(m)})$, $m = 1, \ldots, M$, where $S_i^{(m)}$ is the sample covariance operator for the group indexed with label i in permutation m. The p-value of the test is the proportion of $d(S_1^{(m)}, S_2^{(m)})$ which are greater than or equal to $d(S_1, S_2)$.

4.3 Some simulations

We now consider simulation studies to explore the behaviour of the different distances with various modifications of the covariance structure. All the curves are simulated on $[0, 1]$ with a Gaussian process with mean $\sin(x)$ and covariance function Σ_1 and Σ_2 respectively. Observations are generated on a grid of $p = 32$ points with six different sample sizes $N = 5, 10, 20, 30, 40, 50$. Each permutation test is performed with $M = 1000$ and the test is repeated for 200 samples, so that we can evaluate the power of the test for different values of sample size and different degrees of violation of the null hypothesis. Fig. 1 shows the covariance function Σ_1 for the first group in all the simulations (where this covariance was obtained from the male curves within the Berkeley growth curve dataset (Ramsay and Silverman, 2005)).

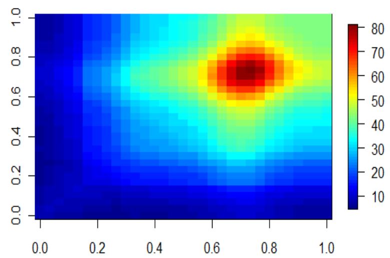

Figure 1: Integral kernel of the true covariance function Σ_1 for reference group.

First simulation: We consider the case in which the first two eigenvalues of Σ_2 are a convex combination of the first two eigenvalues of Σ_1:

$$\begin{aligned} \lambda_1^2 &= \gamma\lambda_2^1 + (1-\gamma)\lambda_1^1 \\ \lambda_2^2 &= \gamma\lambda_1^1 + (1-\gamma)\lambda_2^1 \end{aligned} \tag{3}$$

Figure 2 shows the estimated power for different values of γ and N.

It is worth mentioning that, despite the normality of the data, traditional parametric tests for comparison of covariances can be applied only when $N > p$, i.e. $N = 40, 50$. Indeed, the power is low also in these cases, since the sample size is small with respect to dimension $p = 32$. Note that the square root and Procrustes tests are the most powerful here, and all tests have the correct type I error probability.

Second simulation: We consider now a difference in the total variation between the covariance operators in the two groups, so that

$$\Sigma_2 = (1+\gamma)\Sigma_1.$$

In this case we can also compare the proposed method with the generalisation of the Levene test proposed in Anderson (2006), since this is a procedure to test for differences in multivariate dispersion. In particular we present results for the permutation version of this test to give a fair comparison with our method, even if the simulated data are indeed Gaussian. On the other hand, the spectral distance is not suitable, since it cannot discern between an operator and one of its multiples.

Fig. 3 shows the estimated power for different values of γ and N. For smaller N the Levene test has a high type I error. For larger N and γ, the Procrustes and square root tests are a little more powerful than the kernel L^2 and Levene tests.

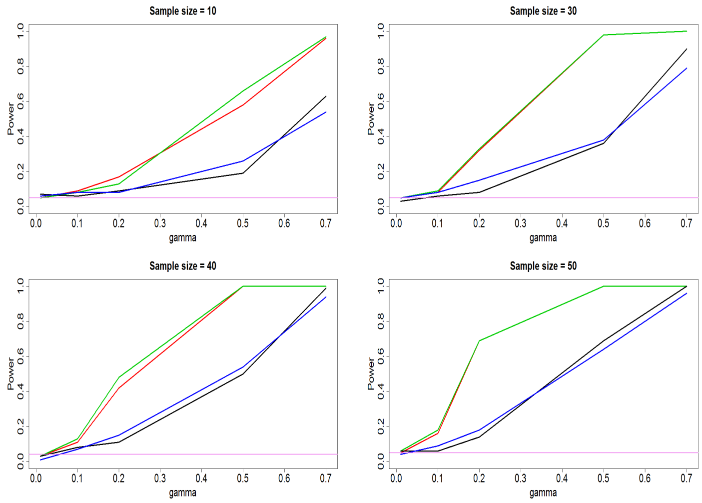

Figure 2: Power estimated via simulation for different values of γ and N obtained with Procrustes size-and-shape distance (red line), Square root distance (green line), Kernel $L^2(\Omega \times \Omega)$ distance (black line) and Spectral distance (blue line). The purple line shows the significance level $\alpha = 0.05$.

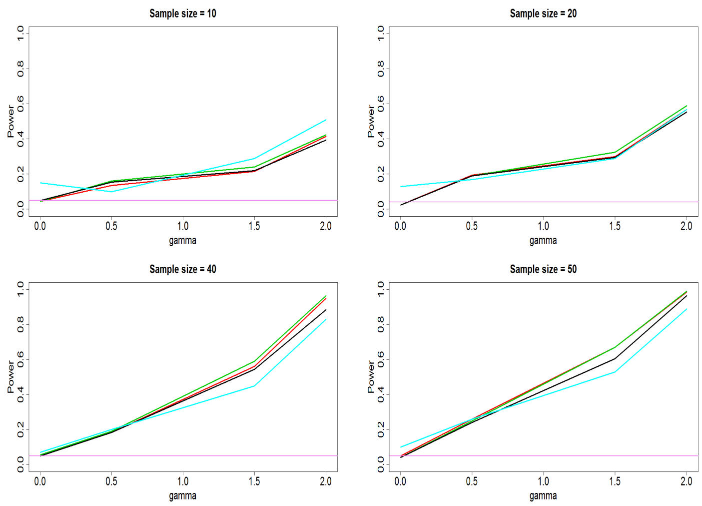

Figure 3: Power estimated via simulation for different values of γ and N obtained with Procrustes size-and-shape distance (red line), Square root distance (green line), Kernel $L^2(\Omega \times \Omega)$ distance (black line) and generalised Levene test (cyan line). The purple line shows the significance level $\alpha = 0.05$.

5 Applications

5.1 Data from AneuRisk Project

We illustrate here a possible application of the inferential technique described above. We consider data that have been collected within the AneuRisk Project, designed to investigate the role of vessel morphology and blood fluid dynamics on the pathogenesis of cerebral aneurysm (http://mox.polimi.it/it/progetti/aneurisk). A detailed description of the problem can be found in Sangalli et al. (2009a).

The AneuRisk data set is based on a set of three-dimensional angiographic images taken from 65 subjects, hospitalised at Niguarda Ca Granda Hospital (Milan) from September 2002 to October 2005, who were suspected of being affected by cerebral aneurysms. Out of these 65 subjects, 33 subjects have an aneurysm at or after the terminal bifurcation of the Internal Carotid Artery (ICA) (Upper group), 25 subjects have an aneurysm along the ICA (Lower group), and 7 subjects have not had any aneurysm (No-aneurysm group). In general, Upper group subjects are those with the most dangerous aneurysms; for this reason and for qualitative consideration, statistical analysis conducted in Sangalli et al. (2009a) joins the Lower and No-aneurysm groups in a single group, to be contrasted to the Upper group. Here we want to explore possible differences between Lower and No-aneurysm groups looking at vessel radius and curvature and their covariance operators.

Starting from the angiographies of the 65 patients, estimates of vessel radius and curvature are obtained with the procedure described in Sangalli et al. (2009b), resulting in a free-knots regression splines reconstruction of radius and curvatures. Each patient is therefore described by a pair of functions $R(s)$ and $C(s)$, where the abscissa parameter s measures an approximate distance along the ICA, from its terminal bifurcation toward the heart (for conventional reasons, this abscissa parameter takes a negative value to highlight that the direction is opposite with respect to blood flow). These curves are defined on different intervals, thus we restrict our analysis to the region which is common to all curves (i.e., for values of abscissa between -25 and -1). Fig. 4 shows radius and curvature for the two groups, while their covariance operators can be seen in Fig. 5. We evaluate the kernels of the covariance operators on an equispaced grid of $p = 65$ points.

We want now to verify the equality of the two groups in terms of covariance structure, since a visual inspection of the covariance operators would seem to indicate differences. A permutation test for equality of radius covariance operators result in a p-value of 0.94 using Procrustes distance, 0.885 with Square root distance and 0.9 for kernel distance. P-values of permutation tests for equality of curvature covariance operators are 0.86 for Procrustes distance, 0.775 for Square root distance and 0.87 for kernel distance. Therefore, there is no statistical evidence for difference of covariance operators between the two groups. Thus, the decision taken in the original analysis to treat the curves as being from a single group is not rejected.

Fig. 6 shows covariance operators for the two groups used in Sangalli et al. (2009a), i.e. patients with aneurysm in the upper part of the artery and patients with aneurysm

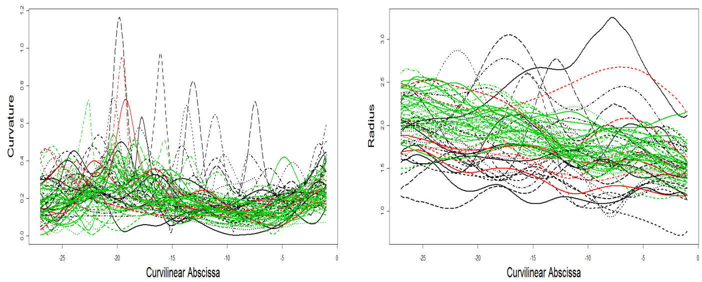

Figure 4: Curvature (left) and Radius(right) for the 65 patients in the range of abscissa common to all curves. Black coloured curves correspond to patients of the Lower group, red coloured curves to patients of the No-aneurysm group and green coloured curves to patients of the Upper group.

in lower part of the artery or no aneurysm at all. If we perform the permutation test on radius covariance operators for these two groups, we find p-values less than 0.0001 for Procrustes distance, Square root distance and kernel distance. For the curvature covariance operators, we obtain p-values of 0.1 for Procrustes distance, 0.02 for Square root distance and 0.02 for kernel distance. Thus, all these distances provide far smaller p-values than in the previous case indicating that the difference between these two groups is worth investigation. However, the evidence is somewhat weak for curvature, as the Procrustes distance is only significant at $p \leq 0.1$, with this distance being free from the arbitrary choice of decomposition.

5.2 Exploring relationships among Romance languages

The traditional way of exploring relationships across languages consists of examining textual similarity. However, this neglects phonetic characteristics of the languages. Here a novel approach is proposed to compare languages on the basis of phonetic structure.

People speaking different Romance languages (French, Italian, Portuguese, Iberian Spanish and American Spanish) are registered while pronouncing words in each language. The output of the registration for each word and for each speaker consists of the intensity of the sound over time and frequencies. The aim is to use this data to explore linguistic hypotheses concerning the relationships between different languages. However, while the temporal aspects of each individual word are important, we will concentrate on frequencies. Previous studies (Aston et al., 2010; Hadjipantelis et al., 2012) have indicated that covariance operators characterise languages well, and these will be the object of study. The operators summarise phonetic information for the language, while disregarding characteristics of singular speakers and words. For the scope of this work, we focus on the covariance operators among frequencies in the log-spectrogram, estimated by putting together all speakers of the language in the data set. The spectrogram is a two dimensional time-frequency image which gives localised time and frequency information across the word. We consider different time points as replicates of the same covariance operator among frequencies. It is clear that this is a significant simplification of the rich structure in the data but in itself can already lead to some interesting conclusions.

Let $f_{ijk}(t) \in L^2(\Omega)$ be a realisation of a random process, where $i = 1, \dots, n$ are different families of curves (i.e. different words), $j = 1, \dots, g$ the groups (i.e. languages) and $k = 1, \dots, K$ the observations (individual speakers). Let $S_{ij}(s,t)$ be the sample covariance operator for each family i in each group j. As mentioned above, the working hypothesis is that the significant information of the different groups are in the family-wise covariances S_{ij} rather than in the individual observations f_{ijk}. Thus, we can also consider S_{ij} as realisation of the random variables Σ_j, taking value in the space of trace class non negative definite self-adjoint operators. We want to explore similarities and differences among groups on the basis of their covariance operators Σ_j.

A pairwise comparison of the groups may be for example performed using as test statistic as the distance between the sample Fréchet mean for each group,

$$T_{jj'} = d(\widehat{\Sigma}_j, \widehat{\Sigma}_{j'}),$$

where the Fréchet means $\widehat{\Sigma}_j$ and $\widehat{\Sigma}_{j'}$ are obtained from the samples $S_{1j}, \dots, S_{n_j j}$ and $S_{1j'}, \dots, S_{n_{j'} j'}$ by minimizing the appropriate distance $d(.,.)$, being therefore estimates of the unknown group covariance operators Σ_j and $\Sigma_{j'}$. As in the previous section, a permutation test can be set up randomizing the assignment of the sample covariance operators to the two groups.

Here some preliminary results are reported, focusing on the covariance operator for the word “one” spoken across the different languages by a total of 23 speakers across the five languages. This word is similar in each language (coming from the common latin root), but different enough to highlight changes across the languages (American Spanish: uno; French: un; Iberian Spanish: uno; Italian: uno; Portuguese: um). This also highlights that in this case, textual analysis is difficult as three of the languages have the same textual representation.

Fig. 7 shows the covariance operator estimated for each language via Fréchet averaging along time, using the square root distance. Fig. 8 shows dissimilarity matrix among average covariance operators for each language and the corresponding dendrogram. Indeed, it seems that focusing on the covariance operator allows the capture of some significant information about languages. Relationships among covariance operators have features which are expected by linguistic hypotheses, such as strong similarity between French and Italian and between American Spanish and Iberian Spanish or the fact that Portuguese is far from the pair Italian–French. However, not all our conclusions directly support textual analysis. The distance of Portuguese from both Spanish languages is greater than expected. Moreover, for historical reasons American Spanish is expected to be nearer than Iberian Spanish to French and Italian, but the covariance structures indicate this is reversed. Thus, as this analysis is currently based on the word “one”, providing further assessment using a much larger corpus will be of significant interest, and is the subject of ongoing work.

Extrapolation of covariance operators

A particularly interesting application of the analysis is to provide insight into the change of the frequency structure along the path of language evolution. This would be inherently linked to extrapolation based on the distances we have proposed. A path can be defined, for example, as the shortest path connecting the languages, using as distance the distance between frequency covariance operators. All the distances proposed in the paper lead to the same path,

$$\text{Italian} \leftrightarrow \text{French} \leftrightarrow \text{IberianSpanish} \leftrightarrow \text{AmericanSpanish} \leftrightarrow \text{Portuguese}.$$

Here, we want to compare the frequency covariance structure of a language to those obtained by extrapolating covariance operators of "previous" languages in the evolutionary path. As was seen in Fig 7, the Portuguese language presents a very different covariance structure with respect to the other Romance languages. Thus, it would be of interest to compare its frequency covariance operator with the one extrapolated from the covariances of the two Spanish languages, to see if this kind of covariance was expected (and a linear model of distance appropriate).

Initially, one idea could be to do extrapolation based on kernels,

$$s(s,t)(x) = \frac{1}{x}(s_{SA}(s,t) + x(s_{SA}(s,t) - s_{SI}(s,t))), \tag{4}$$

choosing as x the kernel distance between Portuguese language and Spanish American. However, just as the space of positive definite matrices is not Euclidean, extrapolation based on kernel distances does not result in a valid kernel for a covariance operator (i.e. the associated integral operator is not non-negative definite, see Fig 9). If we apply the extrapolation procedure in the space of square root operators, the problem of negative eigenvalues is solved with the inverse (square) transformation, which makes all the eigenvalues positive. While extrapolating to large negative values and then squaring them would likely introduce additional variation with somewhat questionable meaning, this procedure, at least, provides a valid covariance operator.

Thus, using the square root mapping, we extrapolate the square root covariance operator

$$S(x) = \frac{1}{x}\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}})\}^*\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}})\}, \tag{5}$$

where x can be chosen as $||(S_P)^{\frac{1}{2}} - (S_{SA})^{\frac{1}{2}}||_{HS}$. Fig. 9 shows the integral kernel of the obtained covariance operator. It can be seen that the high variability of Portuguese in the high frequency region is somehow expected from the extrapolation of the Spanish languages covariances.

Similar results can be obtained using a Procrustes approach on the mapped space, i.e. aligning $(S_{SA})^{\frac{1}{2}}$ and $(S_{SI})^{\frac{1}{2}}$ through the optimal unitary operator $\widetilde{R}$ which minimises the square Procrustes distance and then extrapolating:

$$S(x) = \frac{1}{x}\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}}\widetilde{R})\}\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}}\widetilde{R})\}^*, \tag{6}$$

where x can be chosen as $d_P(S_{SA}, S_P)$. The Procrustes alignment makes the two operators nearer in the space of Hilbert-Schmidt operators. Thus, as we expect, the extrapolation is more stable (as the arbitrary nature of the decomposition is removed) and to yield less artifacts, in this case large positive eigenvalues resulting from squared negative eigenvalues. Fig. 9 shows the extrapolated covariance operator for Portuguese for the three methods proposed and indeed the Procrustes estimate seems to provide a better result with somewhat smaller variance in places due to the Procrustes rotation.

However, further rigorous development of this idea would require a more complete understanding of the underlying spaces induced by the linearisation of the path and this would provide considerable scope for future statistical development.

6 Conclusions and further development

In this work the problem of dealing with covariance operators has been investigated. The choice of the appropriate metric is crucial in the analysis of covariance operators, and as such some suitable metrics have been proposed and their properties have been highlighted. In particular two metrics have been applied to infinite dimensional covariance operators: the square root operator distance and the Procrustes size-and-shape distance. Both these metrics rely on the mapping of the covariance operators to a suitable space of Hilbert-Schmidt operators where a linear distance, namely the Hilbert-Schmidt norm of the difference of the operators, is appropriate. The square root operator distance uses the square root map defined in (1), while the Procrustes distance allows for unitary transformations in the space of Hilbert-Schmidt operators, thus taking into account the arbitrariness of the representation.

On the basis of an appropriate metric, statistical methods can be developed to deal with covariance operators in a functional data analysis framework. Here the notable cases of estimating the average from a sample of covariance operators and testing the equality of covariance structure between two groups are illustrated. The latter technique has proved useful for the analysis of the AneuRisk data, where investigating the covariance structures of different groups supports the results of previously published analysis. Moreover, in some applications the covariance operator itself is the object of interest, as shown with the linguistic data of Section 5.2. Using the square root and Procrustes distances between covariance operators of frequencies, some significant phonetic features of Romance languages have been found.

Many other developments are of course possible, for both the theoretical aspects and the practice of the proposed statistical methods. In particular, there is considerable scope for development of the consistency properties for estimators proposed in Section 4 and of necessary conditions for (local) convergence of the Procrustes algorithm in infinite dimensions, as well as a rigorous analysis of the proposed extrapolation procedure. The analysis of linguistic data is only preliminary and more significant results are expected when many words are taken simultaneously into account.

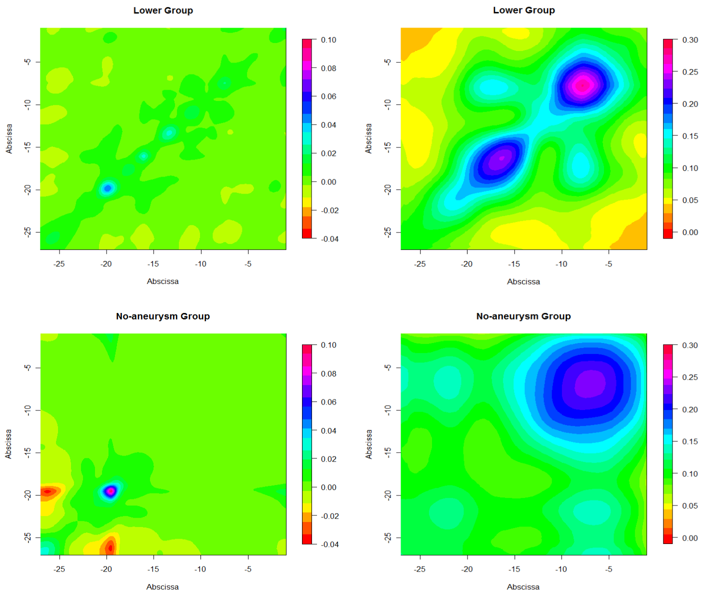

Figure 5: Covariance operator for curvature (left) and radius (right) for the for the Lower (first row) and No aneurysm (second row) groups.

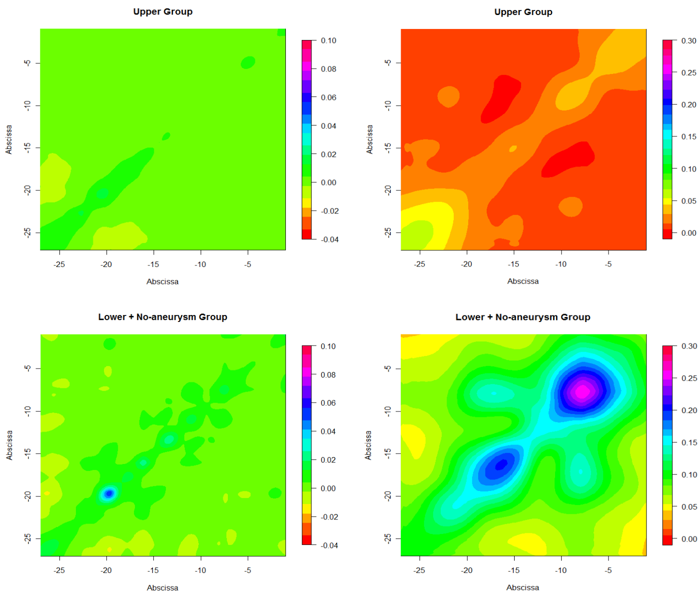

Figure 6: Covariance operator for curvature (left) and radius (right) for the Upper (first row) and Lower or No aneurysm (second row) groups.

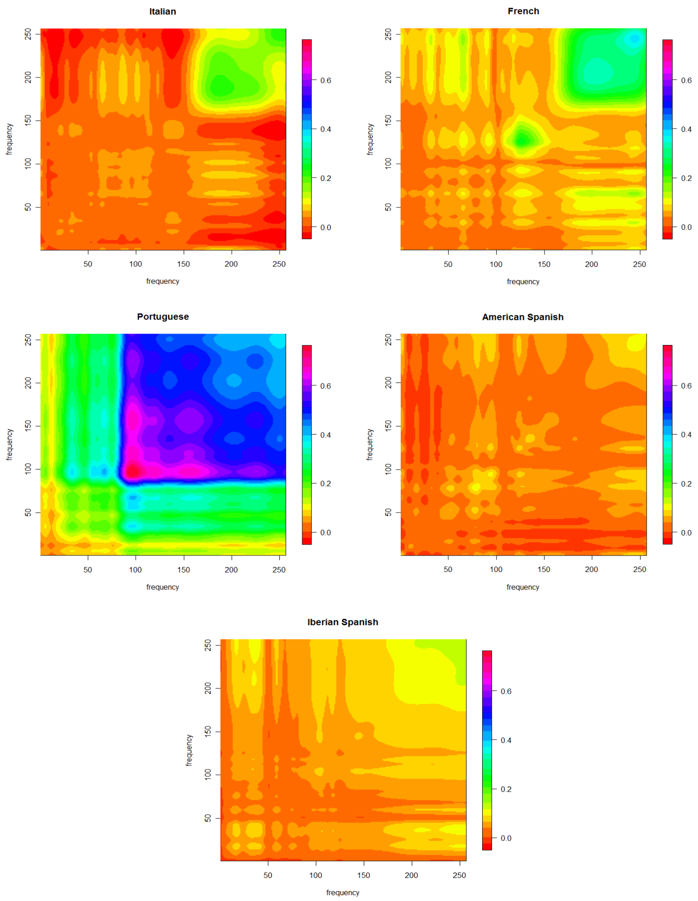

Figure 7: Fréchet average along time of covariance operators of log-spectrogram among frequencies for five romance languages.

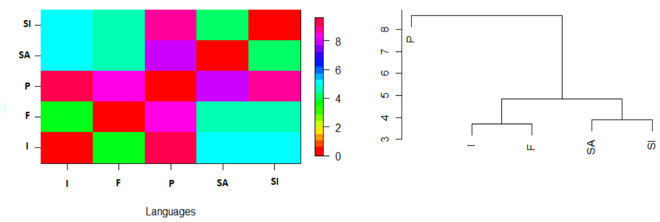

Figure 8: Distance matrix among Fréchet average above, obtained with Square root distance, along with a corresponding dendrogram indicating possible linguistic relationships.

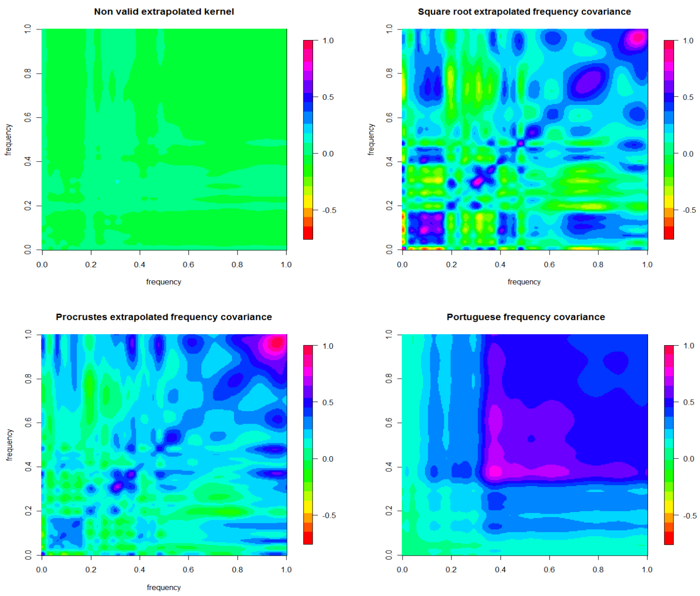

Figure 9: First row: Kernel extrapolated for Portuguese form the two Spanish languages kernels using equation (4) (left). It is not a valid kernel for covariance operators, since the associated integral operator is not non negative definite. Covariance operator extrapolated for Portuguese from the two Spanish languages with the square root mapping of equation (5) (right). Second row: Covariance operator extrapolated for Portuguese from the two Spanish languages with the Procrustes alignment in equation (6) (left) and sample frequency covariance for Portuguese language (right).

Acknowledgements

The authors are very much indebted to Prof. John Coleman for providing the Romance language data and to Pantelis Hadjipantelis for preprocessing it.

References

Anderson, M.J. (2006), "Distance-Based Tests for Homogeneity of Multivariate Dispersions", Biometrics **62**, 245-253.

Aston, J. A. D., Chiou, J.-M. and Evans, J. P. (2010), "Linguistic Pitch Analysis using Functional Principal Component Mixed Effect Models". *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **59**: 297-317.

Baker, C. R. (1970), "Mutual Information for Gaussian Processes", *SIAM Journal on Applied Mathematics*, **19**, 451–458.

Bosq, D. (2000), Linear processes in function spaces, Springer, New York.

Dryden, I. L. and Mardia, K. V. (1998), Statistical Shape Analysis. Wiley, Chichester.

Dryden, I.L., Koloydenko, A., Zhou, D. (2009), "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging", *The Annals of Applied Statistics*, **3**, 1102-1123.

El Karoui, N. (2008), "Operator norm consistent estimation of large-dimensional sparse covariance matrices", *The Annals of Statistics*, **36**, 2717–2756.

Ferraty, F. and Vieu, P. (2006), Nonparametric functional data analysis: theory and practice, Springer Series in Statistics, Springer.

Fremdt, S., Steinebach, J.G., Horvth, L. and Kokoszka, P. (2012), "Testing the Equality of Covariance Operators in Functional Samples", *Scandinavian Journal of Statistics*, in press.

Good, P.I. (2005), Permutation, Parametric and Bootstrap Tests of Hypotheses, Third Edition. Springer, New York.

Groisser, D. (2005), "On the convergence of some Procrustean averaging algorithms", *Stochastics: An International Journal of Probability and Stochastic Processes*, **77**, 31-60.

Gower, J. C. (1975), "Generalized Procrustes analysis", *Psychometrika*, **40**,33–50.

Hadjipantelis, P. Z., Aston, J. A. D., and Evans, J. P. (2012), "Characterizing fundamental frequency in Mandarin: A functional principal component approach utilizing mixed effect models", *Journal of the Acoustical Society of America*, in press.

Karcher, H. (1977), "Riemannian center of mass and mollifier smoothing", *Communications on Pure and Applied Mathematics*, **30**, 509–541.

Kendall, W.S. (1990), "Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence", *Proceedings of the London Mathematical Society*, **61**, 371–406.

Le, H. (2001), "Locating Frchet means with application to shape spaces", *Advances in Applied Probability*, **33**, 324-338.

Panaretos, V. M., Kraus, D. and Maddocks, J.H. (2010). "Second-Order Comparison of Gaussian Random Functions and the Geometry of DNA Minicircles.", *Journal of the American Statistical Association (Theory and Methods)*, **105**: 670-682.

Ramsay, J.O. and Silverman, B.W. (2005), Functional data analysis (2nd ed.), Springer, New York.

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009a), "A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery", *Journal of American Statistical Association*, **104**, 37-48.

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009b), "Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines", *Journal of the Royal Statistical Society Ser. C, Applied Statistics*, **58**, 3, 285-306.

Ten Berge, J. M. F. (1977), "Orthogonal Procrustes rotation for two or more matrices",*Psychometrika*, **42**, 267–276.

Zhu, K. (2007), Operator theory in function spaces (2nd ed.), American Mathematical Society.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

35/2012 Pigoli, D.; Aston, J.A.D.; Dryden, I.L.; Secchi, P.
Distances and Inference for Covariance Functions

34/2012 Menafoglio, A.; Dalla Rosa, M.; Secchi, P.
A Universal Kriging predictor for spatially dependent functional data of a Hilbert Space

33/2012 Motamed, M.; Nobile, F.; Tempone, R.
Analysis and computation of the elastic wave equation with random coefficients

32/2012 Formaggia, L.; Fumagalli, A.; Scotti A.; Ruffo, P
A reduced model for Darcy s problem in networks of fractures

31/2012 Bonizzoni, F.; Buffa, A; Nobile, F:
Moment equations for the mixed formulation of the Hodge Laplacian with stochastic data

30/2012 Beck, J.; Nobile, F.; Tamellini, L.; Tempone, R.;
Convergence of quasi-optimal Stochastic Galerkin Methods for a class of PDES with random coefficients

29/2012 Chen, P.; Quarteroni, A.; Rozza, G.
Stochastic Optimal Robin Boundary Control Problems of Advection-Dominated Elliptic Equations

28/2012 Canuto, C.; Nochetto, R.H.; Verani, M.
Contraction and optimality properties of adaptive Legendre-Galerkin methods: the 1-dimensional case

27/2012 Pigoli, D.; Secchi,P.
Estimation of the mean for spatially dependent data belonging to a Riemannian manifold

26/2012 Tamellini, L.; Le Maitre, O.; Nouy, A.
Model reduction based on Proper Generalized Decomposition for the Stochastic steady incompressible Navier-Stokes equations