

Smooth Principal Component Analysis over two-dimensional manifolds with an application to Neuroimaging

Eardi Lila^{*1,3}, John A. D. Aston^{†2}, and Laura M. Sangalli^{‡3}

¹Cambridge Centre for Analysis, University of Cambridge

²Statistical Laboratory, DPMMS, University of Cambridge

³MOX, Dipartimento di Matematica, Politecnico di Milano

Abstract

Motivated by the analysis of high-dimensional neuroimaging signals located over the cortical surface, we introduce a novel Principal Component Analysis technique that can handle functional data located over a two-dimensional manifold. For this purpose a regularization approach is adopted, introducing a smoothing penalty coherent with the geodesic distance over the manifold. The model introduced can be applied to any manifold topology, can naturally handle missing data and functional samples evaluated in different grids of points. We approach the discretization task by means of finite element analysis and propose an efficient iterative algorithm for its resolution. We compare the performances of the proposed algorithm with other approaches classically adopted in literature. We finally apply the proposed algorithm to resting state functional magnetic resonance imaging data, showing that a cross-validation approach justifies the presence of the penalization term in the analysis of human brain connectome data.

1 Introduction

The recent growth of data arising from neuroimaging has led to profound changes in the understanding of the brain. Neuroimaging is a multidisciplinary activity and the role of statistics in its success should not be underestimated. Much of the work to date has been to determine how to use statistical models in high-dimensional settings that arise out of such imaging modalities as functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG). However, it is becoming increasingly clear that there is now a need to incorporate more and more complex information about brain structure and function into the statistical analysis in order to take understanding of the brain to the next level.

^{*}el425@cam.ac.uk

[†]j.aston@statslab.cam.ac.uk

[‡]laura.sangalli@polimi.it

Considerable amounts of the brain signal captured, for example, by fMRI arise from the cerebral cortex. The cerebral cortex is the highly convoluted thin sheet where most neural activity is focused. It is natural to represent this thin sheet as a 2D surface embedded in a 3D space, structured with a 2D geodesic distance, rather than the 3D Euclidean distance within the volume. In fact, functionally distinct areas may be close to each other if measured with Euclidean distance, but due to the highly convoluted morphology of the cerebral cortex, their 2D geodesic distance along the cortical surface can be far greater. While early approaches to the analysis of hemodynamic signals ignore the morphology of the cortical surface, it has now been well established [Glasser et al. (2013) and references therein] that is beneficial to analyze neuroimaging data through the processing of the signals on the cortical surface using surface-constrained techniques. Classical tools such as non-parametric smoothing models have already been adapted to deal with this kind of data, see e.g. Chung et al. (2014).

The goal of the present paper is to introduce a novel Principal Component Analysis (PCA) technique suitable for working with functional signals distributed over curved domains and specifically over two-dimensional smooth Riemannian manifolds, such as the cortical surface.

The cortical surface can be extracted from structural Magnetic Resonance Imaging (MRI), a non-invasive scanning technique to visualize the internal structure of the brain, rendering it as a 3D image with high spatial resolution. The signal of interest, which we want to analyse with respect to the surface, comes from fMRI, which detects a Blood Oxygen Level Dependent (BOLD) signal [Ogawa et al. (1990)] as a series of repeated measurements in time, yielding a time series of 3D images. An increased neural activity in a particular area of the brain causes an increased demand for oxygen. As the fMRI signal is related to changes in the relative ratio of oxy- to deoxy-hemoglobin, due to their differing magnetic properties, the signal captured within an fMRI scan is considered to be a surrogate for neural activity and is used to produce activation maps or investigate brain functional connectivity. The fMRI signal of each individual related to the neuronal activity in the cerebral cortex is generally mapped on a common template cortical surface, to allow multi-subject statistical analysis.

In this paper, in particular, we will focus our attention on functional connectivity (FC). FC maps, on the cortical surface, can be constructed computing the pairwise correlation between all vertex's fMRI time-series and the mean time-series of a region of interest. This will result in one FC map for each subject that provides a clear view of areas to which the region of interest is functionally connected.

In practice, the template cortical surface is represented by a triangulated surface that can be considered a discrete approximation of the underlying smooth compact two-dimensional Riemannian manifold $\mathcal{M} \subset \mathbb{R}^3$ modelling the cortical surface. Each resting state functional connectivity map can be represented by a function $x_i : \mathcal{M} \rightarrow \mathbb{R}$. Once we have the correlation maps on the cortical surface we want to study how the phenomena varies from subject to subject. A statistical technique that enables to do that is PCA. It is natural to contextualize this task in the framework of Functional Data Analysis [Ramsay and Silverman (2005)].

The rest of the article is organized as follows. In Section 2 we establish the formal theoretical properties of Functional PCA (FPCA) to the case of random functions whose domain is a manifold \mathcal{M} . In Section 3 we introduce a novel FPCA model and propose an algorithm for its resolution. We then give some simulation results in Section 4, indicating the performance of our methodology, as compared to other methods in literature. We then return to the FC maps example in Section 5, to consider how the surface based PCA analysis might be used in this case and make some concluding remarks in Section 6.

2 Functional principal component analysis

Consider the space of square integrable functions on \mathcal{M} : $L^2(\mathcal{M}) = \{f : \mathcal{M} \rightarrow \mathbb{R} : \int_{\mathcal{M}} |f(p)|^2 dp < \infty\}$ with the inner product $\langle f, g \rangle_{\mathcal{M}} = \int_{\mathcal{M}} f(p)g(p)dp$ and norm $\|f\|_{\mathcal{M}} = \sqrt{\int_{\mathcal{M}} |f(p)|^2 dp}$. Consider the random variable $X : \Omega \rightarrow L^2(\mathcal{M})$, with mean $\mu = \mathbb{E}[X]$ and a finite second moment, i.e. $\int_{\mathcal{M}} \mathbb{E}[X^2] < \infty$. Moreover assume that its covariance function $K(p, q) = \mathbb{E}[(X(p) - \mu(p))(X(q) - \mu(q))]$ is square integrable. Mercer's Lemma [Riesz and Sz.-Nagy (1955)] guarantees the existence of a non-increasing sequence (κ_j) of eigenvalues of K and an orthonormal sequence of corresponding eigenfunctions (ψ_j) , such that

$$\int_{\mathcal{M}} K(p, q)\psi_j(p)dp = \kappa_j\psi_j(q), \quad \forall q \in \mathcal{M}. \quad (1)$$

Moreover $K(p, q) = \sum_{j=1}^{\infty} \kappa_j \psi_j(p)\psi_j(q)$ for each $p, q \in \mathcal{M}$. Thus X can be expanded as $X = \mu + \sum_{j=1}^{\infty} \varepsilon_j \psi_j$, where the random variables $\varepsilon_1, \varepsilon_2, \dots$ are uncorrelated and are given by $\varepsilon_j = \int_{\mathcal{M}} \{X(p) - \mu(p)\}\psi_j(p)dp$. This is also known as the Karhunen-Loève (KL) expansion of X .

The collection (ψ_j) defines the strongest modes of variation in the random function X and are called Principal Component (PC) functions. In fact ψ_1 is such that

$$\psi_1 = \underset{\phi : \|\phi\|_{\mathcal{M}}=1}{\operatorname{argmax}} \int_{\mathcal{M}} \int_{\mathcal{M}} \phi(p)K(p, q)\phi(q)dpdq,$$

while ψ_m , for $m > 1$, solves an analogous problem with the added constraint of ψ_m being orthogonal to the previous $m - 1$ functions $\psi_1, \dots, \psi_{m-1}$, i.e.

$$\psi_m = \underset{\substack{\phi : \|\phi\|_{\mathcal{M}}=1 \\ \langle \phi, \psi_j \rangle_{\mathcal{M}}=0 \quad j=1, \dots, m-1}}{\operatorname{argmax}} \int_{\mathcal{M}} \int_{\mathcal{M}} \phi(p)K(p, q)\phi(q)dpdq.$$

The random variables $\varepsilon_1, \varepsilon_2, \dots$ are usually called PC scores.

Another important property of PC functions is the best M basis approximation. In fact, for any fixed $M \in \mathbb{N}$, the first M PC functions of X satisfies

$$(\psi_i)_{i=1}^M = \underset{(\{\phi_m\}_{m=1}^M : \langle \phi_m, \phi_l \rangle = \delta_{ml})}{\operatorname{argmin}} \mathbb{E} \int_{\mathcal{M}} \left\{ X - \mu - \sum_{m=1}^M \langle X, \phi_m \rangle \phi_m \right\}^2, \quad (2)$$

where δ_{ml} is the Kronecker delta, such that $\delta_{ml} = 1$ for $m = l$ and 0 otherwise.

Suppose x_1, \dots, x_n are n smooth samples from X . Usually, for each of these functions, only noisy evaluations $x_i(p_j)$ on a fixed discrete grid of points p_1, \dots, p_s are given. In this setting, we will now recall the two standard approaches to FPCA. The pre-smoothing approach and the regularized PCA approach.

The pre-smoothing approach is based on the two following steps. In the first step the observations associated to each function are smoothed, in order to obtain smooth representations of x_1, \dots, x_n . Then the sample mean $\bar{x} = n^{-1} \sum_i x_i$ and the sample covariance $\hat{K}(p, q) = \frac{1}{n} \sum_{i=1}^n (x_i(p) - \bar{x}(p))(x_i(q) - \bar{x}(q))$ are used to estimate μ and K respectively. Applying the orthonormal basis expansion to \hat{K} is then possible to write $\hat{K}(p, q) = \sum_{j=1}^{\infty} \hat{\kappa}_j \hat{\psi}_j(p) \hat{\psi}_j(q)$ with $p, q \in \mathcal{M}$. The sequence $\hat{\psi}_1, \hat{\psi}_2, \dots$ is usually treated as an approximation of ψ_1, ψ_2, \dots . The estimates $\hat{\psi}_1, \hat{\psi}_2, \dots$ are computed through the characterization $\int_{\mathcal{M}} \hat{K}(p, q) \hat{\psi}_j(p) dp = \hat{\kappa}_j \hat{\psi}_j(q)$, which is solved by the discretization of the problem on a fine grid or by the basis expansion of estimated smooth functions. In the case where the domain is an interval of the real line, a theoretical study on the accuracy of $\hat{\psi}_j$ as an estimate of ψ_j is offered for example in [Hall and Hosseini-Nasab \(2006\)](#).

Define the $n \times s$ matrix $\mathbf{X} = (x_i(p_j))$, the column vector $\boldsymbol{\mu} = (\frac{1}{n} \sum_{i=1}^n x_i(p_j))$ of length s , the $n \times M$ matrix $\mathbf{A} = (\langle X_i, \phi_m \rangle)$ and the $s \times M$ matrix $\boldsymbol{\Phi} = (\phi_m(p_j))$. Let $\mathbf{1}$ denote the column vector of length n with all entries equal to 1. The empirical counterpart of the objective function in (2) becomes

$$\frac{1}{n} \|\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T - \mathbf{A}\boldsymbol{\Phi}^T\|_F^2, \quad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm, defined as the square root of the sum of the squares of its elements. This last formulation gives a natural way to deal with the fact that only pointwise and noisy evaluations $x_i(p_j)$, $i = 1, \dots, n$, $j = 1, \dots, s$ of the underlying functional samples are usually available. However, it does not incorporate any information on the smoothness of the functional data. In fact, considering the Singular Value Decomposition (SVD) of $\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T$, it can be shown that the minimizing arguments of (3) are $\hat{\boldsymbol{\Phi}} = \mathbf{V}$ and $\hat{\mathbf{A}} = \mathbf{U}\mathbf{D}$, thus the obtained formulation is a multivariate PCA applied to the data-matrix \mathbf{X} .

The regularized PCA approach consists on adding a penalization term to the classic formulation of the PCA, in order to recover a desired feature of the estimated underlying functions. In particular the formulation (3) has shown a great flexibility for this purpose. Examples of models where a sparseness property is assumed on the data are offered for instance in [Jolliffe et al. \(2003\)](#); [Zou and Hastie \(2005\)](#); [Shen and Huang \(2008\)](#). In the specific case of functional data analysis the penalization term usually encourages the PC functions to be smooth. Examples of PCA models that explicitly incorporates a smoothing penalization term are given by [Rice and Silverman \(1991\)](#); [Silverman \(1996\)](#); [Huang et al. \(2008\)](#). The cited works deal with functions whose domain is a limited interval in \mathbb{R} . [Zhou and Pan \(2014\)](#) recently proposed a smooth FPCA for two-dimensional functions on irregular planar domains. Their approach is based on a mixed effects model that specifies the PC functions as bivariate splines on triangulations

and the PC scores as random effects. Here we propose a FPCA model that can handle real functions observable on a two-dimensional manifold. In particular we shall consider a smoothing penalty operator, coherent with the 2D geodesic distances on the manifold. This leads to the definition of a model that can fully exploit the information about the geometry of the manifold.

3 Smooth FPCA over two-dimensional manifolds

Suppose the sample of n functions $x_i : \mathcal{M} \rightarrow \mathbb{R}$ is observed at a fixed set of points p_1, \dots, p_s in \mathcal{M} (this will be relaxed later). Let f be a real valued and twice differentiable function and $\mathbf{u} = \{u_i\}_{i=1, \dots, n}$ a n -dimensional real column vector. We propose to estimate the first PC function $\hat{f} : \mathcal{M} \rightarrow \mathbb{R}$ and the associated PC scores vector $\hat{\mathbf{u}}$, by solving the equation:

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^s (x_i(p_j) - u_i f(p_j))^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f \quad (4)$$

where $\Delta_{\mathcal{M}}$ is the Laplace-Beltrami operator for functions defined over the manifold \mathcal{M} . The Laplace-Beltrami operator is a generalization of the standard Laplacian to the case of functions defined on surfaces in Euclidean spaces. It is related to the local curvature of f on \mathcal{M} . The parameter λ controls the trade-off between the empirical term of the objective function and roughness penalizing term. The $\mathbf{u}^T \mathbf{u}$ term is justified by some invariance considerations on the objective function as done in the case of one dimensional domains, in [Huang et al. \(2008\)](#). Consider the transformation $(\mathbf{u} \rightarrow c\mathbf{u}, f \rightarrow \frac{1}{c}f)$, with c a constant, and the transformation $(\mathbf{X} \rightarrow c\mathbf{X}, \mathbf{u} \rightarrow c\mathbf{u})$, where $\mathbf{X} = (x_i(p_j))$. Then the objective function in (4) is invariant with respect to the first transformation, while the empirical and the smoothness terms are re-scaled by the same coefficient with the second transformation.

The subsequent PCs can be extracted sequentially by removing the preceding estimated components from the data matrix \mathbf{X} . This allows the selection of a different penalization parameter λ for each PC estimate. We will refer to this model as Smooth-Manifold FPCA (SM-FPCA).

3.1 Iterative algorithm

Here we present the numerical algorithm for the resolution of the model introduced above. Our approach for the minimization of the functional (4) can be summarized in the following two steps:

- Splitting the optimization in a finite dimensional optimization in \mathbf{u} and an infinite-dimensional optimization in f ;
- Approximating the infinite-dimensional solution thanks to a Surface Finite Element discretization.

Let \mathbf{f}_s be the vector of length s such that $\mathbf{f}_s = (f(p_1), \dots, f(p_s))^T$. The expression in (4) can be rewritten as

$$(\hat{\mathbf{u}}, \hat{f}) = \operatorname{argmin}_{\mathbf{u}, f} \|\mathbf{X} - \mathbf{u}\mathbf{f}_s^T\|_F^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \quad (5)$$

A normalization constraint must be considered in this minimization problem to make the representation unique, as in fact multiplying \mathbf{u} by a constant and dividing f by the same constant does not change the objective function. For simplicity of implementation, we here optimize in \mathbf{u} under the constraint $\|\mathbf{u}\|_2 = 1$, and leave the infinite-dimensional optimization in f unconstrained.

Our proposal for the minimization of the criterion (5) is to alternate the minimization of \mathbf{u} and f in an iterative algorithm:

Step 1 Estimation of \mathbf{u} given f . For a given f , the minimizing \mathbf{u} of the objective function in (5) is

$$\mathbf{u} = \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{f}_s\|_2^2 + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f}, \quad (6)$$

and the minimizing unitary-norm vector \mathbf{u} is

$$\mathbf{u} = \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{X}\mathbf{f}_s\|_2}. \quad (7)$$

Step 2 Estimation of f given \mathbf{u} . For a given \mathbf{u} , solving (5) with respect to f is equivalent to finding the f that minimizes

$$J_{\lambda, \mathbf{u}}(f) = \mathbf{f}_s^T \mathbf{f}_s + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f - 2\mathbf{f}_s^T \mathbf{X}^T \mathbf{u}. \quad (8)$$

Step 1 is basically the classical expression of the score vector given the loadings vector, where in this case the loading vector is given by \mathbf{f}_s , the evaluations of the PC function in p_1, \dots, p_s . The problem in Step 2 is not trivial, consisting in an infinite-dimensional minimization problem. Let z_j denote the j th element of the vector $\mathbf{X}^T \mathbf{u}$, then minimizing the functional in (8) is equivalent to minimizing

$$\sum_{j=1}^s \left(z_j - f(p_j) \right)^2 + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f \quad (9)$$

This problem consists in estimating a smooth scalar field f defined on a manifold, starting from noisy observations z_j at points p_j . In the special case where \mathcal{M} is a sphere or a sphere-like surface, that is $\mathcal{M} = \{\sigma(v) = \rho(v)v : v \in S\}$ where $S \subset \mathbb{R}^3$ is the unit sphere centered at the origin and, this smoothing problem has been considered for instance by Wahba (1981) and Alfeld et al. (1996). Moreover, the functional (9) is considered

for instance by [Ettinger et al. \(2015\)](#) and [Dassi et al. \(2015\)](#). Here \mathcal{M} is respectively a manifold homeomorphic to an open ended cylinder and a manifold homeomorphic to a sphere. In the later two works the field f is estimated by first conformally recasting the problem to a planar domain and then discretizing it by means of planar finite elements, generalizing the planar smoothing model in [Ramsay \(2002\)](#). Our approach is also based on a Finite Element (FE) discretization, but differently from [Ettinger et al. \(2015\)](#) and [Dassi et al. \(2015\)](#), we construct here a FE space directly on the triangulated surface $\mathcal{M}_{\mathcal{T}}$ that approximates the manifold \mathcal{M} , i.e. we use surface FE, avoiding any flattening step and allowing the formulation to be applicable to any manifold topology.

3.2 Surface Finite Element discretization

Assume, for clarity of exposition only, that \mathcal{M} is a closed surface, as in our motivating application. The case of non-closed surfaces can be dealt with by considering some appropriate boundary conditions as done for instance in the planar case in [Sangalli et al. \(2013\)](#). Consider the linear functional space $H^2(\mathcal{M})$, the space of functions in $L^2(\mathcal{M})$ with first and second weak derivatives in $L^2(\mathcal{M})$. The infinite dimensional part of the estimation problem can be reformulated as follows: find $\hat{f} \in H^2(\mathcal{M})$ such that

$$\hat{f} = \underset{f \in H^2(\mathcal{M})}{\operatorname{argmin}} J_{\lambda, \mathbf{u}}(f). \quad (10)$$

Proposition 1. *The solution $\hat{f} \in H^2(\mathcal{M})$ exists and is unique and is such that*

$$\sum_{j=1}^s \varphi(p_j) \hat{f}(p_j) + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}} \varphi \Delta_{\mathcal{M}} \hat{f} = \sum_{j=1}^s \varphi(p_j) \sum_{i=1}^n x_i(p_j) u_i \quad (11)$$

for every $\varphi \in H^2(\mathcal{M})$.

As detailed in the Supplementary Material, the key idea is to minimize $J_{\lambda, \mathbf{u}}(f)$ by differentiating this functional with respect to f . This leads to (11), that characterizes the estimate \hat{f} as the solution of a linear fourth-order problem.

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} . Figure 1 for instance shows the triangulated surface approximating the left hemisphere of a template brain. We then consider the linear finite element space V consisting in a set of globally continuous functions over $\mathcal{M}_{\mathcal{T}}$ that are linear affine where restricted to any triangle τ in \mathcal{T} , i.e.

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

This space is spanned by the nodal basis ψ_1, \dots, ψ_K associated to the nodes ξ_1, \dots, ξ_K , corresponding to the vertices of the triangulation $\mathcal{M}_{\mathcal{T}}$. Such basis functions are lagrangian, meaning that $\psi_i(\xi_j) = 1$ if $i = j$ and $\psi_i(\xi_j) = 0$ otherwise. Setting $\mathbf{f} = (f(\xi_1), \dots, f(\xi_K))^T$ and $\boldsymbol{\psi} = (\psi_1, \dots, \psi_K)^T$, every function $f \in V$ has the form

$$f(p) = \sum_{k=1}^K f(\xi_k) \psi_k(p) = \mathbf{f}^T \boldsymbol{\psi}(p) \quad (12)$$

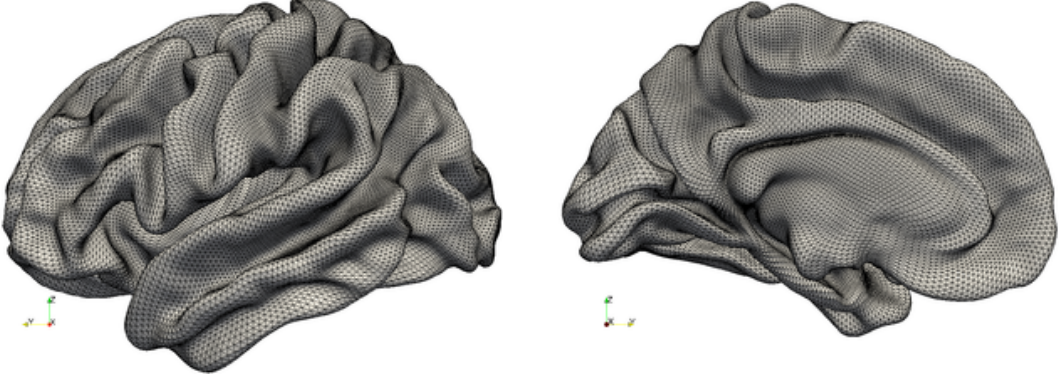


Figure 1: The triangulated surface approximating the left hemisphere of the template brain. The mesh is composed by 32K nodes and by 64K triangles

for each $p \in \mathcal{M}_\tau$. The surface finite element space provides a finite dimensional subspace of $H^1(\mathcal{M})$ [Dziuk (1988)]. To use this finite element space to discretize the infinite-dimensional problem (11), that is well posed in $H^2(\mathcal{M})$, we first need a reformulation of (11) that only involves first order derivatives. This can be obtained by introducing an auxiliary function $g = \Delta_\mathcal{M} f$, splitting the equation (11) into a coupled system of second order problems and finally integrating by parts the second order terms. The details of this derivation can be found in the supplementary material. The discrete estimators $\hat{f}_h, \hat{g}_h \in V$ are then obtained solving

$$\begin{cases} \int_{\mathcal{M}_\tau} \nabla_{\mathcal{M}_\tau} \hat{f}_h \nabla_{\mathcal{M}_\tau} \varphi_h - \int_{\mathcal{M}_\tau} \hat{g}_h \varphi_h = 0 \\ \lambda \int_{\mathcal{M}_\tau} \nabla_{\mathcal{M}_\tau} \hat{g}_h \nabla_{\mathcal{M}_\tau} v_h + \sum_{j=1}^s \hat{f}_h(p_j) v_h(p_j) = \sum_{j=1}^s v_h(p_j) \sum_{i=1}^n x_i(p_j) u_i \end{cases} \quad (13)$$

for all $\varphi_h, v_h \in V$. Define the $s \times K$ matrix $\Psi = (\psi_k(p_j))$ and the $K \times K$ matrices $\mathbf{R}_0 = \int_{\mathcal{M}_\tau} (\psi \psi^T)$ and $\mathbf{R}_1 = \int_{\mathcal{M}_\tau} (\nabla_{\mathcal{M}_\tau} \psi)(\nabla_{\mathcal{M}_\tau} \psi)^T$. Then exploiting the representation (12) of functions in V we can rewrite (13) as a linear system. Specifically the Finite Element solution $\hat{f}_h(p)$ of the discrete counterpart (13) is given by $\hat{f}_h(p) = \psi(p)^T \hat{\mathbf{f}}$ where $\hat{\mathbf{f}}$ is the solution of

$$\begin{bmatrix} \Psi^T \Psi & \lambda \mathbf{R}_1 \\ \lambda \mathbf{R}_1 & -\lambda \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \Psi^T \mathbf{X}^T \mathbf{u} \\ \mathbf{0} \end{bmatrix} \quad (14)$$

Solving (14) leads to

$$\hat{\mathbf{f}} = (\Psi^T \Psi + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1} \Psi^T \mathbf{X}^T \mathbf{u}. \quad (15)$$

Although this last formula is a compact expression of the solution, it is preferable to compute the solution from the linear system (14) due to the sparsity property of its left-hand side. As an example, in the simulations and the application shown in Sections 4-5, respectively less than 1% and less than 0.1% of the elements in the matrix in the left hand side of (14) are different from zero, allowing a very efficient solution of the linear system.

In the model introduced, we assume that all the observed functions $x_i : \mathcal{M} \rightarrow \mathbb{R}$ are sampled on the common set of points $p_1, \dots, p_s \in \mathcal{M}$. Suppose moreover, $p_1, \dots, p_s \in \mathcal{M}$

coincide with the vertices of the approximating triangulated surface $\mathcal{M}_{\mathcal{T}}$. In this particular case, an alternative approach could consist of interpreting the points $p_1, \dots, p_s \in \mathcal{M}_{\mathcal{T}}$ as the nodes of a graph linked by the edges of the triangulation and considering the model (4) with a discrete smoothness operator term instead of the Laplace-Beltrami operator (see e.g. [Belkin and Niyogi \(2001\)](#) for the choice of the penalization term and [Deng Cai et al. \(2011\)](#) for an application to matrix decomposition). However, thanks to its functional nature, the formulation (4) can be easily extended to the case of missing data or sparsely sampled functional data. Suppose now each function x_i is observable on a set of points $p_1^i, \dots, p_{s_i}^i$, then the natural extension of the model (4) becomes

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^{s_i} (x_i(p_j^i) - u_i f(p_j^i))^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f.$$

Following the same procedure, we can achieve an analogous algorithm based on the following two steps.

Step 1 Estimation of the unitary-norm vector \mathbf{u} given f .

$$\mathbf{u} \text{ such that } u_i = \frac{\sum_{j=1}^{s_i} x_i(p_j^i) f(p_j^i)}{\sqrt{\sum_{i=1}^n (\sum_{j=1}^{s_i} x_i(p_j^i) f(p_j^i))^2}}.$$

Step 2 Estimation of f given \mathbf{u} .

$f = \mathbf{f}^T \boldsymbol{\psi}$ with \mathbf{f} such that

$$\begin{bmatrix} \mathbf{L} & \lambda \mathbf{R}_1 \\ \lambda \mathbf{R}_1 & -\lambda \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{D}^T \mathbf{u} \\ \mathbf{0} \end{bmatrix}$$

where

$$\mathbf{L} = \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^{s_i} u_i^2 \psi_1(p_j^i) \psi_1(p_j^i) & \dots & \sum_{i=1}^n \sum_{j=1}^{s_i} u_i^2 \psi_1(p_j^i) \psi_K(p_j^i) \\ \dots & \dots & \dots \\ \sum_{i=1}^n \sum_{j=1}^{s_i} u_i^2 \psi_K(p_j^i) \psi_1(p_j^i) & \dots & \sum_{i=1}^n \sum_{j=1}^{s_i} u_i^2 \psi_K(p_j^i) \psi_K(p_j^i) \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} \sum_{j=1}^{s_1} \psi_1(p_j^1) x_1(p_j^1) & \dots & \sum_{j=1}^{s_n} \psi_1(p_j^n) x_n(p_j^n) \\ \dots & \dots & \dots \\ \sum_{j=1}^{s_1} \psi_K(p_j^1) x_1(p_j^1) & \dots & \sum_{j=1}^{s_n} \psi_K(p_j^n) x_n(p_j^n) \end{bmatrix}$$

3.3 SM-FPCA Algorithm

The algorithm for the resolution of the model SM-FPCA (4) can be summarized in the following steps.

Algorithm 1 SM-FPCA Algorithm

1: Initialization:

(a) Computation of Ψ , \mathbf{R}_0 and \mathbf{R}_1

(b) Perform the SVD: $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$

(c) $\mathbf{f}_s \leftarrow \mathbf{V}[:, 1]$, where $\mathbf{V}[:, 1]$ are the loadings of the first PC

2: Scores estimation:

$$\mathbf{u} \leftarrow \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{X}\mathbf{f}_s\|_2}$$

3: PC function's estimation: \mathbf{f} such that

$$\begin{bmatrix} \Psi^T \Psi & \lambda \mathbf{R}_1 \\ \lambda \mathbf{R}_1 & -\lambda \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \Psi^T \mathbf{X}^T \mathbf{u} \\ \mathbf{0} \end{bmatrix}$$

4: PC function's evaluation:

$$\mathbf{f}_s \leftarrow \Psi^T \mathbf{f}$$

5: Repeat Steps 2–4 until convergence

6: Normalization:

$$\hat{f}(p) \leftarrow \frac{\mathbf{f}^T \boldsymbol{\psi}(p)}{\|\mathbf{f}^T \boldsymbol{\psi}\|_{L^2(\mathcal{M}_T)}}$$

The model (4) is a non-convex minimization problem in (\mathbf{u}, f) . However, in the previous section we proved the existence and uniqueness of the minimizing f given \mathbf{u} and vice-versa. This implies that the objective function is non-increasing under the update rules of the Algorithm 1. Since the first guess of the PC function, given by the SVD, is usually a good starting point, in all our simulations no convergence problem has been detected.

3.4 Parameters selection

The proposed models have a smoothing parameter $\lambda > 0$ that adjusts the trade-off between the fidelity of the estimate to the data via the sum of the squared errors and the smoothness of the solution via the penalty term. The problem of choosing the smoothing parameter is common to all smoothing problems.

The flexibility given by the smoothing parameter can be seen as an advantageous feature; by varying the smoothing parameter the data can be explored on different scales. However, in many cases a data-driven automatic method is necessary. In the following simulations we consider two different criteria. The first approach consists on a K -fold cross validation. The data matrix \mathbf{X} is partitioned by rows into K roughly equal groups. For each group of data $k = 1, \dots, K$ the dataset can be split into a validation set \mathbf{X}^k , composed of the elements of the k th group, and a training set, composed of the remaining elements. For different smoothing parameters, the loading function f^{-k} is estimated from the training dataset. Given the estimated loading function f^{-k} , the

associated score vector \mathbf{u}^k is computed on the validation dataset. Since f^{-k} has been computed on the training dataset, \mathbf{u}^k should be computed on the validation dataset via the formula (6), where $\int_{\mathcal{M}} \Delta_{\mathcal{M}}^2$ can be approximated by $\mathbf{g}^T \mathbf{R}_0 \mathbf{g}$, being $g_h(p) = \psi(p)^T \mathbf{g}$ the auxiliary function approximating $\Delta_{\mathcal{M}} f$. Finally we select the value of the parameter λ that minimizes the following score:

$$CV(\lambda) = \sum_{k=1}^K \frac{\sum_{i=1}^n \sum_{j=1}^s x_i(p_j) - u_i^k f^{-k}(p_j))^2}{np}. \quad (16)$$

The second approach is based on the minimization of a generalized cross-validation (GCV) criteria integrated on the regression step of the iterative algorithm. Denoting $\mathbf{S}(\lambda) = \Psi^T (\Psi^T \Psi + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1} \Psi^T$ the GCV score is defined as

$$\text{GCV}(\lambda) = \frac{1}{s} \frac{\|(\mathbf{I} - \mathbf{S}(\lambda))(\mathbf{X}^T \mathbf{u})\|^2}{(1 - \frac{1}{s} \text{tr}\{\mathbf{S}(\lambda)\})^2}.$$

While the K -fold approach is generally slower, this does not require the inversion of any matrix. This is an advantageous feature, since generally the inverse of sparse matrix is not sparse. It is thus applicable also to datasets \mathbf{X} with a large number of columns s .

Another parameter that must be chosen in the application of the algorithm is the number of PCs to be considered for the dimension reduction of the data. A classical approach is to select this parameter on the basis of cumulated explained variance of the PC. While in the ordinary PC, the scores vectors are uncorrelated and their loadings are orthogonal, in our formulation neither the loadings are explicitly imposed to be orthogonal nor the PC scores to be uncorrelated. Let $\hat{\mathbf{U}}$ be the $n \times k$ matrix such that the columns of $\hat{\mathbf{U}}$ are the first k PC scores vectors. Without the uncorrelation assumption, it is meaningless to compute the total variance explained by the first k PCs by $\text{tr}(\hat{\mathbf{U}}^T \hat{\mathbf{U}})$. To overcome this problem Zou et al. (2004) propose to remove linear dependence between correlated PC scores vectors, by regression projection. Thus they compute the QR decomposition of $\hat{\mathbf{U}}$ as $\hat{\mathbf{U}} = \mathbf{Q}\mathbf{R}$ and define the *adjusted total variance* as $\sum_{j=1}^k \mathbf{R}_{jj}^2$.

4 Simulation studies

In this section we conduct simulations to assess the performance of the SM-FPCA algorithm as compared to other methods.

We consider as domain of the functional observations a triangulated surface $\mathcal{M}_{\mathcal{T}}$ with 642 nodes that approximates the brainstem. On this triangulated surface we generate the orthonormal functions v_1 and v_2 , consisting in two eigenfunctions of the Laplace-Beltrami operator, as shown in Figure 2. These represent respectively the first and second PC functions. We then generate $n = 50$ smooth functions x_1, \dots, x_{50} on $\mathcal{M}_{\mathcal{T}}$ by

$$x_i = u_{i1}v_1 + u_{i2}v_2 \quad i = 1, \dots, n, \quad (17)$$

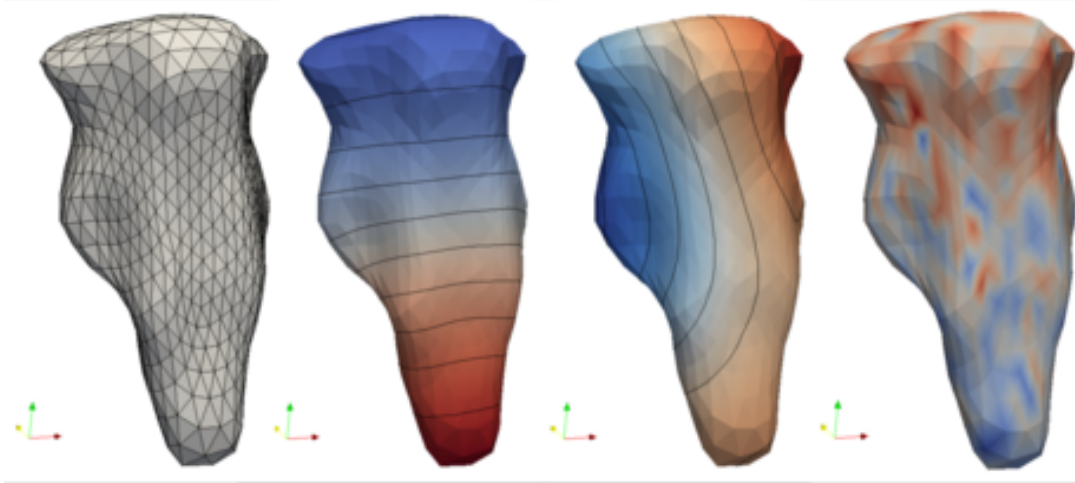


Figure 2: From left to right, the triangulated surface representing the brainstem, a plot of the true first and second PC functions and a plot of a noisy observation on the brainstem.

where u_{i1} , u_{i2} are independent random variables, distributed as $u_{i1} \sim \mathcal{N}(0, \sigma_1^2)$, $u_{i2} \sim \mathcal{N}(0, \sigma_2^2)$, with $\sigma_1 = 4$ and $\sigma_2 = 2$. They represent the PC scores.

The smooth functions x_i are sampled at locations $p_j \in \mathbb{R}^3$ with $j = 1, \dots, s$ coinciding with the nodes of the triangulated surface. Moreover at each of these points has been added to the functions a Gaussian noise with mean zero and standard deviation $\sigma = 0.1$ to obtain the noisy observations denoted with $x_i(p_j)$. We are interested in recovering the smooth PC functions v_1 and v_2 from these noisy observations over $\mathcal{M}_{\mathcal{T}}$. We compare the proposed SM-FPCA technique to two alternative approaches.

The first basic approach we consider is a simple multivariate PCA (MV-PCA) applied to the data-matrix \mathbf{X} . The PC functions are thus obtained by piecewise linear interpolation over the mesh $\mathcal{M}_{\mathcal{T}}$. Finally they are normalized to have unitary norm in $L^2(\mathcal{M}_{\mathcal{T}})$.

A second natural approach is based on a pre-smoothing of the noisy observations that tries to recover the smooth functions x_i , $i = 1, \dots, n$ from their noisy observations $x_i(p_j)$, followed by a MV-PCA on the denoised evaluations of the functions on p_j , $j = 1, \dots, s$. The smoothing problem for a field defined on a Riemannian manifold is not trivial. In this case the smoothing technique applied is Iterated Heat Kernel (IHK) smoothing [Chung et al. (2005)]. The heat kernel smoothing of the noisy observation $x_i(p_j)$, is given by $K_\eta \times x_i(p_j) = \int_{\mathcal{M}} K_\eta(p, q) x_i(p_j) dq$, where η is the smoothing parameter and K_η is the heat kernel, whose analytic expression can be extracted from the eigenfunctions of the Laplace-Beltrami operator. However, for numerical approximation, it can be shown that for η small and for q close to p we have

$$K_\eta(p, q) \approx \frac{1}{(2\pi\eta)^{\frac{1}{2}}} \exp\left[-\frac{d^2(p, q)}{2\eta^2}\right].$$

The desired level of smoothing can be reached after k iterations, thanks to the following property: $K_\eta^k \times f = K_\eta \times \dots \times K_\eta \times f = K_{\sqrt{k}\eta}$. For a fixed bandwidth η , the level of smoothing is determined by an optimal number of iterations selected via the F-test criterion outlined in [Chung et al. \(2005\)](#). In these simulations the bandwidth has been set at $\eta = 2.5$, heuristically selecting the one with the best performance after some initial pilot studies. We will refer to this approach as IHK-PCA.

The proposed SM-FPCA technique is implemented as follows. For each PC we run Algorithm 1 with 15 iterations of the steps 2-4. For the choice of the optimal smoothing parameter λ , both K -fold, with $K = 5$, and GCV approaches have been applied.

The reconstructed PC functions, using the three different approaches are shown in Figures 3-4. It is evident that applying the MV-PCA yields to a reconstruction far from the true, because of the absence of any spatial information. The reconstruction through the IHK-PCA approach and the SM-FPCA model are considerably more satisfactory.

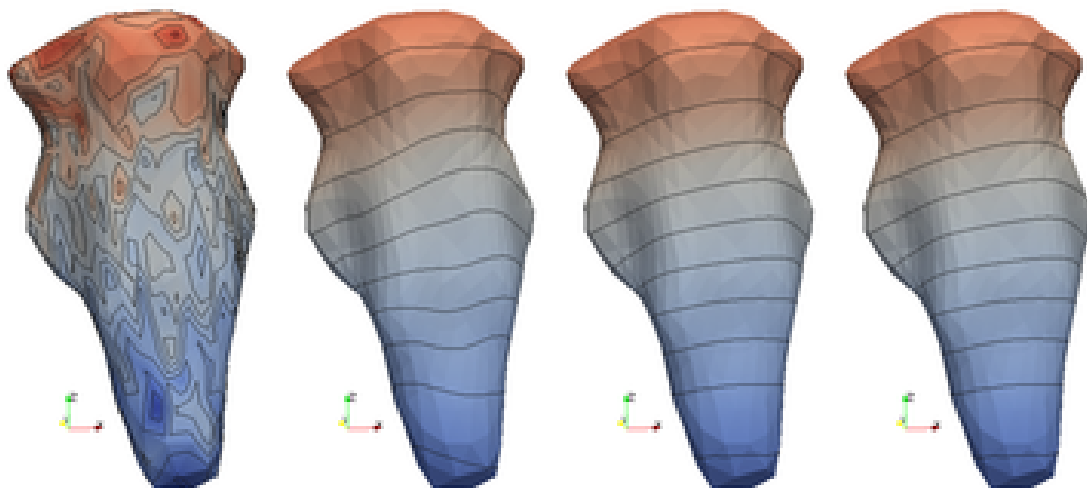


Figure 3: From left to right, contours of the first PC function reconstructed respectively with MV-PCA, IHK-PCA, SM-FPCA GCV and SM-FPCA K-fold. From a visual inspection, MV-PCA shows unsatisfactory results, while a better estimation is achieved by IHK-PCA and SM-FPCA. In particular SM-FPCA shows an almost optimal reconstruction of the contours' horizontal pattern in the original first PC function.

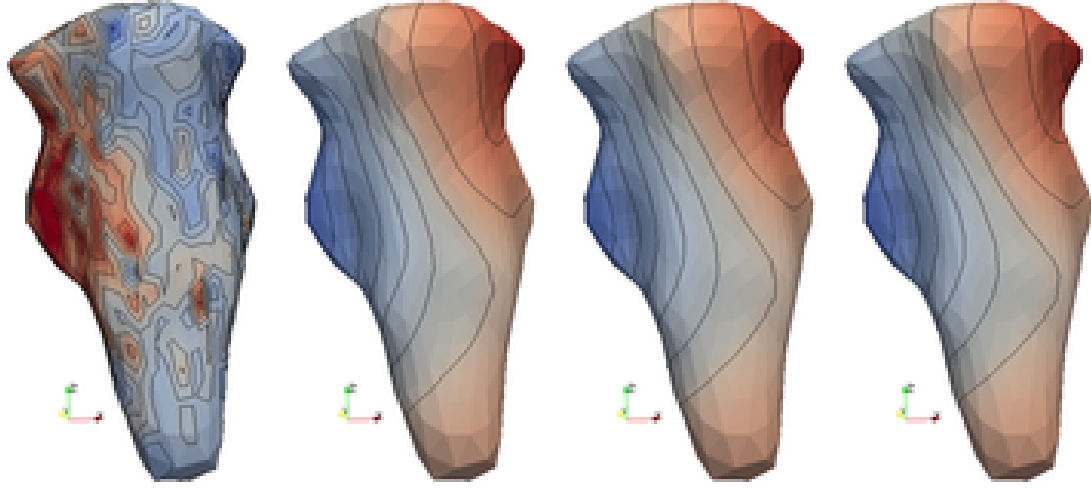


Figure 4: From the left to the right, contours of the second PC function reconstructed respectively with MV-PCA, IHK-PCA, SM-FPCA GCV and SM-FPCA K-fold. A visual inspection confirms that a better estimation is achieved by IHK-PCA and SM-FPCA respect to MV-PCA.

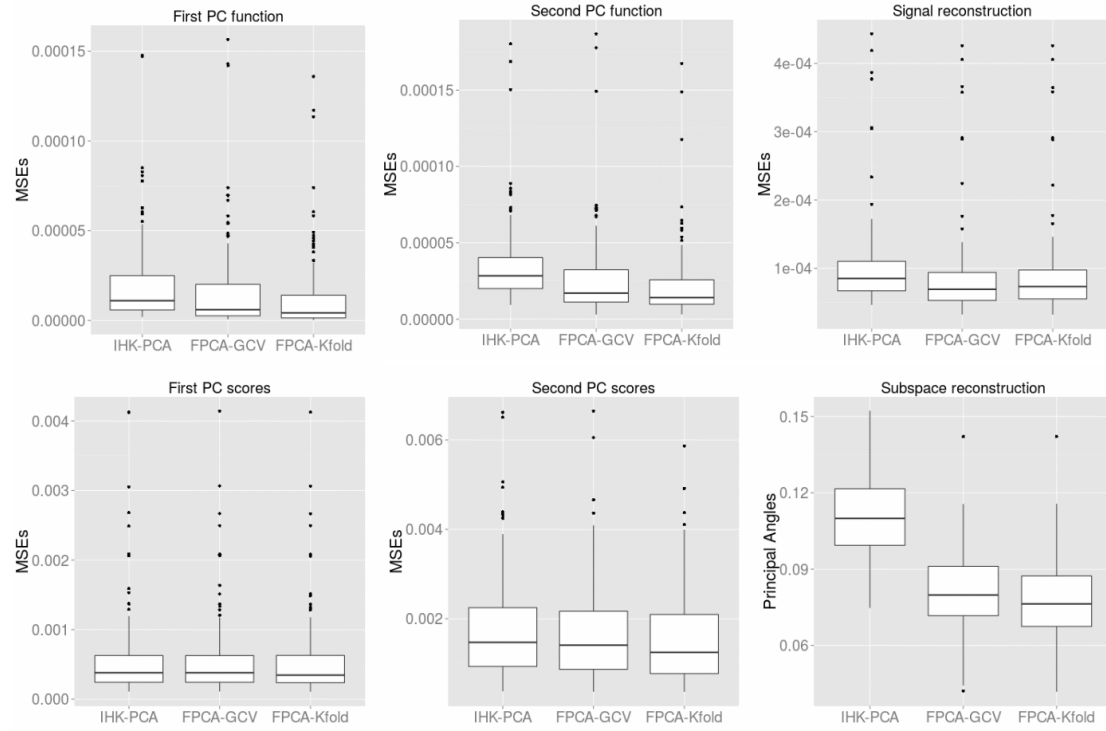


Figure 5: Boxplots summarizing the performance of IHK-PCA and SM-FPCA. For the SM-FPCA both GCV and K -fold have been applied for the selection of the smoothing parameter.

While the poor performance of the MV-PCA is evident, to assess the performance of the other two methods, we apply them to 100 datasets generated as previously detailed. The quality of estimated individual surfaces is then measured using the mean square error (MSE) over all the locations p_j , $j = 1, \dots, s$. MSEs are also used to evaluate the reconstruction of the PC scores vectors. Another performance measure used is the principal angle between the subspace spanned by the estimated PC functions and the subspace spanned by the true PC functions, as used in Shen and Huang (2008). Intuitively, the principal angle measures how similar two subspaces are. For this purpose we construct the $s \times 2$ matrices $\mathbb{V} = (v_i(p_j))$ and $\hat{\mathbb{V}} = (\hat{v}_i(p_j))$, where \hat{v}_i is the i th estimate of the true PC function v_i . Then we compute the orthonormal set of basis $\mathbf{Q}_{\mathbb{V}}$ and $\mathbf{Q}_{\hat{\mathbb{V}}}$ from the QR decomposition of \mathbb{V} and $\hat{\mathbb{V}}$. The principal angle is defined as the angle $\cos^{-1}(\rho)$, where ρ is the minimum singular value of $\mathbf{Q}_{\hat{\mathbb{V}}}^T \mathbf{Q}_{\mathbb{V}}$. The results are summarized in the boxplots in Figure 5, which compares the IHK-PCA and SM-FPCA algorithms with respect to the reconstruction's errors of the PC functions v_1 and v_2 , the PC scores $\mathbf{u}_1 = (u_{i1})$ and $\mathbf{u}_2 = (u_{i2})$, the reconstructed signals $x_i = u_{i1}v_1 + u_{i2}v_2$ and the principle angles between the subspaces spanned by the true and estimated PC functions.

The boxplots highlight the fact that SM-FPCA provides the best estimates of the PC functions and corresponding scores vectors. The K -fold approach to the choice of the smoothing parameter in the SM-FPCA model performs better than the GCV approach.

5 Application

The data set which we consider in this paper arises from the Human Connectome Project Consortium [Essen et al. (2012)], which is collecting data such as structural scans, resting-state and task-based functional MRI scans, and diffusion-weighted MRI scans from a large number of healthy volunteers to help elucidate the understanding of normal brain function. Many preprocessing considerations have already been resolved in the so called minimally preprocessed dataset. Among the various preprocessing pipelines applied to the HCP original data, of particular interest for us is the one named *fMRISurface* [Glasser et al. (2013)]. This pipeline provides a transformation of the 3D structural MRI and 4D signal from the functional MRI scan, so to enable the application of statistical analysis techniques on brain surfaces. For each subject, the personal cortical surface is extracted as a triangulated surface from the structural MRI and to each vertex of this mesh is associated a BOLD time-series derived from the BOLD signal of the underlying gray-matter ribbon. The cortical surfaces extracted are aligned to a template cortical surface generated from the cortical surfaces of 69 healthy adults. In practice this cortical surface is represented by two triangulated surfaces with 32k vertices, one for each hemisphere. In Figure 1 the left hemisphere is shown. Through this anatomical transformation map, the patients' BOLD time-series, on the cortical surface, are coherently located to the vertices of the template cortical surface. The fMRI signal used for our analysis has been acquired in absence of any task and for this reason is also called resting state fMRI. Finally each time-series is filtered to the band of frequencies $[0.009, 0.08]$ Hz.

As already mentioned in Section 1, a classic approach in the study of the resting state

fMRI, is to exploit the time dimension of the data, for the extraction of a connectivity measure among the different parts of the cortical surface. A standard choice for this purpose is the computation of the temporal correlation. It first consists of identifying a Region of Interest (ROI) on the cortical surface. For each subject, all the time-series in the ROI are used to find a representative mean time-series. To each vertex of the cortical surface we associate the pairwise correlation of the time-series located in that vertex with the time-series representative of the ROI. Finally each correlation value is transformed using Fisher's r-to-z transformation, yielding a resting state functional connectivity (RSFC) map for each subject. The total number of subjects considered for this analysis is 491.

For the choice of the ROI, we consider the cortical parcellation derived in [Gordon et al. \(2014\)](#), where a group-average boundary map of the cortical surface is derived from resting state fMRI (Figure 6). The identified cortical areas are unlikely to correspond the individual parcellation of each subject, since they are derived from a group average study. However, they can serve as a reasonable ROIs in individual subjects. The parcel that served as ROI in the following analysis is visible in red in Figure 6. For the chosen

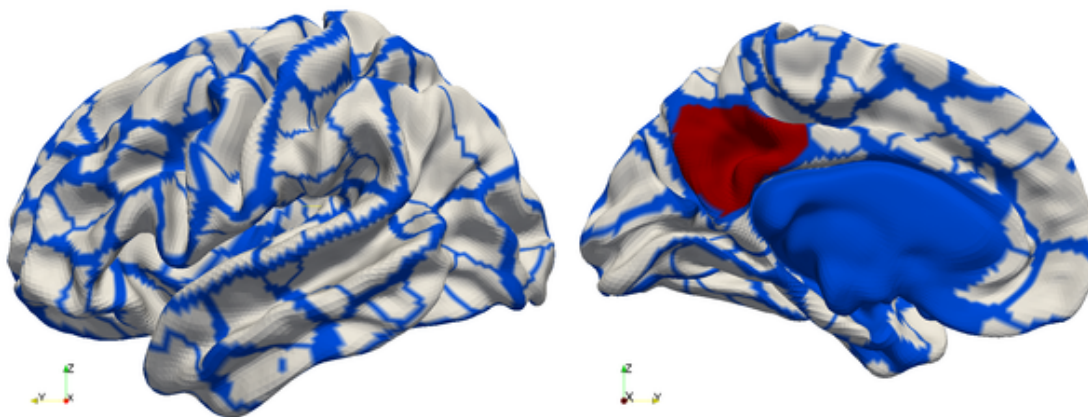


Figure 6: Parcellation of the cortical surface derived in [Gordon et al. \(2014\)](#). In red the Region of Interest chosen for the computation of the RSFC maps. This region is localized on an area of the cerebral cortex called precuneus.

ROI, a snapshot of the RSFC map of one subject is shown in Figure 7.

The mean RSFC map is shown in Figure 8. As expected high correlation values are visible inside the ROI. The mean RSFC over 491 subjects shows a variability coherent with the parcellation, in the sense that the vertices inside each parcel show similar values. We wish now to understand which are the main modes of variation of this RSFC maps among the different subjects, by applying a PCA.

The first three PC functions, estimated with SM-FPCA, are shown in Figures 9-10-11 as compared to the PC functions derived from MV-PCA and IHK-PCA. The choice of

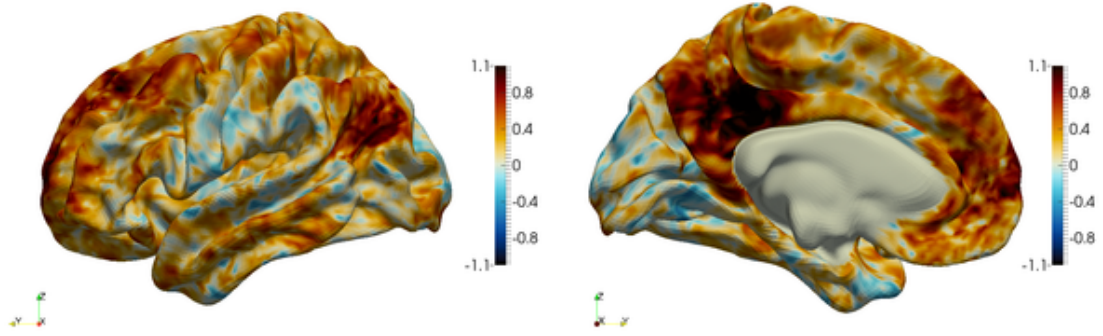


Figure 7: A snapshot of the RSFC map of one subject.

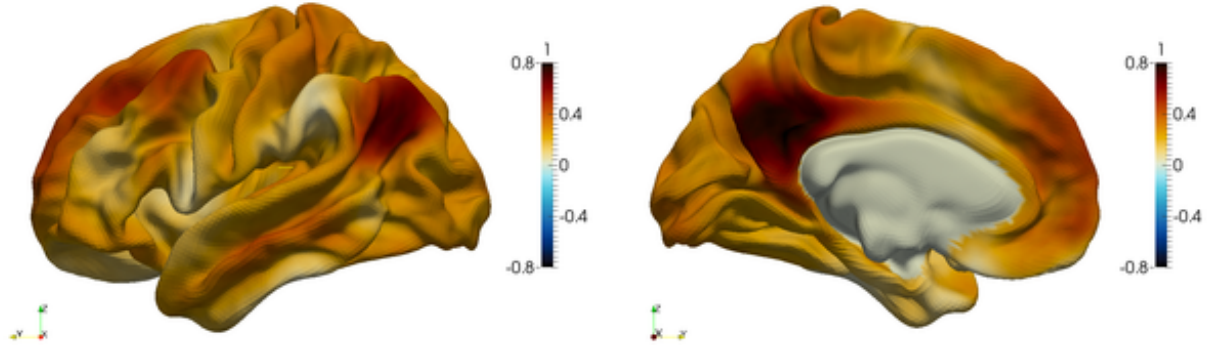


Figure 8: The mean RSFC map computed over 491 subjects. As expected, high correlation values are visible inside the ROI.

the smoothing parameter for the SM-FPCA is based on the K -fold cross validation, with $K = 5$.

The PC functions estimated from the MV-PCA shows an excessive variability, since the sample size is not sufficiently large to deal with the extremely high dimensionality of the data. In fact, even recent attempts to model the subject variability from resting state fMRI leads to the conclusion that spatial mismatches, introduced by the alignment problem, are one of the biggest sources of currently observable differences between subjects [Harrison et al. (2015)]. This registration process can result in misalignments, due to the lack of functional regions being perfectly coincident or due to situations where the local topology is strongly different among subjects. These misalignments can introduce fictitious effects on the computed PC functions. Data misalignment is a well known problem in FDA [Ramsay and Silverman (2005)]. For functional data with one-dimensional

domains, typical approaches are based on shifting or (monotone) transformations of the domain of each function. But neither shifting nor monotonic transformations make sense on a generic non-euclidean domain, so it is not clear how to generalize the standard FDA approaches. The introduction of a smoothing penalty in the PCA model should reduce the variability effects due to misalignment. In fact the smoothing parameter in the SM-FPCA algorithm can be seen as a further degree of freedom that allows a multiscale analysis, meaning that by increasing the smoothing penalty parameter is possible to constrain the results to show only the macroscopical effects of the phenomena and to remove the artifacts introduced by the preprocessing steps.

The PC functions estimated through IHK-PCA and SM-FPCA seem both to resolve the problem of the presence of some residual noise in the PC function estimates. However, we would like to emphasize the fact that, for instance, the third PC function computed with IHK-PCA shows some differences with the one computed with SM-FPCA. This is probably due to the fact that an individual pre-smoothing approach with the presented data tends to delete part of the information contained in the single RSFC map, while the SM-FPCA model incorporates all the noisy data as is and introduces a smoothness penalization directly on the PC function estimates. Contrary to MV-PCA and IHK-PCA, in all three PC functions, SM-FPCA shows a satisfactory level of smoothness, without deleting sharper changes in some locations, which may indicate the presence of a boundary between two cortical areas, since it is well known that the cortical surface is organized in a number of interacting cortical areas.

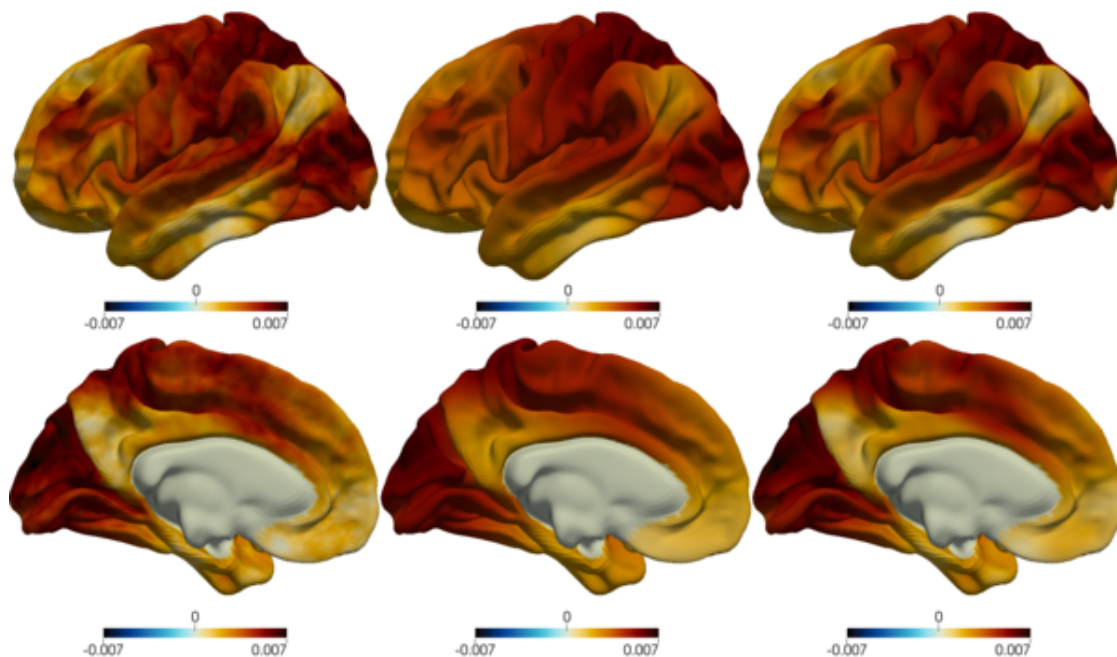


Figure 9: From left to right, two views of the first PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.

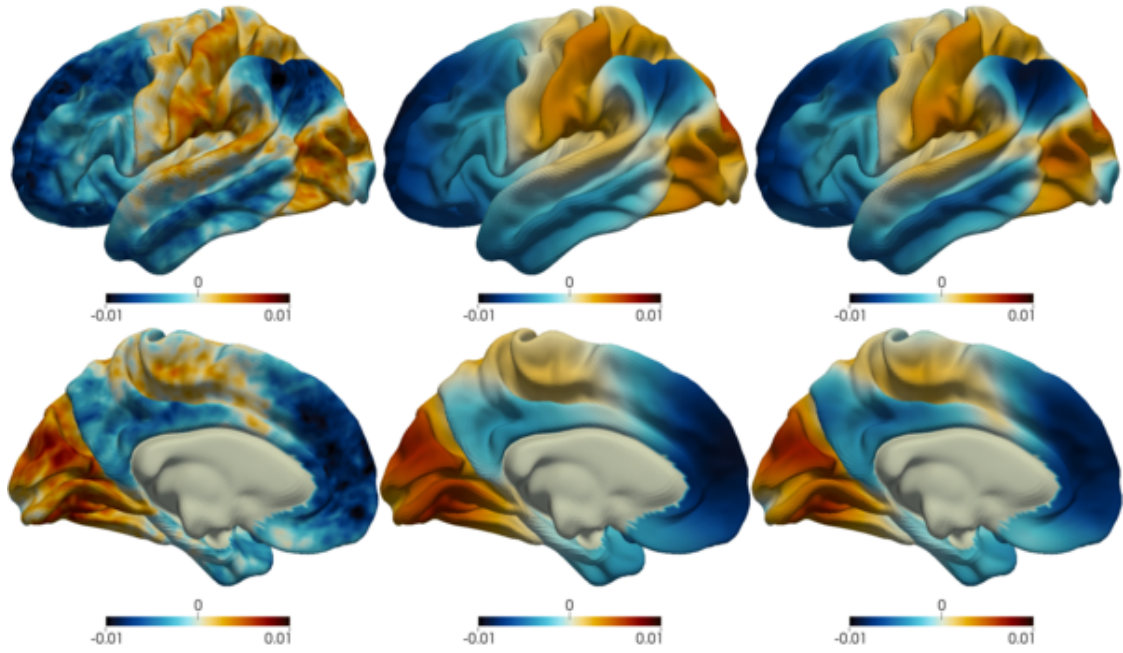


Figure 10: From left to right, two views of the second PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.

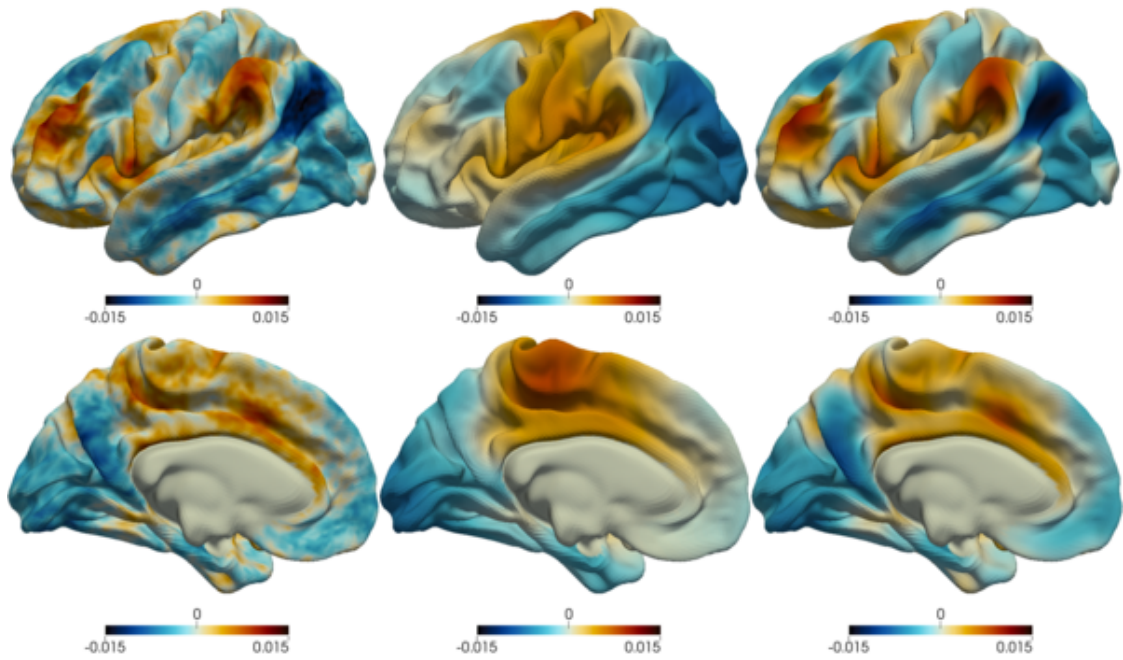


Figure 11: From left to right, two views of the third PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.

For the purpose of interpretation of the PC functions we might prefer to plot the functions $\mu \pm 2\sigma f$, where μ denotes the mean RSFC map, σ denotes the standard deviation of the PC scores vector and f denotes the associated PC function. In Figure 12 we show the described plot for the first PC function. We can observe that while the high correlation value in the ROI and inferior parietal are in first approximation preserved from subject to subject, a high variability between subjects can be observed in the areas surrounding the ROI and the inferior parietal, which is understood due to individual inter-subject differences [Buckner et al. (2008) and references therein].

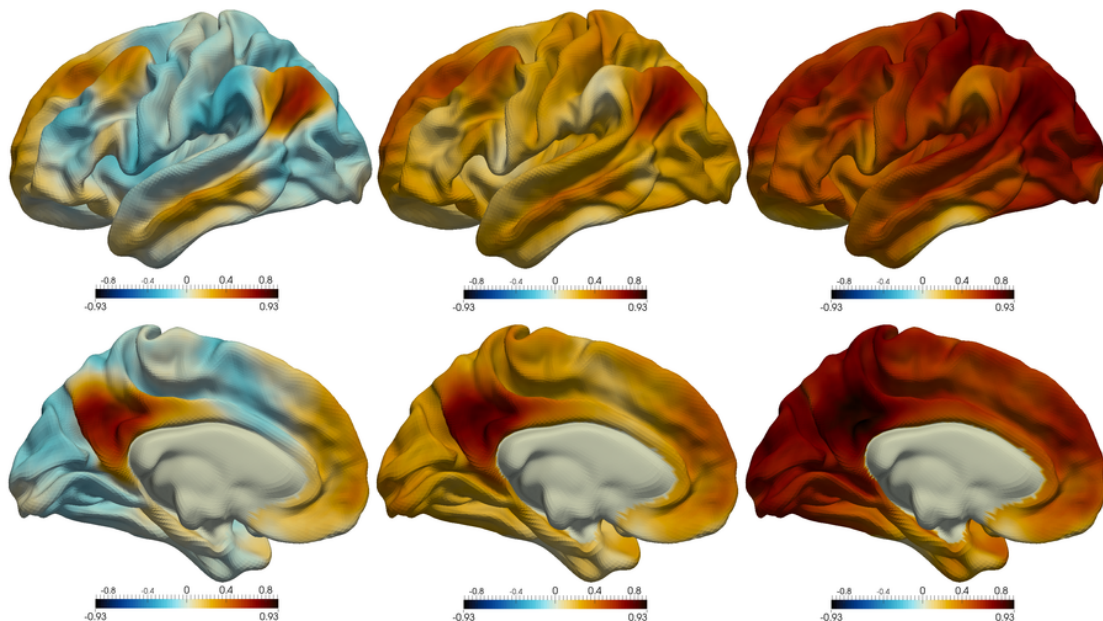


Figure 12: From left to right, two views of $\mu - 2\sigma f$, μ , $\mu + 2\sigma f$, where μ denotes the mean RSFC map, σ denotes the standard deviation of the first PC scores vector and f denotes the first PC function.

6 Discussion

In this paper we introduce a novel PCA technique that can handle functional data located over a two-dimensional manifold. The adopted approach is based on a regularized PCA model. In particular a smoothness penalty term that measures the curvature of a function over a manifold is introduced and the estimation problem is solved via an iterative algorithm that uses finite elements. The motivating application is the analysis the RSFC maps over the cortical surface, derived from fMRI. In this setting the adoption of a MV-PCA suffers of the high-dimensionality of the data with respect to the relatively small sample size. The adoption of an approach based on individual pre-smoothing of the functional samples, followed by a MV-PCA, gives smooth estimates of the PC functions; However, this pre-smoothing step tends to remove also part of useful information from

the original data. The proposed SM-FPCA instead returns smooth PC functions that nevertheless are able to capture localized features of the estimated PC function.

A further important feature of SM-FPCA is its computational efficiency. The most computationally intensive operation is the resolution of the linear system in the iterative algorithm. However this linear system enjoys two important properties. The first is the independence between its dimension, related to the number of nodes of the triangular mesh, and the number of point-wise observations available for each functional sample or the sample size. In fact, since its resolution time depends mostly on the mesh size, a mesh simplification approach [Dassi et al. (2015)] could be adopted to speed up the algorithm. The second and most fundamental property is the sparsity of the linear system. The use of a sparse solver allows an efficient computation of the solution. For instance, in the final application the dimension of the linear system is $64K \times 64K$. Despite its dimension, the solving time is less than a second. The application of the entire algorithm, for a fixed smoothing parameter, with 15 iterations is less than 15 seconds on a Intel Core i5-3470 3.20GHz workstation, with 4 GB of RAM.

A Appendices

A.1 Well-posedness of the estimation problem (10)

Proof. Proposition 1. We exploit a characterization theorem [Braess (2007), chapter 2] which states that if G is a symmetric, positive definite, bilinear form on a vector space L , and F is a linear functional on L , then v is the unique minimizer of

$$G(v, v) - 2F(v)$$

in V if and only if

$$G(v, \varphi) = F(\varphi) \quad \text{for all } \varphi \in L. \quad (18)$$

Moreover, there is at most one solution to problem (18).

The desired result follows from application of the above theorem considering the vector space $L = H^2(\mathcal{M})$, the symmetric, positive definite, bilinear form $G(f, \varphi) := \sum_{j=1}^p \varphi(p_j) f(p_j) + \lambda \int_{\mathcal{M}} \Delta \varphi \Delta f$ and the linear functional $F(f) = \sum_{j=1}^p f(p_j) \sum_{i=1}^n x_i(p_j) u_i$. Positive definiteness of the form G , in $H^2(\mathcal{M})$, is shown by the following argument. Suppose that $G(f, f) = 0$ for some $f \in H^2(\mathcal{M})$; then $\int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f = 0$ and $\sum_{j=1}^p f(p_j)^2 = 0$. Each element $f \in H^2(\mathcal{M})$ can be written such that, for any $p \in \mathcal{M}$, $f(p) = \tilde{f}(p) + c$, with $\tilde{f} \in U = \{\tilde{f} \in H^2(\mathcal{M}) : \int_{\mathcal{M}} \tilde{f} = 0\}$ and c a constant. The solution of $\Delta_{\mathcal{M}} \tilde{f} = 0$ in U exists unique and is $\tilde{f} = 0$ [Dziuk and Elliott (2013)]. Thus $\int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f = 0$ for $f \in H^2(\mathcal{M})$ implies that $f(p) = c$, for any $p \in \mathcal{M}$, then $\sum_{j=1}^p f(p_j)^2 = pc^2$. But $pc^2 = 0$ if and only if $c = 0$, so $f(\cdot) = 0$. Consequently, G is positive definite on $H^2(\mathcal{M})$.

The estimator \hat{f} is thus

$$\sum_{j=1}^p \varphi(p_j) \hat{f}(p_j) + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}} \varphi \Delta_{\mathcal{M}} \hat{f} = \sum_{j=1}^p \varphi(p_j) \sum_{i=1}^n x_i(p_j) u_i \quad (19)$$

for every $\varphi \in H^2(\mathcal{M})$.

□

A.2 Reformulation of the estimation problem

The problem of finding $f \in H^2(\mathcal{M})$ that satisfies condition (19) for every $\varphi \in H^2(\mathcal{M})$ can be rewritten as the problem of finding $(\hat{f}, g) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$ that satisfies:

$$\begin{cases} \sum_{j=1}^p \varphi(p_j) \hat{f}(p_j) + \lambda \int_{\mathcal{M}} (\Delta \varphi) g = \sum_{j=1}^p \varphi(p_j) \sum_{i=1}^n x_i(p_j) u_i \\ \int_{\mathcal{M}} v g - \int_{\mathcal{M}} v (\Delta \hat{f}) = 0 \end{cases} \quad (20)$$

for all $(\varphi, v) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$. In fact, if the pair of functions $(\hat{f}, g) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$ satisfies condition (20) for all $(\varphi, v) \in H^2(\mathcal{M}) \times L^2(\mathcal{M})$, then \hat{f} also satisfies problem (19). In contrast, if $\hat{f} \in H^2(\mathcal{M})$ satisfies problem (19), then the pair $(\hat{f}, \Delta \hat{f})$ automatically satisfies the two equations in problem (20). Owing to integration by part and to the fact that \mathcal{M} has no boundaries, we get:

$$\begin{aligned} \int_{\mathcal{M}} (\Delta_{\mathcal{M}} \varphi) g &= - \int_{\mathcal{M}} \nabla_{\mathcal{M}} \varphi \nabla_{\mathcal{M}} g \\ \int_{\mathcal{M}} v (\Delta_{\mathcal{M}} \hat{f}) &= - \int_{\mathcal{M}} \nabla_{\mathcal{M}} v \nabla_{\mathcal{M}} \hat{f} \end{aligned}$$

Now, asking the auxiliary function g and of the test functions v to be such that $g, v \in H^1(\mathcal{M})$, the problem of finding $\hat{f} \in H^2(\mathcal{M})$ that satisfies (19) for each $\varphi \in H^2(\mathcal{M})$ can be reformulated as finding $(\hat{f}, g) \in (H^1(\mathcal{M}) \cap C^0(\mathcal{M})) \times H^1(\mathcal{M})$

$$\begin{cases} \sum_{j=1}^p \varphi(p_j) \hat{f}(p_j) + \lambda \int_{\mathcal{M}} \nabla \varphi \nabla g = \sum_{j=1}^p \varphi(p_j) \sum_{i=1}^n x_i(p_j) u_i \\ \int_{\mathcal{M}} v g - \int_{\mathcal{M}} \nabla v \nabla \hat{f} = 0 \end{cases} \quad (21)$$

for all $(\varphi, v) \in (H^1(\mathcal{M}) \cap C^0(\mathcal{M})) \times H^1(\mathcal{M})$; Moreover, the theory of problems of elliptic regularity ensure that such \hat{f} still belongs to $H^2(\mathcal{M})$ [Dziuk and Elliott (2013) and reference therein]. Finally the discrete estimators $\hat{f}_h, \hat{g}_h \in V \subset H^1(\mathcal{M})$ are obtained solving

$$\begin{cases} \int_{\mathcal{M}_{\mathcal{T}}} \nabla_{\mathcal{M}_{\mathcal{T}}} \hat{f}_h \nabla_{\mathcal{M}_{\mathcal{T}}} \varphi_h - \int_{\mathcal{M}_{\mathcal{T}}} \hat{g}_h \varphi_h = 0 \\ \lambda \int_{\mathcal{M}_{\mathcal{T}}} \nabla_{\mathcal{M}_{\mathcal{T}}} \hat{g}_h \nabla_{\mathcal{M}_{\mathcal{T}}} v_h + \sum_{j=1}^s \hat{f}_h(p_j) v_h(p_j) = \sum_{j=1}^s v_h(p_j) \sum_{i=1}^n x_i(p_j) u_i \end{cases}$$

for all $\varphi_h, v_h \in V$. A generic function in V can be written as the linear combination of the finite number of basis spanning V . This allows the solution $\hat{f}_h(p) = \psi(p)^T \hat{\mathbf{f}}$ to be characterized by the linear system (14) in the original paper.

A.3 Simulation on the sphere

Here we present some further simulation studies on a domain \mathcal{M} that is a sphere centered on the origin and with radius $r = 1$, approximated by the triangulated surface $\mathcal{M}_{\mathcal{T}}$ in Figure 13.

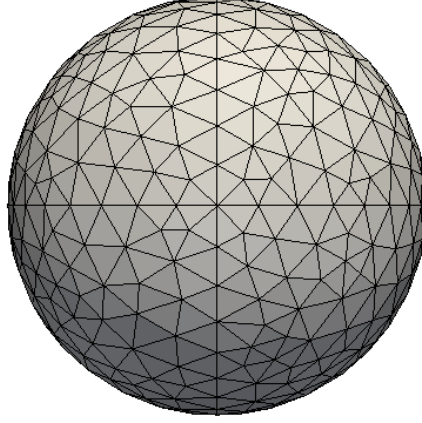


Figure 13: The triangulated surface approximating the sphere with 488 points.

We generate $n = 50$ smooth functions x_1, \dots, x_{50} on $\mathcal{M}_{\mathcal{T}}$ by

$$x_i = u_{i1}v_1 + u_{i2}v_2, \quad i = 1, \dots, n$$

where v_1 and v_2 represent the two PC functions with expressions

$$\begin{cases} v_1(x, y, z) = \frac{1}{2} \sqrt{\frac{15}{\pi}} \frac{xy}{r^2} \\ v_2(x, y, z) = \frac{3}{4} \sqrt{\frac{35}{\pi}} \frac{xy(x^2 - y^2)}{r^4} \end{cases}$$

and u_{i1} , u_{i2} represent the PC scores, generated independently and distributed as $u_{i1} \sim N(0, \sigma_1^2)$, $u_{i2} \sim N(0, \sigma_2^2)$ with $\sigma_1 = 4$, $\sigma_2 = 2$. The PC functions are two components of the Spherical Harmonics basis set, so they are orthonormal on the sphere, i.e. $\int_{\mathcal{M}} v_i^2 = 1$ for each $i \in \{1, 2\}$ and $\int_{\mathcal{M}} v_i v_k = 0$ for each $i \neq k$ with $i, k \in \{1, 2\}$. The PC functions are plotted in Figure 14. The functions x_i are sampled at locations coinciding with the nodes of the mesh in Figure 13. At these locations, a Gaussian white noise with standard deviation $\sigma = 0.1$ has been added to the true function x_i . We are then interested in recovering the smooth PC functions v_1 and v_2 from these noisy observations.

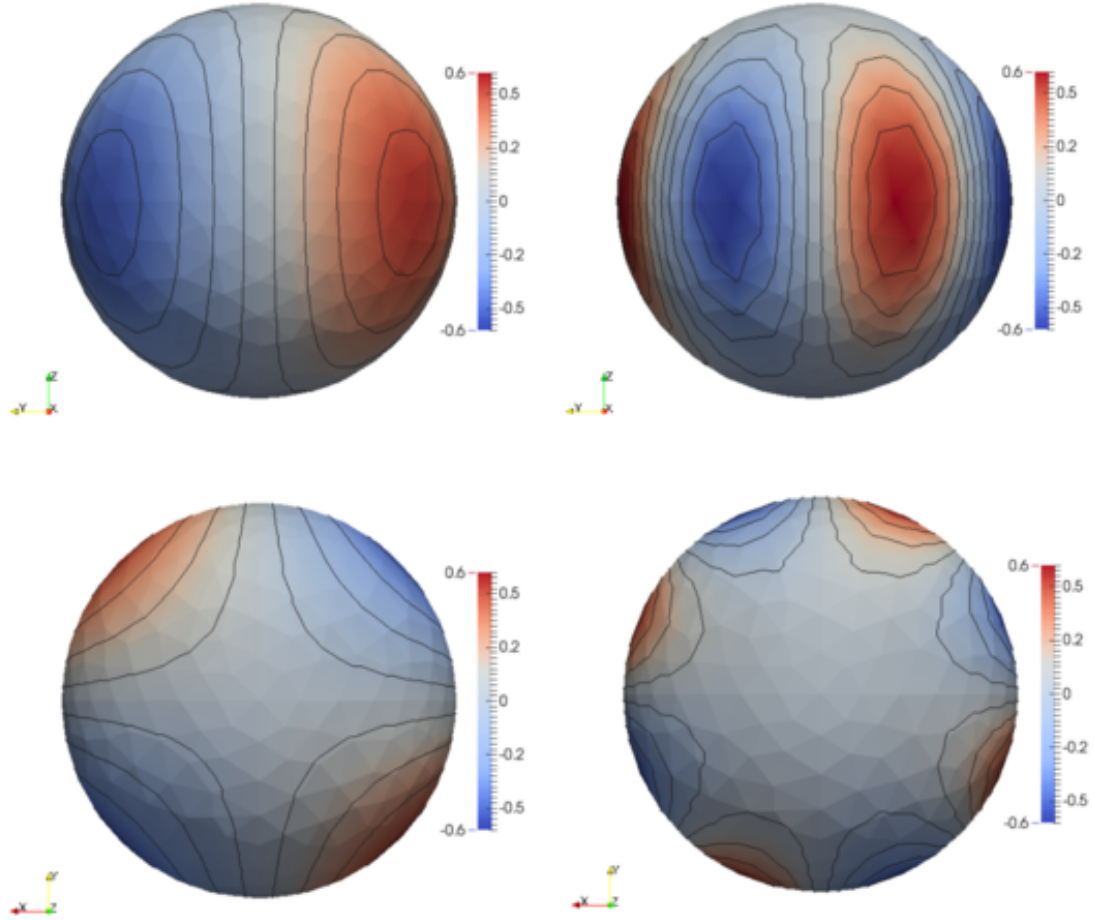


Figure 14: From the left to the right a plot of the true first and second PC functions.

We apply the proposed SM-FPCA method, choosing the optimal smoothing parameter λ , both with the K-fold and with GCV. We compare to the approach based on pre-smoothing followed by MV-PCA on the denoised evaluations of the functions at the locations p_j , $j = 1, \dots, p$. In this case, the smoothing techniques used is Spherical Splines [Wahba (1981)], using the implementation in the R package *mgcv*. The smoothing parameter choice is based on the GCV criterion. We will refer to this approach as SSpline-PCA. The results are summarized in Figure 15.

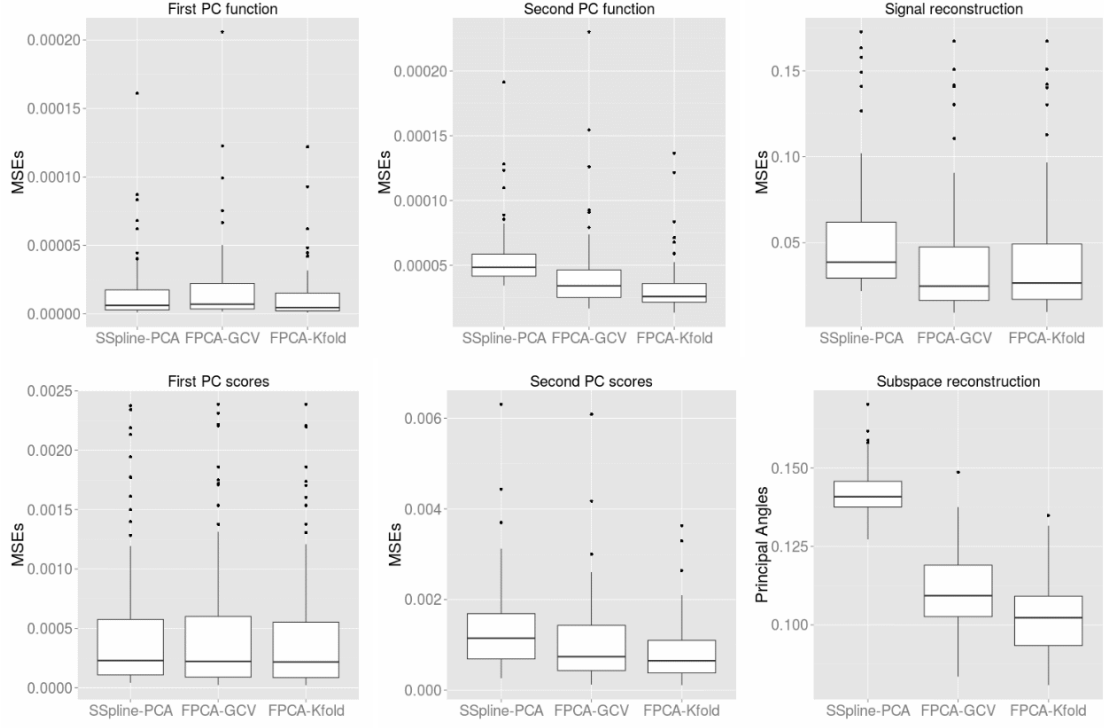


Figure 15: Boxplots summarizing the performance of SSpline-PCA and SM-FPCA. For the SM-FPCA both GCV and K -fold has been applied for the selection of the smoothing parameter.

The best estimates of the first two PC functions and corresponding scores are provided by the proposed SM-FPCA with selection of the smoothing parameter base on the K -fold approach. SSpline-PCA does a comparable job on the first principal component, but a significantly worst on the second. A possible explanation for this is the fact that SSpline-PCA tends to over-smooth the data, due to the low signal-to-noise setting of the simulations. This results in good performances for the first PC, but causes a loss of information that worsen the estimation of the second PC. The measure based on the computation of the principal angle between the original space spanned by $\{v_i\}_{i=1,2}$ and the estimated PC functions $\{\hat{v}_i\}_{i=1,2}$ emphasizes the good performance of the introduced algorithm.

References

- P. Alfeld, M. Neamtu, and L. L. Schumaker. Fitting scattered data on sphere-like surfaces using spherical splines. *Journal of Computational and Applied Mathematics*, 73(12):5 – 43, 1996.
- M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding

- and clustering. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14*, pages 585–591, 2001.
- D. Braess. *Theory, fast solvers, and applications in elasticity theory*. Cambridge University Press, 2007.
- R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter. The brain’s default network. *Annals of the New York Academy of Sciences*, 1124(1):1–38, 2008.
- M. Chung, J. Hanson, and S. Pollak. Statistical analysis on brain surfaces. Technical report, University of WisconsinMadison, 2014.
- M. K. Chung, S. M. Robbins, K. M. Dalton, R. J. Davidson, A. L. Alexander, and A. C. Evans. Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25(4):1256 – 1265, 2005.
- F. Dassi, B. Ettinger, S. Perotto, and L. M. Sangalli. A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Applied Numerical Mathematics*, 90:111 – 131, 2015.
- Deng Cai, Xiaofei He, Jiawei Han, and T. S. Huang. Graph Regularized Nonnegative Matrix Factorization for Data Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1548–1560, aug 2011.
- G. Dziuk. Finite elements for the beltrami operator on arbitrary surfaces. In S. Hildebrandt and R. Leis, editors, *Partial Differential Equations and Calculus of Variations*, volume 1357 of *Lecture Notes in Mathematics*, pages 142–155. Springer Berlin Heidelberg, 1988. ISBN 978-3-540-50508-2.
- 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. Connectivity Connectivity.
- B. Ettinger, S. Perotto, and L. Sangalli. Spatial regression models over two-dimensional manifolds. *Biometrika*, 2015. doi:10.1093/biomet/asv069.
- M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. V. Essen, and M. Jenkinson. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80(0):105 – 124, 2013.

- E. M. Gordon, T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, and S. E. Petersen. Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex*, 2014.
- P. Hall and M. Hosseini-Nasab. On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1): 109–126, 2006.
- S. J. Harrison, M. W. Woolrich, E. C. Robinson, M. F. Glasser, C. F. Beckmann, M. Jenkinson, and S. M. Smith. Large-scale probabilistic functional modes from resting state fmri. *NeuroImage*, 109(0):217 – 231, 2015.
- 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.
- I. T. Jolliffe, N. T. Trendafilov, and M. Uddin. A modified principal component technique based on the lasso. *Journal of Computational and Graphical Statistics*, 12(3):531–547, 2003.
- S. Ogawa, T. M. Lee, A. R. Kay, and D. W. Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A*, 87(24): 9868–72, 1990.
- J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2nd edition, 2005.
- T. Ramsay. Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):307–319, 2002.
- J. A. Rice and B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(1):pp. 233–243, 1991.
- 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.
- H. Shen and J. Z. Huang. Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, 99(6):1015 – 1034, 2008.
- B. W. Silverman. Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24(1):1–24, 02 1996.
- G. Wahba. Spline interpolation and smoothing on the sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16, 1981.

- L. Zhou and H. Pan. Principal component analysis of two-dimensional functional data. *Journal of Computational and Graphical Statistics*, 23(3):779–801, 2014.
- H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67:301–320, 2005.
- H. Zou, T. Hastie, and R. Tibshirani. Sparse principal component analysis. *Journal of Computational and Graphical Statistics*, 15:2006, 2004.