Statistics on functional data and covariance operators in linear inverse problems

Eardi Lila, Simon Arridge, John A. D. Aston

1Cambridge Centre for Analysis, University of Cambridge
2Centre for Medical Image Computing, University College London
3Statistical Laboratory, DPMMS, University of Cambridge

Abstract

We introduce a framework for the statistical analysis of functional data in a setting where these objects cannot be fully observed, but only indirect and noisy measurements are available, namely an inverse problem setting. The proposed methodology can be applied either to the analysis of indirectly observed functional data or to the associated covariance operators, representing second-order information, and thus lying on a non-Euclidean space. To deal with the ill-posedness of the inverse problem, we exploit the spatial structure of the sample data by introducing a flexible regularizing term embedded in the model. Thanks to its efficiency, the proposed model is applied to MEG data, leading to a novel statistical approach to the investigation of functional connectivity.

1 Introduction

An inverse problem is the process of recovering missing information from indirect and noisy observations. Not surprisingly, inverse problems play a central role in numerous fields such as, to name a few, geophysics (Zhdanov, 2002), computer vision (Hartley and Zisserman, 2003), medical imaging (Arridge, 1999; Lustig et al., 2008) and machine learning (Vito et al., 2005).

Solving a linear inverse problem means finding an unknown $x$, for instance a function or a surface, from a noisy observation $y$, which is a solution to the model

$$y = Kx + \varepsilon,$$

where $y$ and $\varepsilon$ belong to an either finite or infinite dimensional Banach space. The map $K$ is assumed to be known and is usually referred to as the forward operator. The term $\varepsilon$ represents observational error.
Problem 1 is a well-studied problem within applied mathematics. Its main difficulties arise from the fact that, in practical situations, an inverse of the forward operator does not exist, or if it does, it amplifies the noise term. For this reason such a problem is called ill-posed. Consequently, the problem of estimating \( x \) in (1) is generally tackled by minimizing a functional which is the sum of a data (fidelity) term and a regularizing term encoding prior information on the function to be recovered. Alternatively, a Bayesian approach could be adopted (Stuart, 2010). Problem 1 could be regarded as a generalization of the smoothing problem, more common in the statistical literature, where the aim is reconstructing an underlying smooth function from noisy evaluations (see e.g. Ramsay, 2002; Wood et al., 2008; Ettinger et al., 2016).

In this work, we look at the inverse problem from a Functional Data Analysis (FDA) (Ramsay and Silverman, 2005) prospective. Therefore, we establish a framework for performing statistical analysis on indirectly observed data samples \( x_1, \ldots, x_n \), consisting of random functions or surfaces, where each function represents one sample element. The indirect observations are assumed to be generated by the model

\[
y_i = K_i x_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2}
\]

with \( K_1, \ldots, K_n \) a collection of sample specific known forward operators.

However, in many statistical applications, it is of interest to perform statistics on the second order information associated to the functional samples. This is, for instance, the case in a number of areas of neuroimaging, particularly those investigating functional connectivity. Thus, despite Problem 2 being itself of interest, it can also be seen as a simple case of a more complex framework upon which we can build a model for the second order information associated to the functional observations.

Specifically, we consider the problem of performing statistical analysis on the indirectly observed samples \( C_1, \ldots, C_n \), that are covariance operators, expressing the second order properties of the underlying functional data. The indirect observations are covariance operators generated by the model

\[
S_i = K_i \circ C_i \circ K_i^* + \varepsilon_i, \quad i = 1, \ldots, n, \tag{3}
\]

where \( K_i^* \) denotes the adjoint operator and the term \( \varepsilon_i \) models observational error. The term \( K_i \circ C_i \circ K_i^* \) represents the covariance operator of \( K_i X^{(i)} \), with \( X^{(i)} \) an underlying random function whose covariance operator is \( C_i \).

Problem 2 has been classically dealt with by reconstructing each observation independently, or in other words, the underlying statistical model of the data is ignored, and such a problem is formulated as \( n \) separate Problem 1s. However, such an approach can be sub-optimal in particular in a large noise setting, as when estimating one signal, the information from all the other sampled signals is systematically ignored. Problem 3 introduces the additional difficulty that \( \{C_i\} \) and \( \{S_i\} \) live on non-Euclidean spaces. It is moreover not obvious how to introduce a regularization term on the covariance operators \( \{C_i\} \) reflecting, for instance, smoothness assumptions on the underlying functional data.

We tackle the inverse problems introduced here by generalizing the concept of functional Principal Component Analysis (fPCA) to indirectly observed functional samples and covariance operators.
1.1 Motivating application - functional connectivity

In recent years, statistical analysis of covariance matrices has gained a predominant role in medical imaging and in particular in functional neuroimaging. In fact, covariance matrices are the natural objects to represent the brain’s functional connectivity, which can be defined as a measure of covariation, in time, of the cerebral activity among brain regions. While many techniques have been proposed to describe functional connectivity, almost all can be described in terms of a function of a covariance or related matrix.

![Figure 1: On the top left, head model of a subject and superimposition of the 248 MEG sensors positioned around the head, called ‘sensors space’. On the top right, brain model of the same subject represented by a triangular mesh of 8K nodes, which represents the ‘brain space’. On the bottom left, an example of a synthetic signal detected by the MEG sensors. The dots represent the sensors, the color map represents the signal detected by the sensors. On the bottom right, intensity of the reconstructed signal on the triangular mesh of the cerebral cortex.](image)

Covariance matrices representing functional connectivity can be computed from the signals arising from functional imaging modalities. The choice of a specific functional imaging modality is generally driven by the preference to have high spatial resolution signals, and thus high spatial resolution covariance matrices, versus high temporal resolution, and thus the possibility to study the temporal dynamic of the covariance matrices. Functional Magnetic Resonance falls in the first category, while Electroencephalogram (EEG) and Magnetoencephalography (MEG) in the second. However, high temporal resolution does generally come at the price of indirect measurements. In fact, as shown in Figure 1 for the case of MEG data, the signals are in practice detected on the sensors space. It is however of interest to produce results on the associated signals on the cerebral...
cortex, which we will refer to as brain space. The signals on the brain space are functional data whose domain is the geometric representation of the brain and are associated with the neuronal activity on the cerebral cortex. We borrow here the notion of brain space and sensors space from Johnstone and Silverman (1990) and we use it throughout the paper for convenience, however it is important to highlight that the formulation of the problem is much more general than the setting of this specific application.

The signals on the brain space are related to the signals on the sensors space by a forward operator, derived from the physical modeling of the electrical/magnetic propagation, from the cerebral cortex to the sensors. This is generally referred to as the forward problem. For methods like MEG, the forward operator is defined through the solution to a partial differential equation of diffusion type. Such a mapping induces a strong degree of smoothing and consequently the corresponding inverse problem, i.e. the reconstruction of a signal on the brain space from observations in the sensor space, is strongly ill-posed. In fact, signals with fairly different intensities on the brain space, due to the diffusion effect, result in signals with similar intensities in the sensors space. In Figure 1, we show an example of a signal on the brain space and the associated signal on the sensors space.

From a practical perspective, it is crucial to understand how the different parts of the brain interact, which is sometimes known as functional connectivity. A possible way to understand these interactions is by analyzing the covariance function associated to the signals generated from the cerebral activity of an individual on the brain space. More recently, the interest has shifted from this static approach to a dynamic approach. In particular, for a single individual, it is of interest to understand how these covariance functions vary in time. This is a particularly active field, known as dynamic functional connectivity (Hutchison et al., 2013). Another element of interest is understanding how these covariance functions vary among individuals. We believe that FDA should have a central role in addressing these questions, and tackle the problem as it is, without having to compromise on its complexity.

The remainder of this paper is organized as follows. In Section 2 we give a formal description of the problem. We introduce the models associated to Problem 2 and 3 in...
Section 3 and 4 respectively. In Section 5, we perform simulations to access the validity of the estimation framework. In Section 6 we apply the proposed models to the MEG data and we finally give some concluding remarks in Section 7.

2 Mathematical description of the problem

The statistical analysis of data samples that are random functions or surfaces has been well explored in the FDA literature, however, most of those works focus on the setting of fully observed functions. An exception to this is the sparse FDA literature (see e.g. Yao et al., 2005), where instead the functional samples are assumed to be observable only through irregular and noisy evaluations. The work presented here could be seen as a generalization of the sparse FDA setting. We frame the problem in the context of inverse problems and make explicit the links with the inverse problem literature. Often, it will be possible to draw analogies between specific sub-cases of the methodology proposed here and that proposed in the sparse FDA literature.

We now introduce the problem using our driving application as an example. To this purpose, let $M$ be a closed smooth two-dimensional manifold embedded in $\mathbb{R}^3$, which in our application represents the geometry of the cerebral cortex. An example of such a surface is shown on the top right of Figure 1. We denote with $L^2_p(M)$ the space of square integrable functions on $M$. Define $X$ to be a random function with values in a Hilbert functional space $F \subset L^2_p(M)$ with mean $\mu = \mathbb{E}[X]$, finite second moment, and assume the square integrability of its covariance function $C_X (v, v') = \mathbb{E}[(X(v) - \mu(v))(X(v') - \mu(v'))]$.

The associated covariance operator $C_X$ is defined as $C_X g = \int_M C_X (v, v') g(v) dv$, for all $g \in L^2(M)$. Mercer’s Lemma (Riesz and Sz.-Nagy, 1955) guarantees the existence of a non-increasing sequence $\{\gamma_r\}$ of eigenvalues of $C_X$ and an orthonormal sequence of corresponding eigenfunctions $\{\psi_r\}$, such that

$$C_X (v, v') = \sum_{r=1}^{\infty} \gamma_r \psi_r(v) \psi_r(v'), \quad \forall v, v' \in M. \quad (4)$$

As a direct consequence, $X$ can be expanded as $X = \mu + \sum_{r=1}^{\infty} \zeta_r \psi_r$, where the random variables $\{\zeta_r\}$ are uncorrelated and are given by $\zeta_r = \int_M (X(v) - \mu(v)) \psi_r(v) dv$. The collection $\{\psi_r\}$ defines the strongest modes of variation of the random function $X$ and these are called Principal Component (PC) functions. The associated random variables $\{\zeta_r\}$ are called PC scores. Moreover, the defined PC functions are the best finite basis approximation. In fact, for any fixed $M \in \mathbb{N}$, the first $M$ PC functions of $X$ satisfy

$$\langle \psi_i \rangle_{i=1}^{M} = \arg \min_{\{\phi_i\}_{i=1}^{M}} \mathbb{E} \int_M \left\{ X - \mu - \sum_{m=1}^{M} \langle X - \mu, \phi_m \rangle \phi_m \right\}^2, \quad (5)$$

where $\delta_{ml}$ is the Kronecker delta; i.e. $\delta_{ml} = 1$ for $m = l$ and 0 otherwise.
2.1 Principal components of indirectly observed functions

Suppose now that the signals on the sensors space are detected through $s$ sensors. Let \( \{K_i\} \) be a collection of $p \times s$ real matrices, representing the subject specific forward operators relating the signal at $p$ pre-defined points \( \{v_j : j = 1, \ldots, p\} \) on the cortical surface $\mathcal{M}$ with the signal captured by the $s$ sensors. Moreover, define the evaluation operator $\Psi : \mathcal{F} \rightarrow \mathbb{R}^p$ to be a vector-valued functional that evaluates a function $f \in \mathcal{F}$ at the $p$ pre-specified points \( \{v_j \} \subset \mathcal{M} \), returning the $p$ dimensional vector \( (f(v_1), \ldots, f(v_p))^T \).

The operators $\Psi$ and \( \{K_i\} \) are known. However, in the described problem the random function $X$ can be observed only through indirect measurements \( \{y_i \in \mathbb{R}^s : i = 1, \ldots, n\} \) generated from the model

\[
\begin{cases}
  x_i = \mu + \sum_{r=1}^{\infty} \zeta_{i,r} \psi_r \\
y_i = K_i \Psi x_i + \varepsilon_i, \quad i = 1, \ldots, n
\end{cases}
\]

where \( \{x_i\} \) are $n$ independent realizations of $X$, and thus expandible in terms of the PC functions \( \{\psi_r\} \) and the coefficients \( \{\zeta_{i,r}\} \) given by \( \zeta_{i,r} = \int_{\mathcal{M}} (x_i(v) - \mu(v)) \psi_r(v) dv \). The terms \( \{\varepsilon_i\} \) represent observational errors drawn independently from an $s$-dimensional normal random vector, with mean the zero vector and variance $\sigma^2 I_p$, where $I_p$ denotes the $p$-dimensional identity matrix.

Model 6 represents an implementation of the idealized Problem 2. In Figure 3 we give an illustration of the introduced setting. Note that it would not be necessary to define the evaluation operator if the forward operators were defined to be functionals $\{K_i : \mathcal{F} \rightarrow \mathbb{R}^p\}$, relating directly the functional objects on the brain space to the real vectors on the sensors space. It is however the case that the operators $\{K_i\}$ are computed in a matrix form by third part software (see Section 6 for details) for a pre-specified set of points \( \{v_j \} \subset \mathcal{M} \) and it thus convenient to take this into account in the model though the introduction of an evaluation operator $\Psi$.

Here, we consider the problem of estimating the PC functions \( \{\psi_r\} \), or equivalently the eigenfunctions of the covariance operator $C_X$, from the observations \( \{y_i\} \). As already mentioned, in neuroimaging studies, this is often an important task as $C_X$ describes the static functional connectivity of the brain.

The problem of estimating the PC functions \( \{\psi_r\} \) is generally tackled in two steps. In the first step, estimates \( \{\hat{x}_i\} \) of the functions \( \{x_i\} \) are individually computed from the vectors \( \{y_i\} \). In the second step, the covariance function $C_X$ is estimated from \( \{\hat{x}_i\} \) by use of classical estimators, and the associated PCs computed by spectral decomposition of the estimated covariance operator. Being $s \ll p$ and due to the ill-posedness of the inverse problem, reconstructing \( \{x_i\} \) from \( \{y_i\} \) is not straightforward. This is a well known problem in the inverse problem literature and it is generally tackled through estimators of the type

\[
\hat{x}_i = \mathop{\text{arg inf}}_{f \in \mathcal{F}} \|y_i - K_i \Psi f\|^2 + \lambda \mathcal{P}(f), \quad i = 1, \ldots, n,
\]

where $\| \cdot \|$ denotes the Euclidean norm and $\mathcal{P} : \mathcal{F} \rightarrow \mathbb{R}^+$ is a penalty functional, e.g. $\mathcal{P}(f) = \|f\|^2_{L^2}$, the norm of the functional space $\mathcal{F}$. 

The functional $\mathcal{P}$ encodes prior information on the function to be estimated, while the data fidelity term ensures that the resulting estimated function $\hat{x}_i$ is such that $K_i \Psi \hat{x}_i$ is a good approximation to the signal $y_i$ actually detected. The parameter $\lambda$ is chosen to optimally weight the two terms, and many data-driven options are available for this purpose, as for instance, cross-validation or the L-curve method (see, e.g., Vogel, 2002). Typical choices for $\mathcal{P}$ are Sobolev (semi-) norms, which encode smoothness, or the total variation norm, which allows discontinuity but penalizes for excessively oscillating functions. Also, more complex penalty terms could be considered, for instance, by adding terms that encourage the reconstruction to be sparse.

The inverse problem in (7), for $K_i = I_p$, a $p \times p$ identity matrix, reduces to a smoothing problem, for functions on a non-Euclidean domain. Such models are more common in the statistical literature. For instance, smoothing methods that can handle functional data whose domain is a subset of the two-dimensional Euclidean space, dealing with complex boundaries have been proposed in Ramsay (2002); Wood et al. (2008); Sangalli et al. (2013). An extension to functions whose domain is a non-Euclidean manifold has been proposed in Ettinger et al. (2016).

However, a two step estimation for the PC functions $\{\psi_r\}$ can be sub-optimal. The main reason is that in the first step the estimations are made individually for each signal $x_i$, and information from the other sampled signals is systematically ignored. In Section 3 we propose a model for the direct estimation of the PC functions $\{\psi_r\}$ from the the data $\{y_i\}$.

In the case of direct but noisy observations of a signal, previous works on statistical estimation of the covariance function, and associated eigenfunctions, have been made, for instance, in Bunea and Xiao (2015) for regularly sampled functions and in Yao et al. (2005) and Huang et al. (2008) for sparsely sampled functions. Further theoretical properties of the approach in Yao et al. (2005) are studied in Hall et al. (2006). Their approach consists on estimating the smooth covariance function by local least square regression.
Due to the fact that covariance matrices do empirically show a spiked structure along the diagonal, in Yao et al. (2005), two different smoothing steps are applied: one on the diagonal and one on the direction orthogonal to the diagonal. However, this second step implicitly exploits the ordering structure of 1D functional data along rows and columns of the 2D matrix representing the discretized and noisy version of the underlying covariance function. In higher dimensions this structure is inevitably lost. A generalization to functions whose domain is a manifold is proposed in Lila et al. (2016) and appropriate spatial coherence is introduced by penalizing directly the eigenfunctions of the covariance operator to be estimated. In the indirect observations setting, Tian et al. (2012) propose a separable model in time and space for source localization. However, the estimation of PC functions of functional data, from indirect and noisy samples has not been covered yet. Such a model is introduced in Section 3 and plays an important role in the definition of the PC model for covariance operators, whose setting is introduced in the next section.

2.2 Principal components of indirectly observed covariance functions

Suppose now we are given a set of \( n \) covariance functions \( \{C_i : i = 1, \ldots, n\} \), representing the associated covariance operators \( \{C_i : i = 1, \ldots, n\} \) on the brain space. In our driving application, each covariance function \( C_i : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R} \) describes the functional connectivity of the \( i \)th individual or the functional connectivity of the same individual at the \( i \)th time-point. Here we consider the problem of defining and estimating a set of PC covariance functions \( \{C_i\} \), which is a set of covariance functions that enable the description of \( \{C_i\} \) through the ‘linear combinations’ of few components. Such a reduced order description is of interest, for example, in understanding how functional connectivity varies among individuals or over time.

We define a model for the PC covariance functions of \( \{C_i\} \) from the set of indirectly observed covariance matrices, computed from the signal on the sensors space, and thus given by \( \{S_i \in \mathbb{R}^{s \times s}, i = 1, \ldots, n\} \) with

\[
S_i = K_i C_i K_i^T + E_i^T E_i, \quad i = 1, \ldots, n,
\]

(8)

where \( C_i = (C_i(v_j, v_l))_{j,l} \) and \( \{v_j : j = 1, \ldots, p\} \) are the sampling points associated to the operator \( \Psi \). The forward operators \( \{K_i\} \) act on both sides of the covariance functions \( \{C_i\} \), due to the linear transformation \( K_i \Psi \) applied to the signals on the brain space before being detected on the sensors space. The term \( E_i^T E_i \) is an error term, where \( E_i \) is a \( s \times s \) matrix such that each entry is an independent sample of a Gaussian distribution with mean zero and standard deviation \( \sigma \). Model (8) could be regarded as an implementation of the idealized Problem 3, where the covariance operators are represented by the associated covariance functions. An illustration of the setting introduced can be found in Figure 4.

Generally, PCs are defined and computed by seeking linear subspaces that maximize the variance of the data projected on it, or that analogously minimizes the distance of the projected data from the observed data. However, in the case of PCs on the space of covariance functions, a linear subspace, or part of it, is likely to fall outside the
non-Euclidean cone of positive semi-definite operators. In literature, this non-Euclidean structure is accounted for by introducing a proper distance in the space of covariance matrices (Dryden et al., 2009) or covariance operators (Pigoli et al., 2014). In particular, Dryden et al. (2009) introduce a PC model for directly observed covariance matrices, i.e. $K_i = I_p$. Such a model cannot deal with indirectly observed covariance matrices, we thus propose a novel approach for this problem in Section 4.

3 Principal components of indirectly observed functions

The aim of this section is to define a model for the estimation of the PC functions $\{\psi_r\}$ from the observations $\{y_i\}$, defined in (6).

3.1 Model

Let now $z = (z_1, \ldots, z_n)^T$ be a $n$-dimensional real column vector and $H^2(M)$ be the Sobolev space of functions in $L^2(M)$ with first and second distributional derivatives in $L^2(M)$. We propose to estimate $\hat{f} \in H^2(M)$, the first PC function of $X$, and the associated PC scores vector $z$, by solving the equation

$$\hat{z}, \hat{f} = \arg \min_{z \in \mathbb{R}^n, f \in H^2(M)} \sum_{i=1}^n \|y_i - z_i K_i \Psi f\|^2 + \lambda z^T \Delta f, \quad (9)$$

where the Laplace-Beltrami operator $\Delta$, integrated over the manifold $M$, enables a smoothing regularizing effect on the PC function $\hat{f}$, while the data fit term encourages $K_i \Psi f$ to capture the strongest mode of variation of $y_i$. The parameter $\lambda$ controls the trade-off between the data fit term of the objective function and the regularizing term. The second PC function can be estimated by classical deflation methods, i.e. by working
on the residuals \( \{y_i - z_i K_i \Psi \hat{f}\} \), and so on for the subsequent PCs. The proposed model can be interpreted as a regularized least square estimation of the first PC function \( \psi_1 \) in (6), with the terms \( \{z_i\} \) playing the role of estimates of the variables \( \{\zeta_{i1}\} \).

In the simplified case of a single forward operator \( K = K_1 = \ldots = K_n \), the minimization problem (9) can be reformulated in a more classical form. In fact, fixing \( f \) in (9) and minimizing \( z \) gives

\[
    z_i = \frac{y_i^T K_i \Psi f}{\|K_i \Psi f\|^2 + \lambda \int_M \Delta_M^2 f}, \quad i = 1, \ldots, n, \quad (10)
\]

which can then be used to show that the minimization problem (9) is equivalent to maximizing

\[
    \frac{(\Psi f)^T K^T \Psi y^T y K(\Psi f)}{\|K_i \Psi f\|^2 + \lambda \int_M \Delta_M^2 f}, \quad (11)
\]

with \( Y \) a \( n \times s \) real matrix, where the ith row of \( Y \) is the observation \( y_i^T \). This reformulation gives further insights on the interpretation of \( \hat{f} \) in (9). In fact, \( \hat{f} \) is such that \( K \Psi \hat{f} \) maximizes \( \frac{1}{n} Y^T Y \), i.e. the point-wise estimate of the covariance matrix in the sensors space. The term \( z^T z \) in (9), places the regularization term \( \lambda \int_M \Delta_M^2 f \) in the denominator of the equivalent formulation (11). Thus, \( \hat{f} \) is regularized by the choice of norm in the denominator of (11), in a similar fashion to the classic functional principal component formulation of Silverman (1996). Ignoring the spatial regularization, the point-wise evaluation of the PC function \( \Psi f \) in (11) can be interpreted as the first PC vector computed from the dataset of backprojected data \( [K_1^T y_1, \ldots, K_n^T y_n]^T \), similarly to what proposed in Dobriban et al. (2017) in the context of optimal prediction.

3.2 Algorithm

Here we propose a minimization approach for the objective function in (9), which we approach by alternating the minimization of \( z \) and \( f \) in an iterative algorithm. In (9), a normalization constraint must be considered to make the representation unique, as in fact multiplying \( z \) by a constant and dividing \( f \) by the same constant does not change the objective function. We optimize in \( z \) under the constraint \( \|z\|_2 = 1 \), which leads to a normalized version of the estimator (10)

\[
    z_i = \frac{y_i^T K_i \Psi f}{\sqrt{\sum_{i=1}^n y_i^T K_i \Psi f}}, \quad i = 1, \ldots, n. \quad (12)
\]

For a given \( z \), solving (9) with respect to \( f \) will turn out to be equivalent to solving an inverse problem. Specifically, consider now a triangulated surface \( \mathcal{M}_\mathcal{T} \), union of the finite set of triangles \( \mathcal{T} \), giving an approximated representation of the manifold \( \mathcal{M} \). We then consider the linear finite element space \( V \) consisting in a set of globally continuous functions over \( \mathcal{M}_\mathcal{T} \) that are affine where restricted to any triangle \( \tau \) in \( \mathcal{T} \), i.e.

\[
    V = \{v \in C^0(\mathcal{M}_\mathcal{T}) : v|_\tau \text{ is affine for each } \tau \in \mathcal{T}\}.
\]

10
This space is spanned by the nodal basis \( \phi_1, \ldots, \phi_\kappa \) associated to the nodes \( \xi_1, \ldots, \xi_\kappa \), corresponding to the vertices of the triangulation \( \mathcal{M}_T \). Such basis functions are Lagrangian, meaning that

\[
\phi_i(p, \xi_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}
\]

for every function \( f \in V \) has the form

\[
f(v) = \sum_{k=1}^\kappa f(\xi_k) \phi_k(v) = c^T \phi(v)
\]

for all \( v \in \mathcal{M}_T \). To ease the notation, we assume that the \( p \) points \( \{v_j\} \) associated with the evaluator function \( \Psi \) coincide with the nodes of the triangular mesh \( \xi_1, \ldots, \xi_\kappa \), and thus we have that the coefficients \( c \) are such that \( c = \Psi f \) for any \( f \in V \). Consequently, we are assuming the forward operators \( \{K_i\} \) to be \( \kappa \times s \) matrices, relating the \( \kappa \) points on the \( i \)th subject cortical surface, in one-to-one correspondence to \( \xi_1, \ldots, \xi_\kappa \), to the \( s \)-dimensional signal detected on the sensors of the \( i \)th subject.

Let now \( M \) and \( A \) be the mass and stiffness \( \kappa \times \kappa \) matrices defined in the Appendix. Than, the solution of (9), in the discrete space \( V \), is given by the following proposition.

**Proposition 1.** The Surface Finite Element solution \( \hat{f}_h \in V \) of model (9), for a given unitary norm vector \( z \), is \( \hat{f}_h = \hat{c}^T \phi \) where \( \hat{c} \) is the solution of

\[
\hat{c} = \left( \sum_{i=1}^n z_i^2 K_i^T K_i + \lambda AM^{-1} A \right)^{-1} \sum_{i=1}^n z_i K_i^T y_i.
\]

Equation (14) has the form of a penalized regression, where the penalty operator resulting from the discretization procedure is \( AM^{-1} A \).

**Algorithm 1 Inverse fPCA Algorithm**

1: Initialization:
   (a) Computation of \( M \) and \( A \)
   (b) Initialize \( z \), the scores vector associated to the first PC function
2: PC function’s estimation:
   Compute \( c \) such that
   \[
   \left( \sum_{i=1}^n z_i^2 K_i^T K_i + \lambda AM^{-1} A \right) c = \sum_{i=1}^n z_i K_i^T y_i
   \]
   \[
   f_h \leftarrow c^T \phi
   \]
3: Scores estimation:
   \[
   z_i \leftarrow \frac{y_i^T K_i \Psi f_h}{\sqrt{\sum_{i=1}^n y_i^T K_i \Psi f_h}}, \quad i = 1, \ldots, n
   \]
4: Repeat Steps 2–3 until convergence
The sparsity of the linear system (14), namely the amount of zero values entries, depends on the sparsity of its components. The matrices $M$ and $A$ are very sparse, however $M^{-1}$ it is not, in general. In the numerical analysis of Partial Differential Equations literature, the matrix $M^{-1}$ is generally replaced with the sparse matrix $\hat{M}^{-1}$, where $\hat{M}_{jj} = \sum_j M_{jt}$. The penalty operator $AM^{-1}A$ approximates very well the behavior of $AM^{-1}A$.

Moreover, in the case of single subjects longitudinal studies, we have a single forward operator $K = K_1 = \ldots = K_n$ common to all the observed signals, and consequently equation (14) can be rewritten as the sparse overdetermined system

$$K \sqrt{\lambda \hat{M}^{-1/2}} A c = \begin{bmatrix} Y^T z \\ 0 \end{bmatrix},$$

(15)

to be interpreted in a least-square sense. A sparse QR solver can be finally applied to efficiently solve the linear system (15).

In Algorithm 1 we summarize the main algorithmic steps to compute the PC functions and associated PC scores for indirectly observed functions.

### 3.3 Eigenfunctions of indirectly observed covariance operators

Suppose now we are in the case of a single forward operator $K = K_1 = \ldots = K_n$. Combining Steps 2–3 of Algorithm 1, and moving the normalization step from $z_i$ to $f_h$, we obtain the iterations

$$(K^T K + \lambda A\hat{M}^{-1}A)c = K^T \sum_{i=1}^n (y_i y_i^T) K \Psi f_h$$

$$f_h \leftarrow c^T \phi; f_h \rightarrow \frac{f_h}{\|f_h\|}.$$  

The obtained algorithm depends on the data only through $\sum_{i=1}^n (y_i y_i^T)$ that up to a constant is the covariance matrix computed on the sensors space. The proposed algorithm can thus be applied to situations where the observations $\{y_i\}$ are not available, but we are given only the associated $s \times s$ covariance matrix $S$ on the sensors space, computed from $\{y_i\}$.

### 4 Principal components of indirectly observed covariance functions

In this section we introduce a PC model for covariance operators, on the brain space, given a sample of covariance matrices, computed from indirectly measured signals on the sensors space. To this purpose, Algorithm 1 offers a valid first approach to deal with the problem. In fact, as explained in Section 3.3, this could be applied independently to each covariance matrix $S_i$, on the sensors space, to have a low rank reconstruction of the associated covariance function $C_i$ through its eigenfunctions on the brain space. The
vectorized matrix form of the reconstructed covariances could then be used to compute
the main modes of variation of the covariance functions in the brain space. However, in
the application considered in this work, the covariance functions, in a matrix form, are
represented by $24K \times 24K$ matrices, and due to their size not more than few of such
objects can be allocated on the memory of a workstation to compute the associated PCs.

This motivates the approach pursued in this section, where instead of independently
reconstructing the latent covariance functions, we formulate a model that allows to op-
timally represent the latent covariance functions by means of a set of PC covariance
functions and PC covariance scores. However, PCs are defined for linear spaces and are
not suitable to represent the non-Euclidean cone of positive semi-definite matrices. We
account for the non-Euclidean structure by introducing a proper distance in the space of
covariance matrices. To this purpose, in Section 4.1 we first briefly review the distances,
on the space of covariance matrices and covariance operators, that have been proposed
in the literature.

4.1 Metrics for the non-Euclidean space of covariance matrices

In Arsigny et al. (2006), the log-Euclidean distance of two positive-definite matrices $S_1$
and $S_2$ is defined as $d_{\text{log}}(S_1, S_2) = \| \log(S_1) - \log(S_2) \|_F$, where $\| \cdot \|_F$
is the Frobenius norm and $\log(\cdot)$ the matrix logarithm, i.e. $\log(S) = V \log(D)V^T$ with $S = VDV^T$ its spectral
decomposition and $\log(D)$ denoting the diagonal matrix whose entries are the logarithms
of the entries of $D$. Pennec et al. (2006) introduce an affine invariant Riemannian met-
ric for positive definite matrices defined as $d_{\text{Riem}}(S_1, S_2) = \| S_1^{-1/2}S_2S_1^{-1/2} \|_F$, where
$S_1^{-1/2} = VD^{-1/2}V^T$. A further option is the Cholesky distance $d_{\text{chol}}(S_1, S_2) = \| \text{chol}(S_1) - \text{chol}(S_2) \|_F$, where $\text{chol}(S)$ denotes either the upper or lower triangular Cholesky decom-
position of a positive-definite matrix $S$. The Cholesky distances cannot be generalized
to the case of positive semi-definite matrices due to the presence of 0-valued eigenvalues,
nor to the case of covariance operators, due to the infinite sequence of eigenvalues tending
to zero (Pigoli et al., 2014; Dryden et al., 2009).

The Cholesky distance could be regarded as a particular case of the class of distances
$d(S_1, S_2) = \| L_1 - L_2 \|_F$ derived from a symmetric decomposition of the type $S_k = L_k^T L_k$
with $k = 1, 2$. Another example of these decompositions is the square root decomposition
$L_k^{1/2} = V_k D_k^{1/2} V_k^T$. However, there is an infinite number of choices for such symmetric
decompositions, as in fact for any $R \in O(s)$, an element of the orthogonal group, $S = (RL)^T RL = L^T L$. Different choices of $R$ lead to different metrics. Motivated by this
argument, Dryden et al. (2009) introduce the non-Euclidean size-and-shape metric, which
for two covariance matrices $S_1$ and $S_2$, is defined as

$$d_S(S_1, S_2)^2 = \inf_{R \in O(s)} \| L_1 - RL_2 \|_F^2,$$

with $L_k$ denoting a generic decomposition of $S_k$ such that $S_k = L_k^T L_k$ for $k = 1, 2$, as for instance the Cholesky decomposition or the matrix square root decomposition.

The minimizing $R \in O(s)$ in (16), sometimes also referred as the Procrustes solution for

13
matching $L_1$ to $L_2$, is given by
\[ \hat{R} = WU^T, \tag{17} \]
where $U$ and $W$ are the left and right eigenvector matrices of the singular value decomposition of $L_1^T L_2 = W \Lambda U^T$.

Intuitively, the distance introduced removes the effect of any specific orthogonal re-parametrization from the element $L_2$ of the decomposition. Equivalently, the shape metric decomposition could be interpreted as the operation of seeking a representing data matrix $L_2$ such that the resulting associated empirical covariance matrix $L_2^T L_2$ is equal to the observed covariance matrix $S_2$. There are however an infinite number of matrices $RL_2$, with $R \in O(s)$, whose covariance equals $S_2$. The Procrustes distance chooses the re-parametrization $\hat{R}$ that minimizes the Frobenius distance of the representing data matrices $L_1$ and $RL_2$. Finally, note that in Dryden et al. (2009) the symmetric decomposition is equivalently defined as $S_k = L_k^T L_k$, instead of $S_k = L_k^T L_k$. Here we opt for the latter choice as $L_1$ can be interpreted as a classical data-matrix and it is thus clear that each column represents a set of observations in a specific spatial location.

### 4.2 Model

Consider now $n$ sample covariance matrices $S_1, \ldots, S_n$, each of size $s \times s$, representing $n$ different connectivity maps on the sensors space. Three of such covariance matrices, associated to three different individuals, are shown in Figure 2. Recall moreover that we denote with $\mathcal{M}$ the brain surface template and with $\{K_i \in \mathbb{R}^{p \times s}\}$ the set of subject specific forward operators, relating the signal at the $p$ pre-specified points $\{v_j\}$ on the cortical surface $\mathcal{M}$ with the signal detected on the $s$ sensors.

The aim of this section is to introduce a model for the first PC covariance function of the covariance functions $\{C_i\}$, on the brain space, associated to the actually observed covariance matrices $\{S_i\}$, on the sensors space. The matrices $\{S_i\}$ are related to the covariance functions $\{C_i\}$ through formula (8) that we recall here being
\[ S_i = K_i C_i K_i^T + E_i^T E_i, \quad i = 1, \ldots, n, \]
with $C_i = (C_i(v_j, v_l))_{j,l}$ and $\{v_j\}$ the sampling points associated to the operator $\Psi$. The first PC covariance function should represent the main mode of variation of the covariance functions $\{C_i\}$.

The first step in defining such a model is introducing a proper metric in the space of covariance matrices on the sensors space. Here, we opt for the size-and-shape distance defined in Section 4.1. So let $L_i$ be such that $L_i^T L_i = S_i$ for all $i = 1, \ldots, n$. In the practical implementation, we compute the matrices $\{L_i\}$ from the square-root decompositions of $\{S_i\}$. The size-and-shape distance on the space of covariance matrices, induces a concept of Fréchet sample mean $\hat{\Sigma} = \hat{\Pi}^T \hat{\Pi}$ defined as
\[ \hat{\Pi} = \arg \inf_{\Pi} \sum_{i=1}^{n} \inf_{R \in O(p)} \|R_i L_i - \Pi\|^2, \]

14
the solution of which can be found using the Generalized Procrustes Algorithm (Dryden and Mardia, 2016). Furthermore, the samples \( \{S_i\} \) can be represented in terms of the elements

\[ V_i = \hat{R}_i L_i - \hat{\Pi}, \]

with \( \{V_i\} \) elements in the unconstrained space of \( s \times s \) real matrices and \( \hat{R} \) given by (17).

From a geometric prospective \( \{V_i\} \) can be regarded as the coordinates, in the tangent space centered on \( \hat{\Sigma} = \hat{\Pi}^T \hat{\Pi} \), associated to \( \{S_i\} \). On the elements of the tangent space, the (Euclidean) Frobenius distance is used as an approximation to the non-Euclidean size-and-shape metric on the space of covariance matrices.

Introduce now \( \hat{f} \in H^2(\mathcal{M}) \) and \( \{\check{z}_i \in \mathbb{R}^s : i = 1, \ldots, n\} \), given by the following model:

\[
\begin{align*}
\{\check{z}_i\}, \hat{f} &= \arg\min_{\{z_i \in \mathbb{R}^s : f \in H^2(\mathcal{M})\}} \sum_{i=1}^{n} \|V_i - z_i(K_i \Psi f)^T\|_F^2 + \lambda \sum_{i=1}^{n} \|z_i\|^2 \int_\mathcal{M} \Delta_M^2 f. \tag{18}
\end{align*}
\]

We define the first PC covariance function to be the function \( F : \mathcal{M} \times \mathcal{M} \to \mathbb{R} \) such that \( F(v, v') = \hat{f}(v) \hat{f}(v') \), with \( v, v' \in \mathcal{M} \), which can be also denoted with \( F = \hat{f} \otimes \hat{f} \).

The empirical term of the equation (18) seeks for an approximation of the tangent space element \( V_i \) with a rank-one matrix \( z_i(K_i \Psi f)^T \). Thus, recalling the definition of the Procrustes size-and-shape metric, \( \hat{f} \) and \( \check{u} \) represent (up to tangent space approximation) a minimizer of

\[
\sum_i d_S(\check{S}_i, \|z_i\|^2 K_i \check{F} K_i^T)^2 + \lambda \sum_{i=1}^{n} \|z_i\|^2 \int_\mathcal{M} \Delta_M^2 f, \tag{19}
\]

with

\[ F_{ji} = F(v_j, v_i) = f(v_j) f(v_i), \]

where \( \{\check{S}_i\} \) denote the demeaned (in the size-and-shape distance) covariance matrices \( \{S_i\} \) and \( \{v_j\} \) denotes the sampling points associated to the operator \( \Psi \). Comparing formula (19) with (8), leads to the interpretation of \( F \) as the first PC covariance function of \( \{C_i\} \), where each underlying covariance function \( C_i \) is approximated by \( \|z_i\|^2 F \), i.e. the product of a subject specific constant \( \|z_i\|^2 \) and a component \( F \) common to all the observations. This is a well known property of the first PC, and could be regarded as a generalized version of the property (5) to PC of covariance functions. The coefficient \( \|z_i\|^2 \) defines the \( i \)th element of the first PC covariance scores vector. Finally, the regularizing term in (18) introduces spatial coherence on the estimated \( \hat{f}(v) \) and thus on the estimated first PC covariance function \( F = \hat{f}(v) \hat{f}(v') \), with \( v, v' \in \mathcal{M} \).
Algorithm 2 Inverse Covariance fPCA Algorithm

1: Tangent space covariance representations
   (a) Compute the representations $V_1, \ldots, V_n$ from $S_1, \ldots, S_n$ as

   $\hat{\Pi} = \arg\inf_{\Pi} \sum_{i=1}^{n} \inf_{R_i \in O(p)} \| R_i L_i - \Pi \|^2,$

   $V_i \leftarrow \hat{R}_i L_i - \hat{\Pi}, \quad i = 1, \ldots, n.$

2: Initialization:
   (a) Computation of $M$ and $A$
   (b) Initialize $\{z_i\}_{i=1}^{n}$, the scores of the first PC

3: PC function’s estimation from model (18):
   Compute $c$ such that

   $\hat{z}_i = \frac{c^T \phi}{\sqrt{\sum_{i=1}^{n} \| z_i \|^2}}, \quad i = 1, \ldots, n$

4: Scores estimation from model (18):

   $z_i \leftarrow V_i K_i \Psi f_h, \quad i = 1, \ldots, n$

4.3 Algorithm

Analogously to model (9), the resolution of equation (19) is approached in an iterative
fashion. We set $\sum_{i=1}^{n} \| z_i \|^2 = 1$ in the estimation procedure. This leads to the estimates
of $\{z_i\}$, given $f$, that are

$z_i = \frac{\tilde{z}_i}{\sqrt{\sum_{i=1}^{n} \| z_i \|^2}}, \quad i = 1, \ldots, n,$

with

$\tilde{z}_i = V_i K_i \Psi f_h, \quad i = 1, \ldots, n.$

In the discrete space $V$ introduced in Section 3.2, the estimate of $f$ given $\{z_i\}$ is instead
given by the following proposition.
Proposition 2. The Surface Finite Element solution \( \hat{f}_h \in V \) of model (18), given the vectors \( \{z_i\} \), is \( \hat{f}_h = \hat{c}^T \phi \) where \( \hat{c} \) is the solution of

\[
\hat{c} = \left( \sum_{i=1}^{n} \left\| z_i \right\|^2 K_i^T K_i + \lambda A M^{-1} A \right)^{-1} \sum_{i=1}^{n} K_i^T V_i^T z_i.
\]

Algorithm 2 contains a summary of the estimation procedure of the PC covariance functions and associated scores.

The choice to define the first PC covariance function to be a rank one (i.e. separable) covariance function \( F = \hat{f} \otimes \hat{f} \) is mainly driven by the following reasons. Firstly, rank-one covariance functions are easier to be interpreted due to their limited degrees of freedom. Secondly, on a rank one covariance function \( F = \hat{f} \otimes \hat{f} \) spatial coherence can be imposed by regularizing \( f \), as in fact done for the model (18), and this is fundamental in a setting of indirectly observed covariance functions. Finally, due to their size, might not be possible to store the entire covariance functions on the brain space, instead Model 18 allows an efficient approximation of such covariance functions in terms of their rank-one components.

5 Simulations

In this section, we perform simulations to assess the performances of the proposed algorithms. To reproduce as closely as possible the application setting, the cortical surfaces and the forward operators are taken from the MEG application described in Section 6. The details on the extraction and computation of such objects are left to the same section. For the same reason, the signals on the brain space considered here are vector-valued functions, specifically functions from the brain space \( M \to \mathbb{R}^3 \), as is the case in the MEG application. The proposed methodology can be trivially extended to successfully deal with this case, as shown in the following simulations.

5.1 PC components of indirectly observed functions

We consider \( M_T \) to be a triangular mesh, with \( 8K \) nodes, representing the cortical surface geometry of a subject, as shown on the left panel of Figure 1. Each of the \( 8K \) nodes will represent the discrete set of locations \( \{v_j\} \) associated to the sampling operator \( \Psi \). The locations of the nodes \( \{v_j\} \) on the brain space, the location of the 241 detectors on the sensors space and a model of the subject’s head, enable the computation of a forward operator \( K \) describing the relation between the signal generated on the locations \( \{v_j\} \), on the brain space, and the signal detected on the 241 sensors in the sensors space. In practice, the signal on each node \( v_j \) is described by a three dimensional vector, characterized by an intensity and a direction, while the signal detected on the sensors space is a scalar signal. Thus, the forward operator is a \( 241 \times 24K \) matrix.

Now, we produce synthetic data following the generative model (6). Specifically, on \( M_T \), we construct the three \( L^2 \) orthonormal vector-valued functions \( \{\psi_r = (\psi_{r,1}, \psi_{r,2}, \psi_{r,3}) : r = 1, 2, 3\} \), with \( \psi_r : M_T \to \mathbb{R}^3 \). These represent the PC functions to be estimated.
Figure 5: From top to bottom the components and the energy maps of the PC functions $\psi_1$, $\psi_2$ and $\psi_3$.

Figure 6: From left to right, the energy map of a generated function $x_i$, the associated signal $y_i$ on the sensors space with respectively no additional error, Gaussian error of standard deviation $\sigma = 5$ and Gaussian error of standard deviation $\sigma = 10$.

In Figure 5 we show the three components of $\{\psi_r\}$ and the associated energy maps $\{\|\psi_r\|^2 : r = 1, 2, 3\}$, with $\| \cdot \|$ denoting the Euclidean norm in $\mathbb{R}^3$. We then generate $n = 50$ smooth vector-valued functions $\{x_i\}$ on $\mathcal{M}_T$ by

$$x_i = z_{i1}\psi_1 + z_{i2}\psi_2 + z_{i3}\psi_3 \quad i = 1, \ldots, n,$$

where $\{z_{i1}\}, \{z_{i2}\}, \{z_{i3}\}$ are i.i.d realizations of the three independent random variables $z_r \sim N(0, \sigma_r^2) : r = 1, 2, 3$, with $\sigma_1 = 6$, $\sigma_2 = 3$ and $\sigma_3 = 1$.

The functions $\{x_i\}$ are sampled at the $8K$ nodes, and the forward operator is applied to the sampled values, producing a collection of vectors $\{y_i\}$ each of dimension 241, the number of active sensors. Moreover, on each entry of the vectors $\{y_i\}$, we add Gaussian noise with mean zero and standard deviation $\sigma$, for different choices of $\sigma$, to reproduce different signals to noise ratio regimes.

In the following, we compare the proposed PC function model in Section 3 to an alternative approach. In fact, as already mentioned, the individual functions $\{x_i\}$ could
be estimated from \( \{y_i\} \) by use of classical inverse problem estimators. Here, we adopt the estimates \( \{\hat{x}_i\} \) defined as

\[
\hat{x}_i = \arg \min_{f=(f_1,f_2,f_3): f_1, f_2, f_3 \in H^2(M)} \sum_{i=1}^{n} \|y_i - K \Psi f\|^2 + \lambda \int_M \|\Delta_M f\|^2, \quad i = 1, \ldots, n, \tag{21}
\]

where each \( \hat{x}_i \) is defined in such a way it balances a fitting term and a regularization term, which due to the fact that \( f \) is vector-valued, with a slight abuse of notation, is defined to be

\[
\Delta_M f = \begin{bmatrix}
\Delta_M & 0 & 0 \\
0 & \Delta_M & 0 \\
0 & 0 & \Delta_M
\end{bmatrix}
\begin{bmatrix}
f_1 \\
f_2 \\
f_3
\end{bmatrix},
\]

with \( \{f_l : l = 1, 2, 3\} \) denoting the components of \( f \). The same penalty operator is also adopted to generalize to vector-valued functions the PC models introduced in Sections 3-4. The constant \( \lambda \) is chosen by \( K \)-fold cross-validation, with \( K = 2 \). Once we obtain the estimates \( \{\hat{x}_i\} \) we can compute the estimated PC functions \( \{\hat{\psi}_r\} \) by applying classical multivariate PC analysis on the reconstructed objects \( \hat{x}_i \).

The estimates are compared to the proposed PC function model, as described in Algorithm 1, with 15 iterations. The smoothing coefficient \( \lambda \) here is also chosen by \( K \)-fold cross-validation, with \( K = 2 \). To evaluate the performances of the two approaches, we generate 100 datasets as previously detailed. The quality of the estimated \( r \)th PC function is then measured with \( \sum_{l=1}^{3} \|\nabla_M (\psi_{r,l} - \hat{\psi}_{r,l})\|^2 \). The operator \( \nabla_M \) is the gradient operator on the manifold \( M \) and could be regarded as a generalization to manifolds of the gradient operator as defined for a 2-dimensional Euclidean domain. The results are summarized in the boxplots in Figure 7, for two different signal to noise ratios, where the Gaussian noise has standard deviation \( \sigma = 5 \) and \( \sigma = 10 \). In Figure 6 we show an example of a signal on the brain space corrupted with the specified noise levels.

The boxplots highlight the fact that the proposed approach provides better estimates of the PC functions in particular in a low signal to noise ratio regime, i.e. the estimation of the third PC function. More surprisingly is the stability of the estimates of the proposed algorithm across the generated datasets, as opposed to the naive approach, which returns multiple particularly unsatisfactory reconstructions. An example of such reconstructions is shown in Figure 8.

5.2 PC components of indirectly observed covariance functions

In this section, we consider \( M_T \) to be a \( 8K \) nodes triangular mesh, this time representing a template geometry of the cortical surface, which is shown in Figure 9. This contains only the geometric features common to all subjects. Moreover, each subject’s cortical surface is also represented by a \( 8K \) nodes triangular surface, which is used, together with the locations of the 241 detectors on the sensors space, and the head model, to compute a forward operator \( K_i \) for the \( i \)th subject. The \( 8K \) nodes of each subject’s triangular mesh are in correspondence with the \( 8K \) nodes of the template mesh \( M_T \). This allows the model to be defined on the template \( M_T \).
Figure 7: On the left, a summary of the results in a medium signal-to-noise ratio regime. On the right, a summary of the results in a low signal-to-noise ratio regime. Each boxplot compares the reconstruction errors obtained by applying the two steps naive method with those obtained by applying Algorithm 1.

Figure 8: On the first row the energy maps of the true three PC components to be estimated, on the second row the estimations given by the two steps naive method, and on the third row the reconstructions obtained by applying Algorithm 1.

As in the previous Section, we construct three functions, $L^2$ orthonormal in $M_T$ 

\[ \{ \psi_r = (\psi_{r,1}, \psi_{r,2}, \psi_{r,3}) : r = 1, 2, 3 \} \]. The energy maps of \( \{ \psi_r \} \) are shown in Figure 10. We generate synthetic data from model (8) as follows:

\[
C_i = \sum_{r=1}^{3} z_{ir}^2 \psi_r \otimes \psi_r = \sum_{r=1}^{3} z_{ir}^2 \begin{bmatrix} \psi_{r,1} \otimes \psi_{r,1} & \psi_{r,1} \otimes \psi_{r,2} & \psi_{r,1} \otimes \psi_{r,3} \\ \psi_{r,2} \otimes \psi_{r,1} & \psi_{r,2} \otimes \psi_{r,2} & \psi_{r,2} \otimes \psi_{r,3} \\ \psi_{r,3} \otimes \psi_{r,1} & \psi_{r,3} \otimes \psi_{r,2} & \psi_{r,3} \otimes \psi_{r,3} \end{bmatrix},
\]
where \( z_{i1}, z_{i2}, z_{i3} \) are i.i.d realizations of the three independent random variables \( \{ z_r \sim N(0, \sigma_r^2) : r = 1, 2, 3 \} \), with \( \sigma_1 = 6, \sigma_2 = 5 \) and \( \sigma_3 = 4 \). The matrix-valued form of the covariance functions arises from the fact that the observed functions on the brain space are vector-valued. Subsequently, we construct the point-wise evaluations matrices \( C_i \in \mathbb{R}^{24K \times 24K} \), from which the correspondent covariance matrices on the sensors space are defined as

\[
S_i = K_i C_i K_i^T + E_i^T E_i, \quad i = 1, \ldots, n.
\]

The term \( E_i^T E_i \) is an error term, where \( E_i \) is a \( s \times s \) matrix with each entry that is an independent sample from a Gaussian distribution with mean zero and standard deviation 5. We then apply Algorithm 2 with, 15 iterations, feeding in input \( \{ S_i \} \). The results are shown in Figure 10, in terms of energy maps of the reconstructed functions \( \{ \hat{\psi}_r \} \). These are a close approximation of the underlying functions \( \{ \psi_r \} \). The fidelity measure \( \sum_{l=1}^3 \| \nabla_M(\psi_r,l - \hat{\psi}_r,l) \|^2 \) of such estimates is \( 1.6 \times 10^{-3} \), \( 1.4 \times 10^{-3} \) and \( 1 \times 10^{-2} \), for \( \psi_1 \), \( \psi_2 \) and \( \psi_3 \) respectively, which is comparable in term of order of magnitude to the results obtained in the case of PCs of indirectly observed functions. Across the generation of multiple datasets, results are stable, with the exception of few situations where the cross-validation approach suggests a penalization coefficient \( \lambda \) that under-smoothes the solution, due to very similar associated signals on the sensors space of the

\[
\begin{align*}
\psi_1 & \quad \psi_2 & \quad \psi_3 \\
\text{Original} & \quad \text{Inverse} & \quad \text{Covariance} \\
\text{fPCA} & \quad \text{fPCA} & \quad \text{fPCA}
\end{align*}
\]
under-smoothed solution and the real solution. However, the cross-validation is only a possible approach to the choice of the penalization constant, and many other options have been proposed in the inverse problems literature, (see, e.g., Vogel, 2002).

6 Application

In this section, we apply the developed models to the publicly available HCP Young Adult dataset (Essen et al., 2012). This dataset comprises multi-modal neuroimaging data such as structural scans, resting-state and task-based functional MRI scans, and resting-state and task-based MEG scans from a large number of healthy volunteers. In the following, we briefly review the pre-processing pipeline, applied to such data by the HCP, to ultimately facilitate their use.

6.1 Pre-processing

For each individual a high-resolution 3D structural MRI scan has been acquired. This returns a 3D image describing the structure of the gray and white matter in the brain. Gray matter consists mostly of neuronal cell bodies, and it is the source of most of our neuronal activity. White matter is made of axons connecting the different parts of the gray matter. If we exclude the sub-cortical structures, gray matter is mostly distributed at the outer surface of the cerebral hemispheres. This is also known as the cerebral cortex.

By segmentation of the 3D structural MRI, it is possible to separate gray matter from white matter, in order to extract the cerebral cortex structure. Subsequently a mid-thickness surface, interpolating the mid-points of the cerebral cortex, can be estimated, resulting in a 2D surface embedded in a 3D space that represents the geometry of the cerebral cortex. In practice, such a surface, sometimes referred to as cortical surface, is a triangulated surface. Moreover, from the 3D structural MRI, a surface describing the individuals’ head can be extracted. The latter plays a role in the derivation of the model for the electrical/magnetic propagation of the signal from the cerebral cortex to the sensors. An example of the cortical surface of a single subject, is shown on the right panel in Figure 1, instead the associated head surface and MEG sensors positions are shown on the left panel of the same figure.

Moreover, a surface based registration algorithm has been applied to register each of the extracted cortical surfaces to a triangulated template cortical surface, which is shown in Figure 9. Post registration, the triangulated template cortical surface is sub-sampled to a 8K nodes surface. Moreover, the nodes on the cortical surface of each subject are also sub-sampled to a set of 8K nodes in correspondence to the 8K nodes of the template. For each subject, a $248 \times 24K$ matrix, representing the forward operator, has been computed with FieldTrip (Oostenveld et al., 2011) from its head surface, cortical surface and sensors position. Such matrix relates the vector-valued signals in $\mathbb{R}^3$, on the nodes of the triangulation of the cerebral cortex, to the one detected from the sensors, consisting of 248 magnetometer channels.
With the aim of studying the functional connectivity of the brain, for each subject, three 6 minutes resting state MEG scans have been performed, of which one session is used in our analysis. During the 6 minutes, data are collected from the sensors at 600K uniformly distributed time-points. Using FieldTrip, classical pre-processing is applied to the detected signals, such as low quality channels and low quality segments removal. Details of this procedure can be found in the HCP MEG acquisition protocol. Moreover, band pass filtering is applied, limiting the spectrum of the signal to the [12.5, 29]Hz, also known as the low beta waves.

6.2 Analysis

Here we apply the models proposed in this paper to the HCP data. The first part of the analysis focuses on the study of the dynamic functional connectivity of a specific subject. For this purpose, we subdivide the 6 minutes session in 20 intervals. Each of these intervals is used to compute a covariance matrix in the sensors space, resulting in 20 covariance matrices $S_1, \ldots, S_{20}$. The aim is understanding the main modes of variation of the functional connectivity of the subject on its brain space. Thus, Algorithm 2 is applied to $S_1, \ldots, S_{20}$ to find the PC covariance functions, with 20 iterations and $K$-fold cross-validation, with $K = 2$. The energy maps of the estimated $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ resulting from the analysis are shown in Figure 11. These are associated to the first three PC covariance functions $\hat{\psi}_1 \otimes \hat{\psi}_1$, $\hat{\psi}_2 \otimes \hat{\psi}_2$ and $\hat{\psi}_3 \otimes \hat{\psi}_3$, which are in fact difficult to visualize and interpret, so instead it is useful to focus on the energy maps of $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$. In fact, high intensity areas, in yellow, give a good representation of which areas present high variability, in time, for the way their points are functionally connected to each other. These are points where care should be taken in establishing static functional connectivity assessments.

Figure 11: Energy maps of the estimated $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying Algorithm 2 to the covariance matrices computed from the MEG resting state data of a single subject on 20 consecutive time intervals.

The second part of the analysis focuses on applying the proposed methodology to a
multi subject setting. Specifically, 40 different subjects are considered. For each subject, the 6 minutes scan is used to compute a covariance matrix associated to the subject, resulting in 40 covariance matrices $S_1, \ldots, S_{40}$. The template geometry in Figure 9 is used as a model of the brain space. Algorithm 2 is then applied to $S_1, \ldots, S_{40}$ to find the PC covariance functions on the template brain, associated to $S_1, \ldots, S_{40}$. We run the algorithm for 20 iterations, and choose the regularizing parameter by $K$-fold cross-validation, with $K = 2$. The energy maps of the estimated functions $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$, associated to the first three PC covariance functions $\hat{\psi}_1 \otimes \hat{\psi}_1$, $\hat{\psi}_2 \otimes \hat{\psi}_2$ and $\hat{\psi}_3 \otimes \hat{\psi}_3$, are shown in Figure 12. High intensity areas, in yellow, indicate which areas present high variability, between subjects, for the way their points are functionally connected to each other. This opens up the possibility to understand population level variation in functional connectivity, and indeed, whether, just as we need different forward operators for individuals (due to anatomical differences), we should also be considering both population and subject specific connectivity maps when analysing networks such as the default network. It is also of interest to note that the main modes of population variations in functional connectivity are similar in the first mode but different in subsequent modes with respect to the individual subject temporal functional connectivity variation.

Figure 12: Energy maps of the estimated $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying Algorithm 2 to the covariance matrices computed from the MEG resting state data of 40 different subjects.

7 Discussion

In this work we introduce a general framework for the statistical analysis of functional data in an inverse problem context. In particular, two different settings are considered. In the first one, we introduce a model for indirectly observed functional data in an unconstrained space, which outperforms the naive approach of solving the inverse problem individually for each sample. Moreover this plays an important role in the second setting, where we consider the case of samples that are indirectly observed covariance functions, and thus constrained to be positive-definite. We deal with the non-linearity introduced by such constraint by equipping the space with a proper distance and working with
the tangent space representation of the objects, yet incorporating spatial information in their estimation. The proposed methodology is finally applied to the study of brain connectivity from the signals arising from MEG scans.

The models proposed here can be extended in many interesting directions. From an applied prospective, it is of interest to apply the proposed methodology to different settings, not necessarily involving neuroimaging, where studying second order information has been so far prohibitive. From a modeling point of view, it is of interest to take a step further the integration of the inverse problems literature with the statistical approach we adopt in this paper. For instance, penalization terms that have been shown to be successful in the inverse problems literature, e.g. total variation penalization, could be introduced in our models.

Code

All the code and simulation materials are available at http://www.statslab.cam.ac.uk/~jada2/InvCov_code.tar.gz.

Appendices

A Discrete solutions

Proof of Proposition 1. We want to find a minimizer \( \hat{f} \in H^2(\mathcal{M}) \), given \( z \) with \( ||z|| = 1 \), of the objective function in (9):

\[
\sum_{i=1}^{n} \| y_i - z_i K_i \Psi f \|^2 + \lambda z^T z \int_{\mathcal{M}} \Delta_M^2 \hat{f}
\]

\[
\propto (\Psi f)^T (\sum_{i=1}^{n} z_i^2 K_i^T K_i) \Psi f - 2 (\Psi f)^T (\sum_{i=1}^{n} z_i K_i^T y_i) + \lambda \int_{\mathcal{M}} \Delta_M^2 \hat{f}. \quad (22)
\]

An equivalent formulation of a minimizer \( \hat{f} \in H^2(\mathcal{M}) \) of such objective function is given by satisfying the equation

\[
(\Psi \varphi)^T (\sum_{i=1}^{n} z_i^2 K_i^T K_i) \Psi \hat{f} + \lambda \int_{\mathcal{M}} \Delta_M \varphi \Delta_M \hat{f} = (\Psi \varphi)^T (\sum_{i=1}^{n} z_i K_i^T y_i) \quad (23)
\]

for every \( \varphi \in H^2(\mathcal{M}) \) (see Braess, 2007, Chapter 2). Moreover, such minimizer is unique if \( A(\varphi, f) = (\Psi \varphi)^T (\sum_{i=1}^{n} z_i^2 K_i^T K_i) \Psi f + \lambda \int_{\mathcal{M}} \Delta_M \varphi \Delta_M f \) is definite positive. Given that for a closed manifold \( \mathcal{M}, \int_{\mathcal{M}} \Delta_M f = 0 \) iff \( f \) is a constant function (Dziuk and Elliott, 2013), such condition is equivalent to assume that \( \ker(\sum_{i=1}^{n} z_i^2 K_i^T K_i) \), the kernel of \( \sum_{i=1}^{n} z_i^2 K_i^T K_i \), does not contain the subspace of \( p \)-dimensional constant vectors.
Moreover, we can rewrite equation (23) by introducing the auxiliary function \( h \in L^2(\mathcal{M}) \) as

\[
\begin{align*}
\begin{cases}
(\Psi \varphi)^T \left( \sum_{i=1}^{n} z_i^2 K_i^T K_i \right) \Psi \hat{f} + \lambda \int_{\mathcal{M}} \Delta_M \varphi h = (\Psi \varphi)^T \left( \sum_{i=1}^{n} z_i K_i^T y_i \right) \\
\int_{\mathcal{M}} \nabla_M \hat{f} \cdot \nabla_M v - \int_{\mathcal{M}} h v = 0
\end{cases}
\end{align*}
\]

(24)

for all \((\varphi, v) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})\). Now, asking \( h, v \) to be such that \( h, v \in H^1(\mathcal{M}) \), the Sobolev space of functions in \( L^2(\mathcal{M}) \) with first distributional derivatives in \( L^2(\mathcal{M}) \), by integration by parts we can rewrite (24) as the problem of finding \((\hat{f}, \hat{g}) \in (H^1(\mathcal{M}) \cap C^0(\mathcal{M})) \times H^1(\mathcal{M})\)

\[
\begin{align*}
\begin{cases}
(\Psi \varphi)^T \left( \sum_{i=1}^{n} z_i^2 K_i^T K_i \right) \Psi \hat{f} + \lambda \int_{\mathcal{M}} \nabla_M \varphi \cdot \nabla_M h = (\Psi \varphi)^T \left( \sum_{i=1}^{n} z_i K_i^T y_i \right) \\
\int_{\mathcal{M}} \nabla_M \hat{f} \cdot \nabla_M v - \int_{\mathcal{M}} h v = 0
\end{cases}
\end{align*}
\]

(25)

for all \((h, v) \in (H^1(\mathcal{M}) \cap C^0(\mathcal{M})) \times H^1(\mathcal{M})\). The operator \( \nabla_M \) is the gradient operator on the manifold \( \mathcal{M} \). The gradient operator \( \nabla_M \) is such that \((\nabla_M v)(p)\), for \( v \) a smooth real function on \( \mathcal{M} \) and \( p \in \mathcal{M} \), takes value on the tangent space on \( p \). We denote with \( \cdot \) the scalar product in the tangent space.

Now that equation (25) has been written in terms of first order derivatives only, we discretize the equation by looking for a solution in the discrete space \( V \subset H^1(\mathcal{M}) \), i.e. finding \( \hat{f}, \hat{g} \in V \)

\[
\begin{align*}
\begin{cases}
(\Psi \varphi)^T \left( \sum_{i=1}^{n} z_i^2 K_i^T K_i \right) \Psi \hat{f} + \lambda \int_{\mathcal{M}} \nabla_M \varphi \cdot \nabla_M h = (\Psi \varphi)^T \left( \sum_{i=1}^{n} z_i K_i^T y_i \right) \\
\int_{\mathcal{M}} \nabla_M \hat{f} \cdot \nabla_M v - \int_{\mathcal{M}} h v = 0
\end{cases}
\end{align*}
\]

(26)

for all \( h, v \in V \).

Define now the \( \kappa \times \kappa \) matrices \((M)_{jl} = \int_{M^T} \phi_j \phi_l\) and \((A)_{jl} = \int_{M^T} \nabla_M \phi_j \cdot \nabla_M \phi_l\). Note that requiring (26) to hold for all \( h, v \in V \) is equivalent to require that (26) holds for all \( h, v \) that are basis elements of \( V \), thus exploiting the basis expansion formula (13) we can characterize (26) with the solution of the linear system

\[
\begin{bmatrix}
\sum_{i=1}^{n} z_i^2 K_i^T K_i & \lambda A \\
A & -M
\end{bmatrix}
\begin{bmatrix}
\hat{c} \\
\hat{q}
\end{bmatrix} =
\begin{bmatrix}
\sum_{i=1}^{n} z_i K_i^T y_i \\
0
\end{bmatrix},
\]

(27)

where \( \hat{c} \) and \( \hat{q} \) are the basis coefficients of \( f \in V \) and \( g \in V \), respectively. Solving (27) in \( \hat{c} \) leads to

\[
\left( \sum_{i=1}^{n} z_i^2 K_i^T K_i + \lambda A M^{-1} A \right) \hat{c} = \sum_{i=1}^{n} z_i K_i^T y_i.
\]

(28)
Proof of Proposition 2. We want to find a minimizer \( \hat{f} \in H^2(\mathcal{M}) \), given \( \{z_i\} \) with \( \sum_{i=1}^{n} \|z_i\|^2 = 1 \), of the objective function in (9):

\[
\sum_{i=1}^{n} \|V_i - z_i(K_i\Psi f)^T\|^2 + \lambda \sum_{i=1}^{n} \|z_i\|^2 \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f \\
x(\Psi f)^T \left( \sum_{i=1}^{n} \|z_i\|^2 K_i^T K_i \right) \Psi f - 2(\Psi f)^T \sum_{i=1}^{n} K_i^T V_i^T z_i.
\] (29)

Comparing (29) with (22) it is evident that by following the same steps of the Proof of Proposition 1 we obtain the desired result, which is

\[
\hat{c} = \left( \sum_{i=1}^{n} \|z_i\|^2 K_i^T K_i + \lambda AM^{-1} A \right)^{-1} \sum_{i=1}^{n} K_i^T V_i^T z_i.
\]

References

S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41, 1999.

V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

D. Braess. *Theory, fast solvers, and applications in elasticity theory*. Cambridge University Press, 2007.

F. Bunea and L. Xiao. On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to FPCA. *Bernoulli*, 21(2):1200–1230, 2015.

E. Dobriban, W. Leeb, and A. Singer. Optimal prediction in the linearly transformed spiked model. *ArXiv e-prints*, Sept. 2017.

I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis, with Applications in R. Second Edition*. John Wiley and Sons, Chichester, 2016.

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

G. Dziuk and C. M. Elliott. Finite element methods for surface PDEs. *Acta Numerica*, 22:289–396, 2013.
D. V. Essen, K. Ugurbil, E. Auerbach, D. Barch, T. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. Curtiss, S. D. Penna, D. Feinberg, M. Glasser, N. Harel, A. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S. Petersen, F. Prior, B. Schlaggar, S. Smith, A. Snyder, J. Xu, and E. Yacoub. The human connectome project: A data acquisition perspective. NeuroImage, 62(4):2222–2231, 2012.

B. Ettinger, S. Perotto, and L. M. Sangalli. Spatial regression models over two-dimensional manifolds. Biometrika, 103(1):71–88, 2016.

P. Hall, H. G. Müller, and J. L. Wang. Properties of principal component methods for functional and longitudinal data analysis. Annals of Statistics, 34(3):1493–1517, 2006.

R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, New York, NY, USA, 2 edition, 2003. ISBN 0521540518.

J. Z. Huang, H. Shen, and A. Buja. Functional principal components analysis via penalized rank one approximation. Electronic Journal of Statistics, 2:678–695, 2008.

R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. D. Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang. Dynamic functional connectivity: Promise, issues, and interpretations. NeuroImage, 80:360–378, 2013. Mapping the Connectome.

I. M. Johnstone and B. W. Silverman. Speed of estimation in positron emission tomography and related inverse problems. Ann. Statist., 18(1):251–280, 03 1990.

E. Lila, J. A. D. Aston, and L. M. Sangalli. Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging. Ann. Appl. Stat., 10 (4):1854–1879, 12 2016.

M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing mri. IEEE Signal Processing Magazine, 25(2):72–82, March 2008.

R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience, 2011.

X. Pennec, P. Fillard, and N. Ayache. A riemannian framework for tensor computing. International Journal of Computer Vision, 66(1):41–66, Jan 2006.

D. Pigoli, J. A. D. Aston, I. L. Dryden, and P. Secchi. Distances and inference for covariance operators. Biometrika, 101(2):409–422, 2014.

J. O. Ramsay and B. W. Silverman. Functional Data Analysis. Springer Series in Statistics. Springer, 2nd edition, 2005. ISBN 038740080X.
T. Ramsay. Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):307–319, 2002.

F. Riesz and B. Sz.-Nagy. *Functional Analysis*. Frederick Ungar Publishing Co., 1955.

L. M. Sangalli, J. O. Ramsay, and T. O. Ramsay. Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4):681–703, 2013.

B. W. Silverman. Smoothened functional principal components analysis by choice of norm. *The Annals of Statistics*, 24(1):1–24, 02 1996.

A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

T. S. Tian, J. Z. Huang, H. Shen, and Z. Li. A two-way regularization method for MEG source reconstruction. *Annals of Applied Statistics*, 6(3):1021–1046, 2012.

E. D. Vito, L. Rosasco, A. Caponnetto, U. D. Giovannini, and F. Odone. Learning from examples as an inverse problem. *J. Mach. Learn. Res.*, 6:883–904, Dec. 2005.

C. R. Vogel. *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002. ISBN 0898715075.

S. N. Wood, M. V. Bravington, and S. L. Hedley. Soap film smoothing. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(5):931–955, 2008.

F. Yao, H.-g. Müller, and J.-l. Wang. Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

M. S. Zhdanov. *Geophysical inverse theory and regularization problems*. Amsterdam ; Oxford : Elsevier Science, 2002. ISBN 0444510893.