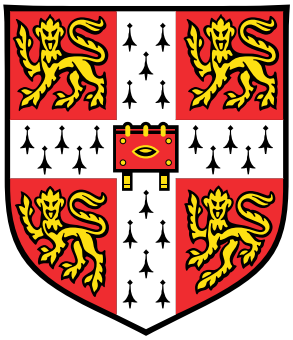
Geometric Functional Data Analysis



Eardi Lila

Emmanuel College
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

July 2019

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text.

Abstract

In this thesis, we introduce a comprehensive framework for the analysis of statistical samples that are functional data with non-trivial geometry. Geometry can interplay with functional data in different forms. The most general setting considered here is that of functional data supported on random non-linear smooth manifolds. This is a situation often encountered in neuroimaging, where modern imaging modalities are now able to produce structural brain representations coupled with functional information. Practitioners have commonly approached the analysis of such data with a two step approach. In the first step the manifolds are registered to a template and in the second step the functional information is analyzed on the template ignoring the registration step. The separation of the two steps precludes studies aimed at understanding how geometric variations relate to functional variations. On the other hand, functional data analysis has mostly developed tools for simplified settings, such as one-dimensional functional samples, limiting their applicability to real data. We formulate a model which is able to jointly represent geometric and functional variations. In this setting, modeling functional information requires the formulation of models able to incorporate structural information on the geometry of the underlying domains, with the aim of mitigating the curse of dimensionality. This is achieved by adopting regularized models involving differential operator penalties. Modeling random smooth manifolds requires the formulation of models constrained to produce 'sensible' shapes, e.g. not self-intersecting. This is achieved by means of diffeomorphic flows. The proposed models have been applied to real data to perform studies able to relate structural changes to functional changes, and specifically, to study associations between brain shape and cerebral cortex thickness. We can also deal with more complex functional samples, themselves constrained to lie in a non-linear subspace. This is for instance the case of covariance operators, describing brain connectivity, which are symmetric and positive semi-definite operators. Thanks to the proposed models, we are able to model connectivity as an 'object' and study its variations in time or across individuals. We also consider further extensions of this framework to the inverse problems setting, which is the setting where each sample is a latent object, and only indirect measurements are available.

To my family

Acknowledgements

I am most grateful to my advisor John Aston for introducing me to the fascinating research topics of my thesis and for sharing his invaluable perspective on the field, which has most definitely shaped me as a researcher. I'm also thankful to my collaborators, Laura Sangalli and Simon Arridge, for their precious inputs. I would also like to thank my thesis examiners, Steve Marron and Lexin Li, for the insightful discussions on this work and the numerous ideas for future work.

I would like to thank my colleagues and now friends Fritz Hiesmayr, without whom office life would have been a solitary experience, Megan Griffin-Pickering, Nicolay Baldin, Kweku Abraham, Mo Dick Wong, Lisa Kreusser, Erlend Riis, Matthias Löffler, Andrew Swan and Sam Thomas.

I am very thankful to all my Emmanuel College friends, with a special thank to Jason and Roberto for constantly radiating their positivity. Thanks to the Emmanuel Boat club and its members for those painful but rewarding early morning outings. I would like to thank my friends from Milan, and particularly, Giacomo, Primo and Matteo - our yearly trip tradition remained intact despite the distance.

A special thank to Delphine, for sharing with me every step of this journey and filling my days with her energy and contagious curiosity.

Finally, I would like to thank my parents and my brother Tony. Without their support and appreciation my academic experience would have not been the same.

Contents

Introduction

Functional Data Analysis (FDA) is the branch of statistics concerned with the analysis of random variables that take values in function spaces. These can represent, for instance, curves, images or shapes. The statistical analysis of samples that are functions comes with challenges unique to this setting. Functional data are in fact intrinsically infinite-dimensional. The covariance operator, the analogue infinite-dimensional object to the covariance matrix, is a compact operator, whose inverse is in general unbounded. Therefore, multivariate approaches that rely on the inverse of the covariance matrix cannot be straightforwardly extended to this setting.

Moreover, in practice, functional samples cannot be fully observed, but only discrete measurements contaminated with observational errors are available. In FDA, it is thus common to deal with the estimation of infinite-dimensional functional quantities, from a finite number of measurements. To this end, some smoothness properties are assumed on the latent functional objects and regularized models are often adopted to enforce these properties.

As opposed to the multivariate setting, where there is a fixed and known correspondence between each entry of the samples, functions can display both amplitude and phase variability. Amplitude variability refers to 'vertical' variation, while phase variability refers to 'horizontal' variation, e.g. shift or more complicated warpings of the domain. Analyzing functional data requires models that can jointly represent amplitude and phase variation, and possibly their relationship.

Despite the fact that FDA is now a mature field, and the issues aforementioned have been to a good extent covered in the literature, functional data have been classically assumed to be smooth functions on a fixed interval of the real line. This has shown to be a great limitation to the applicability of the FDA methodology to modern complex datasets arising, for instance, from medical imaging devices.

In medical imaging, and specifically neuroimaging, functional data are ubiquitous. Structural imaging modalities, such as Magnetic Resonance and Computed Tomography

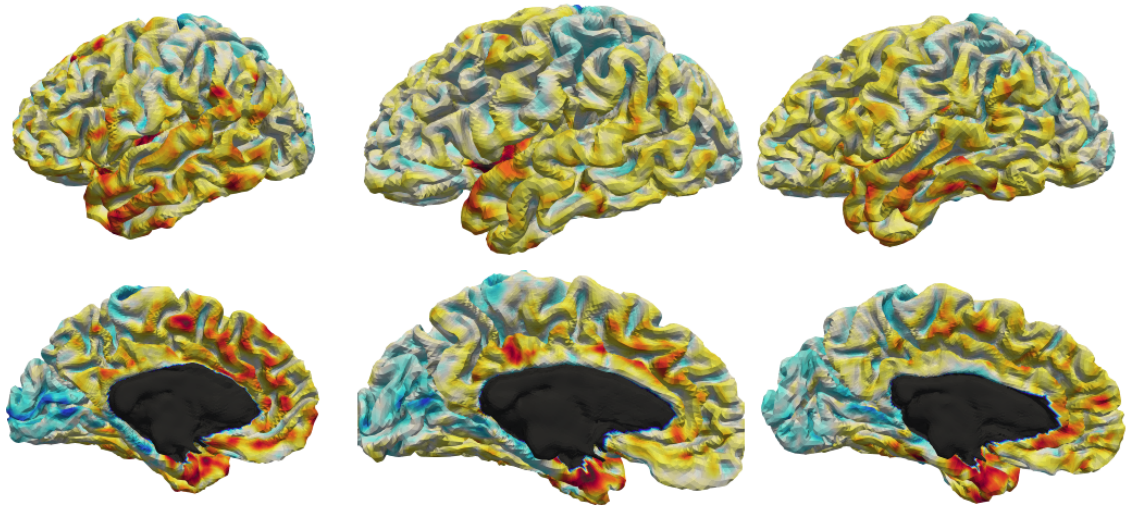


Figure 1 Surface reconstructions of the brain's left hemispheres of three different subjects, with an associated scalar signal representing the cerebral cortex thickness of the subjects. These have been reconstructed from 3D MRI scans of the subjects. The black area is a region which is not part of the brain surface.

provide images that describe the structure of the brain. Functional imaging modalities, such as functional Magnetic Resonance, Electroencephalogram and Magnetoencephalography provide time-variant images describing the activation of the different parts of the brain. These are functional data, which in their most simplified form can be formalized as a population of n samples

$$\{(\mathcal{M}_i, Y_i) \,|\, i = 1, \ldots, n\},$$

with $\{\mathcal{M}_i\}$ a set of potentially non-linear subject-specific domains, representing the structural information and $\{Y_i : \mathcal{M}_i \to \mathbb{R}\}$ a set of functions supported on $\{\mathcal{M}_i\}$ representing the functional information. In Figure 1, we show an example of three of such samples, with $\mathcal{M}_i$ a 2D surface representing the subject-specific cerebral cortex geometry and Y_i a real function representing the cerebral cortex thickness map of the same subject. The analysis of the data in Figure 1, can be contextualized within the more abstract framework of object oriented data analysis (Marron and Alonso, 2014), which is the analysis of complex non-Euclidean objects.

In this thesis, we focus on the following questions. Ignoring for a moment the geometric variability and assuming the data have been registered to an 'average' brain $\mathcal{M}$, how can we perform Functional Principal Components Analysis (FPCA) to optimally represent the functions $\{Y_i : \mathcal{M} \to \mathbb{R}\}$, while incorporating geometric information on the domain $\mathcal{M}$

where the data are supported? This would allow multivariate techniques to be applied to the functional data in a subsequent step, overcoming the non-invertibility of the covariance operator. Moreover, if Y_i are covariance functions describing the connectivity of the brain, can our FPCA methodology deal with the non-trivial geometry of these objects? Can these models deal with missing data, discrete noisy measurements, or indirect measurements? Finally, in its full generality the data in Figure 1 present both subject-specific geometric variations and functional variations. Can we jointly model geometries and functions, and perhaps infer their relationship?

Formally, these problems share many similarities with the FDA problems mentioned at the beginning of this section. For instance, the problem of modeling geometric and functional variations could be regarded as a generalization of the problem of modeling amplitude and phase variability of functional data. However, in FDA these questions have been explored mostly in the simplified setting of 1D functional data supported on a fixed interval. In this thesis, we develop a novel Geometric Functional Data Analysis framework that bridges the current gap between the classical FDA literature and the complexity of modern datasets arising from brain imaging techniques.

Thesis overview

In Chapter 1, we start by giving a brief introduction of the classical FDA setting, mainly focusing on Functional Principal Component Analysis, a crucial tool for the analysis of functional data. In Section 1.2, we follow with a concise review of the main modern differential geometry notions, which play an important role in some of the methodological aspects of this thesis. In Section 1.3, we give an overview on recent developments in the analysis of functional data that present some non-trivial geometries.

In Chapter 2, motivated by the analysis of high-dimensional neuroimaging signals located over the cortical surface, we introduce a novel FPCA 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 the literature. We

finally apply the proposed method to resting state functional magnetic resonance imaging data.

In Chapter 3, 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.

In Chapter 4, we introduce a comprehensive framework for the analysis of functional data, whose domain is a two-dimensional manifold and the domain itself is subject to variability from sample to sample. We formulate a statistical model for such data, that we call Functions on Surfaces, which enables a joint representation of the geometric and functional aspects, and propose an associated estimation framework. We assess the validity of the framework by performing a simulation study and we finally apply it to the analysis of neuroimaging data of cortical thickness, acquired from the brains of different subjects, and thus lying on domains with different geometries.

The works in Chapter 2 and Chapter 4 of this thesis, have now been published. A very preliminary version of the work in Chapter 2 has also appeared in Lila (2014). The work in Chapter 3 has been submitted for publication. The references to these papers are given below.

Lila, E., Aston, J. A. D., and Sangalli, L. M. (2016). Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging. *Annals of Applied Statistics*, 10(4), 1854–1879.

Lila, E., and Aston, J. A. D. (2019). Statistical Analysis of Functions on Surfaces, with an application to Medical Imaging. *Journal of the American Statistical Association*, in press.

Lila, E., Arridge, S., and Aston, J. A. D. (2018). Statistics on functional data and covariance operators in linear inverse problems. *arXiv preprint*, 1806.03954.

Chapter 1

Background

In this chapter, we start by giving a concise introduction to Functional Data Analysis (FDA) as treated in classical textbooks (see, e.g., Horváth and Kokoszka, 2012; Hsing and Eubank, 2015; Ramsay and Silverman, 2005), where the statistical samples are assumed to be elements of the space of square-integrable functions supported on $[0,1]$, or more in general on Ω, with $\Omega \subset \mathbb{R}^d$ a hyper-cube in $\mathbb{R}^d$. In Section 1.1, we first consider the setting of fully observable functional samples. We then present recent advances in the setting where only discrete and noisy measurements of the functional samples are given. We follow with an introduction to the setting of indirectly observed statistical samples. This is a generalization of the case of discrete and noisy measurements to more general transformations of the underlying samples rather than simple discrete evaluations.

In Section 1.2, we give an overview of some basic notions of modern differential geometry, which play an important role in extending the FDA methodology to deal with non-conventional geometric functional data, such as functions supported, or taking values, on non-linear spaces. In Section 1.3, we give an overview of the current literature on the statistical analysis of geometric functional data.

1.1 Functional Data Analysis

1.1.1 Basic Setup

Consider the space of square-integrable functions supported on Ω, i.e. $L^2(\Omega) = \{f : \Omega \to \mathbb{R} : \int_\Omega |f(v)|^2 dv < \infty\}$ with the inner product $\langle f, g \rangle = \int_\Omega f(v)g(v)dv$ and norm $\|f\|^2 = \int_{\mathcal{M}} |f(v)|^2 dv$. As already mentioned, in FDA the support Ω is typically a bounded interval of $\mathbb{R}$, e.g. $\Omega = [0,1]$.

Consider now the random variable (sometimes referred to as a random function) X with values in $L^2(\Omega)$ and define its mean $\mu \in L^2(\Omega)$ and covariance $C_X \in L^2(\Omega \times \Omega)$ to be the (continuous) functions

$$\mu(v) = \mathbb{E}[X](v), \qquad v \in \Omega, \tag{1.1}$$

$$C_X(v, v') = \mathbb{E}[(X(v) - \mu(v))(X(v') - \mu(v'))], \qquad v, v' \in \Omega. \tag{1.2}$$

The covariance function induces a covariance operator $\mathscr{C}_X : L^2(\Omega) \to L^2(\Omega)$, defined to be

$$\mathscr{C}_X g = \int_\Omega C_X(\cdot, v) g(v) \, dv, \qquad g \in L^2(\Omega). \tag{1.3}$$

It is straightforward to check that the covariance function is symmetric, i.e. $C_X(v, v') = C_X(v', v)$ for all $v, v' \in \Omega$, and positive semi-definite, i.e. $\int_{\Omega \times \Omega} C_X(v, v') f(v) f(v') \, dv \, dv' \geq 0$ for any $f \in L^2(\Omega)$. Consequently, the covariance operator is a positive semi-definite operator ($\langle \mathscr{C}_X f, g \rangle \geq 0$ for any $f \in L^2(\Omega)$) and self-adjoint ($\langle \mathscr{C}_X f, g \rangle = \langle f, \mathscr{C}_X g \rangle$ for any $f, g \in L^2(\Omega)$). Moreover, it can also be shown that the covariance operator is compact. Hence, the covariance operator admits the spectral decomposition detailed below.

Proposition 1. *The covariance operator $\mathscr{C}_X$ admits the spectral representation given by*

$$\mathscr{C}_X = \sum_{r=1}^{\infty} \kappa_r \langle \psi_r, \cdot \rangle \psi_r, \tag{1.4}$$

where $\{\kappa_r\}_{r=1}^{\infty}$ is the non-increasing set of real eigenvalues of $\mathscr{C}_X$ and $\{\psi_r\}_{r=1}^{\infty}$ the associated set of orthonormal eigenfunctions, satisfying the set of equations

$$\mathscr{C}_X \psi_r = \kappa_r \psi_r, \tag{1.5}$$

or equivalently,

$$\int_\Omega C_X(\cdot, v) \psi_r(v) \, dv = \kappa_r \psi_r. \tag{1.6}$$

The eigenvualues have zero as the only possible point of accumulation. The operator is said to be of finite rank if the set of non-zero eigenvalues is finite.

Moreover, Mercer's Lemma [Riesz and Szokefalvi-Nagy (1955)] guarantees that the sequence $\{\kappa_r\}$ of eigenvalues of $\mathscr{C}_X$ and the orthonormal sequence of corresponding eigenfunctions $\{\psi_r\}$ are such that

$$C_X(v, v') = \sum_{r=1}^{\infty} \kappa_r \psi_r(v) \psi_r(v'), \qquad v, v' \in \Omega, \tag{1.7}$$

where the convergence is absolute and uniform in $\Omega \times \Omega$.

We have now set the basis to introduce a fundamental tool in FDA: Functional Principal Component Analysis (FPCA). In fact, the spectral representation (1.4) can be used to expand X as

$$X = \mu + \sum_{r=1}^{\infty} \zeta_r \psi_r, \tag{1.8}$$

where the random variables $\zeta_1, \zeta_2, \ldots$ are uncorrelated and are given by $\zeta_r = \langle X - \mu, \psi_r \rangle$. The expansion (1.8) converges uniformly in mean square and is also known as the Karhunen-Loève (KL) expansion of X. This offers a convenient representation of the random variable X, as it disentangles the deterministic infinite-dimensional components $\{\psi_r\}$ from the finite-dimensional random variables $\{\zeta_r\}$, allowing, for instance, multivariate statistical analysis to be performed on a finite truncation of the sequence $\{\zeta_r\}$.

The collection $\{\psi_r\}$ defines the strongest modes of variation in the random function X and these are called **Principal Component (PC) functions**. In fact ψ_1 is such that

$$\psi_1 = \underset{\phi : \|\phi\| = 1}{\arg\max} \int_\Omega \int_\Omega \phi(v) C_X(v, v') \phi(v') dv dv',$$

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 $\psi_1, \ldots, \psi_{m-1}$, i.e.

$$\psi_m = \underset{\substack{\phi : \|\phi\| = 1 \\ \langle \phi, \psi_j \rangle = 0 \quad j = 1, \ldots, m-1}}{\arg\max} \int_\Omega \int_\Omega \phi(v) C_X(v, v') \phi(v') dv dv'.$$

The random variables $\zeta_1, \zeta_2, \ldots$ are called **PC scores**. This property highlights the fact that the KL representation is not only a convenient basis for the random variable X, but its basis elements have an important property that aid exploratory data analysis on the random variable X.

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

$$(\psi_m)_{m=1}^M = \underset{(\{\phi_m\}_{m=1}^M : \langle \phi_m, \phi_l \rangle = \delta_{ml})}{\arg\min} \mathbb{E} \int_\Omega \left\{ X - \mu - \sum_{m=1}^{M} \langle X, \phi_m \rangle \phi_m \right\}^2, \tag{1.9}$$

where δ_{ml} is the Kronecker delta; i.e. $\delta_{ml} = 1$ for $m = l$ and 0 otherwise. Equation (1.9) emphasizes the dimension reduction properties of FPCA and justifies popular approaches

where FPCA is first applied to the functional data and this is followed by a multivariate statistical analysis on the PC scores.

1.1.2 Fully observed functions

Suppose now we are given n samples, $X_1, \ldots, X_n$, that are fully observed independent and identical copies of the random function X taking values in $L^2(\Omega)$. Then, it is natural to estimate the mean function μ and the covariance functions C_X with the associated sample mean and sample covariance function

$$\bar{X}(v) = n^{-1} \sum_{i=1}^{n} X_i(v), \qquad \hat{C}_X(v, v') = \frac{1}{n} \sum_{i=1}^{n} (X_i(v) - \bar{X}(v))(X_i(v') - \bar{X}(v'))$$

for all $v, v' \in \Omega$. The estimates of the PC functions $\{\hat{\psi}_r\}$ and associated eigenvalues $\{\hat{\kappa}_r\}$ are computed through the characterization

$$\int_{\Omega} \hat{C}_X(v, v') \hat{\psi}_r(v) \, dv = \hat{\kappa}_r \hat{\psi}_r(v'),$$

which is in practice solved by discretizing the problem on a fine grid or by basis expansion. The score vectors in the KL expansion (1.8) can then be trivially estimated by integration, i.e.

$$\hat{\zeta}_r = \int_{\Omega} \big(X_i(v) - \hat{\mu}(v)\big) \hat{\psi}_r(v) \, dv.$$

Assuming $\mathbb{E}\|X\|^4 < \infty$ and relying on the Central Limit Theorem in Hilbert spaces, consistency and $\sqrt{n}$ rates of convergence have been derived for the empirical mean and covariance (Bosq, 2000; Dauxois et al., 1982; Hall et al., 2006). If moreover $\kappa_1 > \kappa_2 > \ldots$, uniform convergence of the empirical covariance implies consistency of the empirical eigenvalues and eigenfunctions (the latter, up to a sign factor) (Bosq, 2000).

1.1.3 Discrete measurements

In the previous section, the functional samples were assumed to be fully observable. Although this could be a reasonable assumption for very densely sampled functions, it is usually the case that, for the ith functional sample, only noisy evaluations on a discrete grid of points $v_{i1}, \ldots, v_{ip_i}$ are given, i.e.

$$Y_{ij} = X_i(v_{ij}) + \varepsilon_{ij}, \qquad j = 1, \ldots, p_i; \quad i = 1, \ldots, n, \tag{1.10}$$

where the term ε_{ij} represents independent observational errors, with mean zero and variance σ^2. Sometimes this setting is referred to as the longitudinal setting.

In the longitudinal setting, the aim is recovering the underlying PC functions and PC scores vectors of the random variable X, from the noisy measurements Y_{ij}. This can be dealt with via three different approaches: the 'Smooth-then-Estimate' approach, the 'Estimate-then-Smooth' approach, as named in Descary (2017), or the 'Regularized-Estimate' approach. The idea of overtacking the classical 'Smooth-then-Estimate' approach dates back to Rice and Silverman (1991). A conceptually similar situation is also encountered, for instance, when estimating smooth and monotone functions (Mammen, 1991; Mammen and Thomas-Agnan, 1999).

Smooth-then-Estimate

The 'Smooth-then-Estimate' approach is based on the two following steps. In the first step, the observations associated to each function are smoothed individually, in order to obtain smooth representations of $X_1, \dots, X_n$. In the second step FPCA is carried out, as described in the previous section.

Smoothing is a very well explored topic in statistics, both from the non-parametric regression prospective (Fan and Gijbels, 1996; Green and Silverman, 1994) and the spatial data analysis prospective (Cressie, 1993). Popular smoothing approaches in FDA are the least-square finite basis approximation and smoothing spline approaches, which we briefly recall here. The smoothing step is executed separately for each functional sample. Hence, we drop here the index i denoting the specific sample and focus on the problem

$$Y_j = X(v_j) + \varepsilon_j, \qquad j = 1, \dots, p.$$

The least-square finite basis approach consists of introducing a smooth K-dimensional basis $\{b_k\}_{k=1}^K$ on the domain Ω, and defining a smooth estimate of X, that is $\hat{X} = \sum_{k=1}^K \hat{c}_k b_k$ with

$$\{\hat{c}_k\} = \underset{\{c_k\}}{\arg\min} \sum_j \left(Y_j - \sum_{k=1}^K c_k b_k(v_j) \right)^2.$$

Typical choices of the basis for one-dimensional domains of the type $\Omega = [0,1]$ are truncated Fourier series and B-splines basis (Ramsay and Silverman, 2005). The amount of smoothness on $\hat{X}$ can be controlled by appropriate choice of the basis function and the truncation level K.

An alternative approach to smoothing consists of formulating the problem in a completely non-parametric fashion, i.e. find a function f such that

$$\hat{X} = \arg\min_{f} \sum_{j} \Big(Y_j - f(v_j) \Big)^2 + \lambda \int_{\Omega} (Df)^2, \tag{1.11}$$

with D an operator measuring the roughness of the function f, which squared and integrated over the domain offers a global roughness penalization term. The constant λ weighs the classical term measuring the least square misfit to the data and the penalization term, and can be chosen either with data driven approaches, such as cross-validation, or can be seen as an additional parameter allowing analysis on different 'scales' (Green and Silverman, 1994). A typical choice for D, for one-dimensional domains, is the second derivative. In this case, the formulation (1.11) is also known as the smoothing spline problem.

The smoothing spline approach is of particular interest to the methodology developed in this thesis. In fact, by developing the necessary geometric tools we will be able to extend formulations of the type (1.11) to more general, and possibly non-linear, domains.

Estimate-then-Smooth

In the case of highly sparse sampling schemes, individual recovery of the functions cannot be pursued as there is no sufficient information to treat them independently. An alternative approach consists of exploiting the unprocessed data to construct point-wise estimates and then proceed with smoothing the point-wise estimates to recover the underlying functional estimates.

For simplicity, suppose for the moment that the functional samples are evaluated on a fixed grid of points $v_1, \dots, v_p \in \Omega$ across samples

$$Y_{ij} = X_i(v_j) + \varepsilon_{ij}, \qquad j = 1, \dots, p; \quad i = 1, \dots, n,$$

with ε_{ij} independent observational error with mean zero and variance σ^2.

From the observations $\{Y_{ij}\}$, at the locations $v_1, \dots, v_p \in \Omega$, we can construct the point-wise estimators

$$\bar{Y}(v_j) = \frac{1}{n} \sum_{i=1}^{n} y_{ij}, \qquad \hat{C}_Y(v_j, v_l) = \frac{1}{n} \sum_{i=1}^{n} \big(Y_{ij} - \bar{Y}(v_j) \big) \big(Y_{il} - \bar{Y}(v_l) \big).$$

The estimator $\bar{Y}(v_j)$ returns an estimate of $\mu(v_j)$, the mean of X at the point v_j. Instead, $\hat{C}_Y(v_j, v_l)$ represents an estimate of $C_X(v_j, v_l)$, the covariance function of X at (v_j, v_l), only for $j \neq l$, as $\text{Cov}(Y_i(v_j), Y_i(v_l)) = C_X(v_j, v_l) + \sigma^2$ for $j = l$ and $\text{Cov}(Y_i(v_j), Y_i(v_l)) = C_X(v_j, v_l)$ for $j \neq l$. This is due to the presence of observational error. A standard strategy (Yao et al., 2005) is to remove the diagonal from $\hat{C}_Y$. Finally, the mean μ and covariance function C_X can be reconstructed by smoothing, respectively, $\bar{Y}(v_j)$ and $\hat{C}_Y$ (the latter post-diagonal removal).

A similar approach can be followed when the sampling scheme involves sample-specific sparse measurements $v_{i1}, \ldots, v_{ip_i}$ for the ith sample. In this case, an estimate of the mean function μ can be obtained by smoothing the measurements $\{(Y_i(v_{ij}), v_{ij}) : i = 1, \ldots, n; j = 1, \ldots, p_i\}$. Analogously, the covariance can be estimated by smoothing the 'raw covariance'

$$\left\{\Big(\big(Y_i(v_{ij}) - \hat{Y}(v_{ij})\big)\big(Y_i(v_{il}) - \hat{Y}(v_{il})\big), v_{ij}, v_{il}\Big) : i = 1, \ldots, n; j, l = 1, \ldots, p_i, j \neq l\right\},$$

where the condition $j \neq l$ ensures that the estimates on the diagonal are discarded.

In the more general version of these estimators, a weight w_i could be attached to each term $Y_i(v_{ij})$ and a weight z_i could attached to each term $\big(Y_i(v_{ij}) - \hat{Y}(v_{ij})\big)\big(Y_i(v_{il}) - \hat{Y}(v_{il})\big)$. These can be chosen to put either equal weights on each observation (see, e.g., Yao et al., 2005) or equal weights on each sample (see, e.g., Li and Hsing, 2010). A unified theory is provided in Zhang and Wang (2016).

The PC functions estimates $\{\hat{\psi}_r\}$ and associated eigenvalues estimates $\{\hat{\kappa}_r\}$ can be computed from the eigenanalysis of the estimated covariance function $\hat{C}_X$. As for the scores predictions, in the sparse setting, the integration formula $\int_\Omega \big(X_i(v) - \hat{\mu}(v)\big)\hat{\psi}_r(v)dv$ might be ineffective, due to the low number of measurements per each curve. Therefore, Yao et al. (2005) propose to reformulate the problem of estimating the rth PC scores ζ_{ir} with its best linear predictor, given the observations $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip_i})^T$. Let $\bar{\mathbf{Y}}_i = (\bar{Y}(v_{i1}), \ldots, \bar{Y}(v_{ip_i}))^T$ then the best linear predictor is

$$\hat{\zeta}_{ir} = \hat{\kappa}_r \hat{\boldsymbol{\psi}}_{ir}^T \hat{C}_Y^{-1}(\mathbf{Y}_i - \bar{\mathbf{Y}}_i),$$

with $\hat{\boldsymbol{\psi}}_{ir} = (\hat{\psi}_r(v_{i1}), \ldots, \hat{\psi}_r(v_{ip_i}))^T$ and $[\hat{C}_Y(v_{ij}, v_{il})]$ a $p_i \times p_i$ matrix such that $\hat{C}_Y(v_{ij}, v_{il}) = \hat{C}_X(v_{ij}, v_{il}) + \delta_{j,l}\hat{\sigma}^2$, where $\hat{\sigma}^2$ is an estimator of the noise variance and $\delta_{j,l} = 1$ if $j = l$ and zero otherwise.

A further extension to situations where the noise term is rough, but has a non-trivial (i.e. non-diagonal) covariance structure, has been considered in Descary and Panaretos (2019).

Regularized-Estimation

Here we present a third approach, which is the most relevant to the methodology presented in this thesis: the Regularized-Estimation approach. This consists in adding a penalization term to the classic formulation of PCA, in order to recover a desired feature of the estimated underlying functions.

In the setting of fully observable samples, Regularized-Estimation has been proposed in Rice and Silverman (1991) and Silverman (1996). These formulations are based on the variational formulation of FPCA formulated as the maximization of the quadratic form

$$\frac{\int_\Omega \int_\Omega \phi(v)\hat{C}_X(v,v')\phi(v')dvdv'}{\int_\Omega \phi^2(v)dv}. \tag{1.12}$$

In particular, for functions supported on $\Omega = [0,1]$, Rice and Silverman (1991) propose to modify the formulation above with the maximization of

$$\frac{\int_\Omega \int_\Omega \phi(v)\hat{C}_X(v,v')\phi(v')dvdv' - \lambda \int_\Omega (\phi''(v))^2 dv}{\int_\Omega \phi^2(v)dv}. \tag{1.13}$$

Instead, Silverman (1996) proposes to maximize the alternative formulation

$$\frac{\int_\Omega \int_\Omega \phi(v)\hat{C}_X(v,v')\phi(v')dvdv'}{\int_\Omega \phi^2(v)dv + \lambda \int_\Omega (\phi''(v))^2 dv}. \tag{1.14}$$

Both formulations introduce a smoothing penalty term which yields smoother PC functions.

Consider now the discrete and noisy measurements setting and assume that the functional samples are evaluated on a fixed grid of points $v_1,\dots,v_p \subset \Omega$, as in model (1.10). Define the $n \times p$ matrix $\mathbb{Y} = (Y_{ij})$, the column vector $\bar{\mathbf{Y}} = (\frac{1}{n}\sum_{i=1}^n Y_{ij})$ of length p, the n-dimensional vector $\mathbf{z}$ and the p-dimensional vector $\boldsymbol{\phi} = (\phi_1(v_j))$. Let $\mathbf{1}$ denote the column vector of length n with all entries equal to 1. The empirical counterpart of the objective function in (1.9), for the first PC function, becomes

$$\frac{1}{n}\|\mathbb{Y} - \mathbf{1}\bar{\mathbf{Y}}^T - \mathbf{z}\boldsymbol{\phi}^T\|_F^2, \tag{1.15}$$

where $\|\cdot\|_F$ is the Frobenius matrix norm, defined as the square root of the sum of squares of its entries. This last formulation gives a natural way to deal with the fact that only discrete and noisy evaluations Y_{ij}, $i = 1,\ldots,n$, $j = 1,\ldots,p$ of the underlying functional samples are available. However, it does not incorporate any information on the smoothness of the functional data. In fact, considering the Singular Value Decomposition (SVD) of $\mathbb{Y} - \mathbf{1}\mu^T = UDV^T$, it can be shown that the minimizing arguments of (1.15) are $\hat{\boldsymbol{\phi}} = [V]_{\cdot,1}$ and $\hat{\mathbf{z}} = [UD]_{\cdot,1}$, respectively the first column of V and UD, which is a multivariate PCA applied to the data-matrix $\mathbb{Y}$.

Formulation (1.15) has shown great flexibility for the purpose of incorporating regularizing terms in the estimation of the PC functions, achieving joint estimation and regularization. In the literature of sparse PCA, where sparse in this case refers to a high number of zero entries, sparsifying regularization terms are introduced in the objective function (1.15) (Jolliffe et al., 2003; Shen and Huang, 2008; Zou and Hastie, 2005). In FDA, smoothing penalization terms are usually adopted. These encourage the underlying PC function ϕ_1 in (1.15) to be smooth. An example of a PCA model, for functions supported on the one-dimensional domain $[0,1] \subset \mathbb{R}$, that explicitly incorporates a smoothing penalization term in (1.15), can be found in Huang et al. (2008). The authors propose to minimize the objective function

$$\frac{1}{n}\|\mathbb{Y} - \mathbf{1}\bar{\mathbf{Y}}^T - \mathbf{z}\boldsymbol{\phi}^T\|_F^2 + \lambda \mathbf{z}^T\mathbf{z}\int_{[0,1]}(\phi_1''(v))^2\,dv, \tag{1.16}$$

with respect to $(\mathbf{z}, \phi_1)$. The subsequent PC functions are estimated by removing the first PC function effect from the data matrix $\mathbb{Y}$ and then re-applying the model.

As already mentioned, most of the present literature focuses on functions supported on domains that are intervals of $\mathbb{R}$. Zhou and Pan (2014) propose a smooth FPCA model for two-dimensional functions on irregular planar domains, by adapting the formulation (1.16). Amini and Wainwright (2012) propose a regularized M-estimator in a Reproducing Kernel Hilbert Space (RKHS) framework. In Chapter 2 we propose a generalization of the model in Huang et al. (2008) to the case of real functions whose domain is a non-linear two-dimensional manifold, leading to a model that can fully exploit the information on the geometry of the manifold domain.

1.1.4 Indirect measurements

In the previous subsections, we explored different extensions of classical FPCA to situations where the functional data are not fully observed, but only noisy discrete measure-

ments are available. In this section, we proceed with a further generalization of this setting to the most general case of indirectly observed functions. For instance, in a number of medical imaging modalities, the samples from our random variable of interest X are not directly observable, but only indirect measurements $\mathbf{Y}_i$ are given. These are assumed to be generated from the model

$$\mathbf{Y}_i = \mathscr{K} X_i + \boldsymbol{\varepsilon}_i,$$

where $\mathbf{Y}_i$ and $\boldsymbol{\varepsilon}_i$ take values in an either finite or infinite dimensional space. The map $\mathscr{K}$ is deterministic and assumed to be known and is usually referred to as the forward operator. The term $\boldsymbol{\varepsilon}_i$ represents observational error. This setting is a generalization of the sparse measurements setting, as taking $\mathscr{K}$ to be a vector-valued functional that evaluates a function at p pre-specified points $\{v_j\} \subset \Omega$, we obtain the generative model in (1.10).

In practice, we will assume that $\mathbf{Y}_i$ and $\boldsymbol{\varepsilon}_i$ take values in the space of real s-dimensional vectors and that the forward operator is of the type $\mathscr{K} : L^2(\Omega) \to \mathbb{R}^s$. Here, we consider the problem of estimating the PC functions $\{\psi_r\}$ of X, or equivalently the eigenfunctions of $\mathscr{C}_X$ (which is the covariance operator of X), from the observations $\{\mathbf{Y}_i\}$. In neuroimaging studies this is often an important task as if X is for instance a functional signal describing brain activity, then $\mathscr{C}_X$ gives a description of the first-order connectivity of the brain.

In analogy with the discrete measurements scenario, there are three possible approaches to PCA in this setting: the 'Reconstruct-then-Estimate' approach, the 'Estimate-then-Reconstruct' approach, and the 'Regularized-Estimation' approach. The most popular is the 'Reconstruct-then-Estimate' approach, where the problem of estimating the PC functions $\{\psi_r\}$ is tackled in two steps. In the first step, estimates $\{\hat{X}_i\}$ of the functions $\{X_i\}$ are individually computed from the vectors $\{\mathbf{Y}_i\}$. In the second step, the covariance function C_X is estimated from $\{\hat{X}_i\}$ by use of classical estimators for fully observed functions, i.e. spectral analysis of the estimated covariance operator. Given that in this approach the reconstruction step is performed independently for each sample, we focus on the problem

$$\mathbf{Y} = \mathscr{K} X + \varepsilon.$$

This 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 estimation of X 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 (see, among others, Cavalier, 2008; Hu and Jacob, 2012;

Lefkimmiatis et al., 2012; Mathé and Pereverzev, 2006; Tenorio, 2001), i.e.

$$\hat{X} = \underset{f \in \mathscr{F}}{\arg\inf} \|\mathbf{Y} - \mathscr{K} f\|^2 + \lambda \mathscr{P}(f), \tag{1.17}$$

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

The functional $\mathscr{P}$ encodes prior information on the function to be estimated, while the data fidelity term ensures that the resulting estimated function $\hat{X}$ is such that $\mathscr{K} f$ is a good approximation of the signal $\mathbf{Y}$ actually detected. The parameter λ 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 $\mathscr{P}$ are Sobolev (semi-) norms, which encode smoothness, or the total variation norm, which allows discontinuity but penalizes for excessively oscillating functions.

Alternatively, when it is important to assess the uncertainty associated with the estimates, a Bayesian approach could be adopted (Repetti et al., 2019; Stuart, 2010). More recently, also the deep convolutional neural network approach has been applied to this setting (Jin et al., 2017; McCann et al., 2017). Also, more complex penalty terms could be considered, for instance, by adding terms that encourage the reconstruction to be sparse.

However, a Reconstruct-then-Estimate approach for the PC functions $\{\psi_r\}$ can be suboptimal. 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. Alternatively an 'Estimate-then-Reconstruct' approach could be adopted, where first the PC components of $\mathbf{Y}$ are estimated, and subsequently the associated PC functions $\{\psi_r\}$ are reconstructed with a formulation of the type (1.17). In Katsevich et al. (2015), driven by an application to cryo-electron microscopy, the authors propose an unregularized estimator for the covariance matrix of indirectly observed functions. This appears to work well in their setting, however, to tackle more general ill-posed inverse problems, a regularized approach is needed. In Chapter 3, we develop a Regularized-Estimation approach to directly estimate the PC functions $\{\psi_r\}$ from the the data $\{\mathbf{Y}_i\}$. A more detailed review of current works in this setting is provided in Section 3.2.1.

1.2 Elements of Differential Geometry

In this section, we give a formal introduction to the main geometric concepts necessary to generalize the methodology presented in the previous section to geometric functional

data. Geometric functional data will be either functional samples supported on non-linear domains, e.g. where Ω is a two-dimensional surface in $\mathbb{R}^3$, or functional data that lie on a non-linear space, e.g. where the samples from X belong to a non-linear subspace of $L^2(\Omega)$.

We focus here on a concise exposure, while trying to develop an intuitive understanding of the objects introduced. A more exhaustive treatment on the subject can be found in classical textbooks (see, e.g. Chavel, 2006; Lee, 1997, 2012).

1.2.1 Differentiable Manifold

A differentiable manifold is a mathematical object modeling generalizations of geometric entities such as curves and surfaces. In practice, the notion of a manifold is used in this thesis to model the geometry of a variety of objects, such as, brain surfaces or more abstract entities such as the space of covariance matrices. The choice to model the space of covariance matrices as a manifold has its limitations, in fact, to handle covariances with different ranks the concept of manifold stratification has been adopted in the literature (see, e.g. Hotz et al., 2013, and references therein).

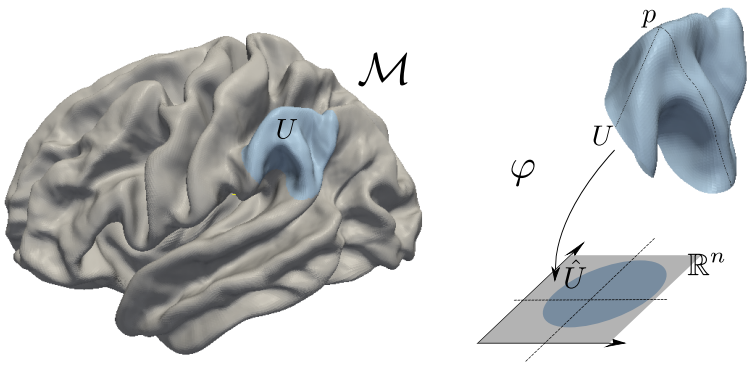


Figure 1.1 A two-dimensional manifold $\mathcal{M}$ embedded in $\mathbb{R}^3$, representing a template of the brain surface, with a coordinate chart (U, φ) offering a local parametrization of the portion $U \subset \mathcal{M}$.

Intuitively, a n-manifold (or a n dimensional manifold) is a set $\mathcal{M}$ with a reference system attached to it, i.e. one-to-one mappings from subsets $U \subset \mathcal{M}$ to $\mathbb{R}^n$. Formally, a n-dimensional manifold is a second-countable Hausdorff space that is locally Euclidean of dimension n (Lee, 2012). This means that for each $p \in \mathcal{M}$ there exists a coordinate chart

on $\mathscr{M}$ that is a pair (U,φ), with U an open subset of $\mathscr{M}$ containing p (the domain of the chart), and $\varphi : U \to \hat{U}$ a homeomorphism (i.e. a one-to-one continuous map) from U to the open subset $\hat{U} = \varphi(U) \subseteq \mathbb{R}^n$. Intuitively, the charts are local parametrizations of the manifold and offer a way to 'navigate' the manifold. In Figure 1.1, we show an example of a chart (U,φ) parametrizing a portion of the manifold $\mathscr{M}$ representing the brain geometry.

We are able to define a real function $f : \mathscr{M} \to \mathbb{R}$ on the manifold $\mathscr{M}$, by associating a real number to each element $p \in \mathscr{M}$, but how do we generalize the concept of a smooth function supported on $\mathscr{M}$? To this purpose, we need to introduce additional structure on $\mathscr{M}$. Let (U,φ) and (V,ψ) be two charts of $\mathscr{M}$ such that there is a point $p \in \mathscr{M}$ that belongs to both U and V. It is natural to define the coordinate transformation from φ to ψ as the function

$$\psi \circ \varphi^{-1} : \hat{U} \to \hat{V}$$

We define an atlas $\mathscr{A}$ for $\mathscr{M}$ to be a collection of charts whose domains cover $\mathscr{M}$ and we say that this is smooth (or that $\mathscr{M}$ is smooth) if any two charts (U,φ) and (V,ψ) in $\mathscr{A}$, such that $U \cap V \neq \emptyset$, have coordinate transformations $\psi \circ \varphi^{-1}$ that are diffeomorphisms, i.e. smooth (C^∞) one to one functions with smooth inverse. Note that $\psi \circ \varphi^{-1}$ is a function from an open subset of $\mathbb{R}^n$ to $\mathbb{R}^n$, hence we have a notion of smoothness for such functions. We define a smooth manifold to be a pair $(\mathscr{M},\mathscr{A})$, where $\mathscr{M}$ is a manifold and $\mathscr{A}$ is a maximal smooth atlas, i.e. a smooth atlas not properly contained in any larger smooth atlas. On a technical note, for smooth manifolds, although an atlas must be included in the definition, multiple different atlases can give "the same" smooth manifold, however two manifolds are equal if and only if they are equal as topological spaces. Requiring the atlas to be maximal, makes the choice of $\mathscr{A}$ unique, in the sense that if $(\mathscr{M},\mathscr{A})$ and $(\mathscr{M},\mathscr{A}')$ are smooth manifolds such that the identity map from $\mathscr{M}$ to $\mathscr{M}$ is a diffeomorphism between them, then $\mathscr{A} = \mathscr{A}'$.

We are now able to define what a smooth function on a smooth manifold is. In fact, suppose $\mathscr{M}$ is a smooth n-manifold, we say that $f : \mathscr{M} \to \mathbb{R}^k$ is a smooth function if for every $p \in \mathscr{M}$ there exists a smooth chart (U,φ) for $\mathscr{M}$ whose domain contains p and such that $\bar{f} = f \circ \varphi^{-1} : \mathbb{R}^n \to \mathbb{R}^k$ is smooth. Analogously, given two smooth manifolds $\mathscr{M}$ and $\mathscr{N}$, a function $F : \mathscr{M} \to \mathscr{N}$ is said to be smooth if for every $p \in \mathscr{M}$ there exist any two charts (U,φ) and (V,ψ), with $p \in U$ and $F(U) \subseteq V$, such that $\psi \circ F \circ \varphi^{-1} : \varphi(U) \to \psi(V)$ is smooth. Given that for $\mathscr{M}$ a smooth manifold, the coordinate transformations are diffeomorphic, the different charts 'agree' on the smoothness of a function f.

1.2.2 Tangent vectors and Tangent spaces

A tangent space $T_p\mathscr{M}$ at a point $p \in \mathscr{M}$ is intuitively the vector space obtained by linearizing $\mathscr{M}$ around p. Consider a parametric curve $\gamma : I \to \mathscr{M}$, with $I \subset \mathbb{R}$, such that $\gamma(a) = p$. In the simpler case where $\mathscr{M}$ is embedded in $\mathbb{R}^l$, i.e. $\mathscr{M} \subset \mathbb{R}^l$, the range of γ also belongs to $\mathbb{R}^l$, and within this linear space we have a meaningful concept of derivative which is

$$\dot{\gamma}(a) = \lim_{h \to 0} \frac{\big(\gamma(a+h) - \gamma(a)\big)}{h}. \tag{1.18}$$

The set of possible 'velocity' vectors $\dot{\gamma}(a)$ defines the tangent space on p. The tangent space on $p \in \mathscr{M}$, for a two-dimensional manifold, is shown in Figure 1.2. However, in general, a manifold does not come with a concept of 'embedding' space $\mathbb{R}^l$ and, in this case, the equation above is not meaningful. This will lead to a more abstract definition of tangent spaces.

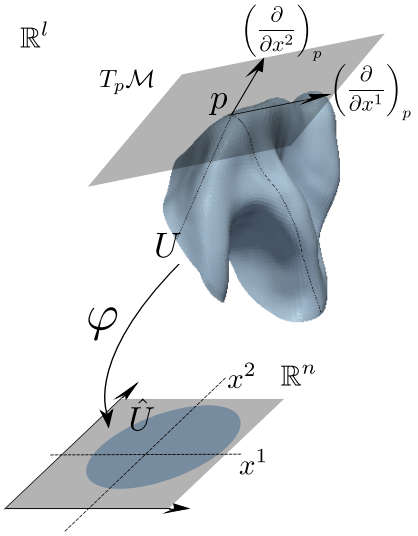


Figure 1.2 Tangent space centered at $p \in \mathscr{M}$.

The main idea consists in relying on a different property of the standard derivative, which is easier generalize to a generic manifold, and use this to define derivatives. Let (U, φ) be a smooth chart containing $p \in \mathscr{M}$. Let x^i be the i-th coordinate of the $\mathbb{R}^n$-valued function $\varphi = (x^1, \ldots, x^n)^T$. Let $f : U \to \mathbb{R}$ be a smooth function. Let $\bar{f} = f \circ \varphi^{-1}$ and define $\bar{\gamma}(t) = (\gamma^1(t), \ldots, \gamma^n(t))$, with $\gamma^i(t) = x^i(\gamma)$. Using the fact that $f(\gamma(t)) = \bar{f}(\bar{\gamma}(t))$, we may

define the directional derivative of f along the curve γ to be

$$\begin{aligned}\frac{d}{dt}f\left(\gamma(t)\right) &= \frac{d}{dt}\bar{f}\left(\bar{\gamma}(t)\right)\\ &= \sum_{i=1}^{n}\left(\frac{\partial \bar{f}}{\partial x^i}\right)_{\bar{\gamma}(t)}\frac{d\gamma^i(t)}{dt}\\ &= \sum_{i=1}^{n}\left(\frac{\partial f}{\partial x^i}\right)_{\gamma(t)}\frac{d\gamma^i(t)}{dt},\end{aligned} \tag{1.19}$$

where we have exploited the representation in local coordinates $\{x^i\}$.

Remark 1. *From now on, in this section, we will rely on the Einstein's convention and omit the summation $\sum$ over repeated indices. With this convention, equation (1.19) becomes*

$$\frac{d}{dt}f\left(\gamma(t)\right) = \left(\frac{\partial f}{\partial x^i}\right)_{\gamma(t)}\frac{d\gamma^i(t)}{dt}.$$

Let us consider this directional derivative as an expression of the function f. In other words, we define the **tangent vector** $\dot{\gamma}(a)$ to be the *operator* that, for a given curve γ, maps smooth functions to the real value $\frac{d}{dt}f(\gamma(t))|_{t=a}$. Then we may rewrite (1.19) as

$$\dot{\gamma}(a) = \left(\frac{d\gamma(t)}{dt}\right)_p := \dot{\gamma}^i(a)\left(\frac{\partial}{\partial x^i}\right)_p, \tag{1.20}$$

where we use the notation $\dot{\gamma}^i(a) = \frac{d\gamma^i(t)}{dt}$. In other words, $\dot{\gamma}(a)$ is a linear combination of operators $\left\{\left(\frac{\partial}{\partial x^i}\right)_p\right\}$, with coefficients $\{\dot{\gamma}^i(a)\}$.

Despite the abstract definition of a tangent vector, defined to be an operator acting on smooth functions, it is important to intuitively think of these objects as 'geometric tangent vectors'. In fact, it can be shown that when the tangent vectors can be defined using (1.18) there is a natural one-to-one correspondence between Equations (1.18) and (1.20).

Consider now all curves passing through the point $p \in \mathscr{M}$. We denote the set of all tangent vectors corresponding to these curves by $T_p\mathscr{M}$:

$$T_p\mathscr{M} = \left\{ c^i\left(\frac{\partial}{\partial x^i}\right)_p : (c^1,\cdots,c^n)^T \in \mathbb{R}^n \right\}.$$

This forms a linear space of dimension n. We call $\left(\frac{\partial}{\partial x^i}\right)_p$ the natural basis with respect to the coordinate system φ. For an element $D \in T_p\mathscr{M}$ and for all smooth real functions f, g and all $a, b \in \mathbb{R}$, D is linear ($D(af+bg) = aD(f) + bD(g)$) and satisfies the Leibniz's rule

$(D(f \cdot g) = f(p)D(g) + g(p)D(f))$, known to hold for the standard derivative. In fact, it is possible to define tangent vectors in terms of these properties. See Lee (2012) for a deeper discussion on possible approaches to define the tangent space.

Now we would like to formally introduce the concept of a vector field on a manifold $\mathcal{M}$, a formalization of the one shown in Figure 1.3. To this purpose, denote with $T\mathcal{M} = \sqcup_{p\in\mathcal{M}} T_p\mathcal{M}$ the tangent bundle, i.e. the disjoint union of tangent spaces. A vector field is a continuous map $X : \mathcal{M} \to T\mathcal{M}$ $(p \mapsto X_p)$, with the property that $X_p \in T_p\mathcal{M}$ for each $p \in \mathcal{M}$. Here, X_p denotes the value of X at p. If $X : \mathcal{M} \to TM$ is a (in general rough) vector field and $(U; \{x^i\})$ is any smooth coordinate chart for $\mathcal{M}$; we can write the value of X at any point $p \in U$ in terms of the coordinate basis vectors:

$$X_p = X^i(p) \left.\frac{\partial}{\partial x^i}\right|_p .$$

This defines n functions $X^i : U \to \mathbb{R}$, called the component functions of X in the given chart. A rough vector field on $\mathcal{M}$ is a (not necessarily continuous) map $X : \mathcal{M} \to T\mathcal{M}$. We are primarily interested in smooth vector fields, the ones that are smooth as maps from $\mathcal{M}$ to $T\mathcal{M}$ (when $T\mathcal{M}$ is given the smooth manifold structure, (see Lee, 2012, for details)). It is standard to denote with $\mathfrak{X}(\mathcal{M})$ (or alternatively with $\Gamma(T\mathcal{M})$) the set of all smooth vector fields on $\mathcal{M}$ and define $fX : \mathcal{M} \to T\mathcal{M}$ as $(fX)_p = f(p)X_p$.

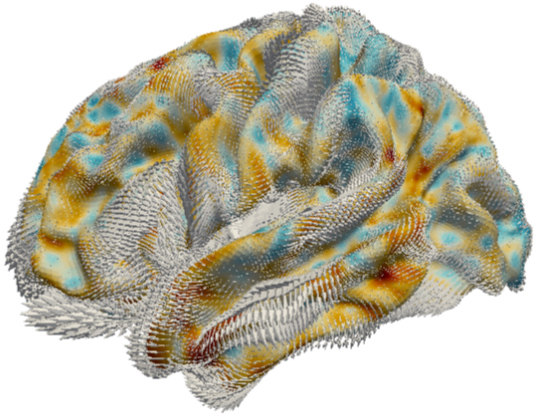


Figure 1.3 A smooth vector field $X \in \mathfrak{X}(\mathcal{M})$ on the manifold $\mathcal{M}$.

For any $X, Y \in \mathfrak{X}(\mathcal{M})$ and any $c \in \mathbb{R}$, the mappings $X+Y : p \mapsto X_p+Y_p$ and $cX : p \mapsto cX_p$ are also members of $\mathfrak{X}(\mathcal{M})$. Hence $\mathfrak{X}(\mathcal{M})$ is a linear space. In addition, for any smooth real function f, the mapping $fX : p \mapsto f(p)X_p$ is a member of $\mathfrak{X}(\mathcal{M})$.

1.2.3 Riemannian metrics

In this section, we enrich smooth manifolds with further structure, with the aim of defining geometric concepts analogous to lengths and angles on linear vector spaces. For manifolds, the appropriate structure is called the Riemannian metric.

Given a smooth manifold $\mathscr{M}$, for each point $p \in \mathscr{M}$, we define an inner product $g_p(D, D')$ for $D, D' \in T_p\mathscr{M}$, to be a real valued map $g_p(D, D')$ that is bi-linear, symmetric, positive-definite and smooth in p. Given a Riemannian metric g on $\mathscr{M}$, we call $(\mathscr{M}, g)$ a Riemannian manifold. In any smooth local coordinates $\left\{\frac{\partial}{\partial x^i}\right\}$, alternatively denoted as $\left\{(\partial_i)_p := \frac{\partial}{\partial x^i}\right\}$, the Riemannian metric g can be written in terms of the $n \times n$ matrix of components $\{g_{ij}; i, j = 1, \cdot, n\}$, with g_{ij} determined by $g_{ij}(p) = g_p(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j})$. If we rewrite the tangent vectors $D, D' \in T_p\mathscr{M}$ in terms of their coordinates as $D = D^i(\partial_i)_p$ and $D' = D'^i(\partial_i)_p$ their inner product can be written as:

$$g_p(D, D') = g_{ij}(p) D^i D'^j,$$

and the length $\|D\|$ of the tangent vector D is given by

$$\|D\|^2 = g_p(D, D) = g_{ij}(p) D^i D^j.$$

We define $g^{ij}(p)$ to be the (i, j)th component of the inverse of the matrix $(g_{ij}(p))_{ij}$ (this also symmetric positive definite), i.e.

$$g_{ij} g^{jk} = \delta_{ik}$$

(in Einstein notation) and with $\delta_{ik} = 1$ if $k = i$ and 0 otherwise.

Let now $\gamma : I \to \mathscr{M}$ be a parametric curve on the Riemannian manifold $\mathscr{M}$. We can finally define its length $\|\gamma\|$ to be

$$\|\gamma\| = \int_I \left\| \frac{d\gamma}{dt} \right\| dt = \int_I \sqrt{g_{ij} \dot{\gamma}^i \dot{\gamma}^j}\, dt,$$

where $\dot{\gamma}^i(t) \in T_{\gamma(t)}\mathscr{M}$ is the derivative of $\gamma^i = x^i \circ \gamma$.

If $\mathscr{M}$ is a sub-manifold of the Riemannian manifold $\mathscr{S}$ (intuitively, a manifold $\mathscr{M}$ that is a 'regular subset' of $\mathscr{S}$), for each point $p \in \mathscr{M}$, we may view $T_p\mathscr{M}$ as a linear subspace of $T_p\mathscr{S}$, and hence an inner product g_p on $T_p\mathscr{S}$ naturally defines an inner product on $T_p\mathscr{M}$. Then, $g|_{\mathscr{M}}$, denoting the induced inner product, defines a Riemannian metric on $\mathscr{M}$. This is a standard choice for n-manifolds embedded in $\mathbb{R}^l$, as the one shown in Figure 1.1. It is

common to say that we identify the tangent vectors of $\mathcal{M}$ with the associated elements of the embedding space $\mathbb{R}^l$.

1.2.4 Covariant derivative

In this section, we generalize the concept of directional derivative to vector fields. This plays an important role in defining 'straight lines' on a manifold $\mathcal{M}$.

Define an affine connection to be a map

$$\nabla : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M}),$$

written as $(X, Y) \mapsto \nabla_X Y$, with the following properties

- Linearity in X: $\nabla_{fX_1+gX_2} Y = f\nabla_{X_1} Y + g\nabla_{X_2} Y$ for $f, g \in C^\infty(\mathcal{M})$
- Linearity in Y: $\nabla_X aY_1 + bY_2 = a\nabla_X Y_1 + b\nabla_X Y_2$ for $a, b \in \mathbb{R}$
- Leibniz rule: $\nabla_X(fY) = f\nabla_X Y + (Xf)Y$ for $f \in C^\infty(\mathcal{M})$.

Despite the fact that a connection is defined by its action on objects globally defined over $\mathcal{M}$ it follows from its definition that $\nabla_X Y|_p$ depends only on the values of Y in a neighborhood of p and the value of X at p. Thus, we have that $\nabla_X Y|_p = \nabla_{X(p)} Y \in T_p\mathcal{M}$. This is a generalization of the directional derivative of Y at p in the direction of the vector X_p. $\nabla_X Y$ is called the covariant derivative of Y in the direction of X.

Let $\{E_i\}$ be a local frame, i.e. a set of local coordinates, for $T\mathcal{M}$ on an open subset $U \subset \mathcal{M}$. Expand $\nabla_{E_i} E_j$ in $\{E_i\}$:

$$\nabla_{E_i} E_j = \Gamma_{ij}^k E_k,$$

with Γ_{ij}^k, $i, j, k = 1, \ldots, n$ functions on U, called the Christoffel symbols. Using the representation of a vector field in the local coordinates $\{E_i\}$ we can expand $X, Y \in \mathfrak{X}$ as $X = X^i E_i$ and $Y = Y^j E_j$, and $\nabla_X Y$ as

$$\nabla_X Y = (XY^k + X^i Y^j \Gamma_{ij}^k) E_k.$$

For a manifold $\mathcal{M}$ embedded in a Euclidean space (this can be generalized, see e.g. Lee (1997)), by requiring that the affine connection ∇ must preserve the metric and must be torsion free, we have that Γ_{ij}^k can be uniquely determined. Under these hypotheses, ∇ is called the Levi-Civita connection.

A vector field along a curve $\gamma : I \to \mathcal{M}$ is a smooth map such that $V(t) \in T_{\gamma(t)}\mathcal{M}$ for every $t \in I$. An example of a vector field along a curve γ is its velocity vector: $\dot{\gamma}(t) \in T_{\gamma(t)}\mathcal{M}$ for each t. Let $\mathcal{M}$ be a manifold with an affine connection ∇, and let γ be a curve in $\mathcal{M}$. A vector field V along a curve γ is said to be parallel along γ with respect to ∇ if $\nabla_{\dot{\gamma}} V(t) = 0$. Note that there is a slight abuse of notation here, as we should instead use $\nabla_{\dot{\gamma}(t)} \tilde{V}$, with $\tilde{V}$ an extension of $\dot{\gamma}$ in the neighborhood of $\gamma(t)$. It can be shown that the particular extension chosen is not relevant. In practice, a connection defines a way to generalize parallel transport on a manifold. In fact, the parallel transport of a vector $u \in T_p\mathcal{M}$ along a curve γ can be defined as the collection of vectors u along the curve γ such that $\nabla_{\dot{\gamma}(t)} u = 0$, where $\dot{\gamma}(t) \in T_{\gamma(t)}\mathcal{M}$. A pictorial representation of the parallel transport of a tangent vector is given in Figure 1.4.

The acceleration of γ is the vector field $\nabla_{\dot{\gamma}}\dot{\gamma}$ along γ. We call γ a geodesic curve, with respect to ∇, if its acceleration is zero: $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$. For any $p \in \mathcal{M}$ and $V \in T_p\mathcal{M}$, there is a unique maximal geodesic (one that cannot be extended to any larger interval) $\gamma : I \to \mathcal{M}$ with $\gamma(0) = p$ and $\gamma(0) = V$, defined on some open interval I. This maximal geodesic is often called the geodesic with initial point p and initial velocity V, and is denoted with γ_V.

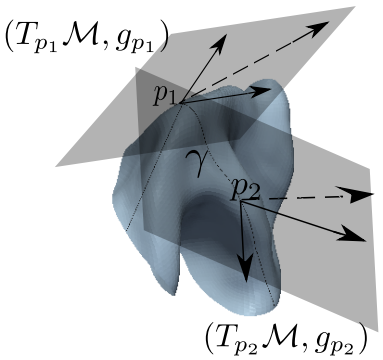


Figure 1.4 Parallel transport of a vector in $T_{p_1}\mathcal{M}$ onto $T_{p_2}\mathcal{M}$ along γ.

We have now defined geodesics, which are curves of zero acceleration, or alternatively (locally) shortest paths. We have also seen how to define lengths of curves. From a statistical prospective, geodesics play a fundamental role, as we can define a distance $d_{\mathcal{M}}(w, w')$ for $w, w' \in \mathcal{M}$ that is the length of the shortest curve between w and w'. Thanks to the concept of geodesic distance it is possible to introduce a geodesic mean for $w_1, \ldots, w_s \in \mathcal{M}$ as

$$\bar{w} = \underset{p \in \mathcal{M}}{\arg\min} \frac{1}{s} \sum_{i=1}^{s} d^2_{\mathcal{M}}(w_i, p),$$

which generalizes the Euclidean mean. The geodesic mean is a particular example of a Fréchet mean, which can instead be defined in any metric space $\mathscr{M}$ equipped with a distance d as

$$\bar{w} = \arg\min_{p \in \mathscr{M}} \frac{1}{s} \sum_{i=1}^{s} d^2(w_i, p),$$

with $w_1, \ldots, w_s \in \mathscr{M}$.

1.2.5 The Exponential and Logarithmic map

Here we define the exponential map, which is important as it allows us to represent elements in $\mathscr{M}$ in terms of associated elements on a tangent space, which is instead a linear space.

Define the subset $\mathscr{E}$ of $T\mathscr{M}$ as

$$\mathscr{E} = \{V \in T\mathscr{M} : \gamma_V \text{ is defined on an interval containing } [0,1]\}.$$

We define $\exp : \mathscr{E} \to \mathscr{M}$, the exponential map, to be

$$\exp(V) = \gamma_V(1).$$

For a $p \in \mathscr{M}$, we define the restricted exponential map $\exp_p$ to be the restriction of exp to the set $\mathbb{E}_p := \mathscr{E} \cap T_p\mathscr{M}$. For any $p \in \mathscr{M}$, there is a neighborhood V of the origin in $T_p\mathscr{M}$ and a neighborhood U of $p \in \mathscr{M}$ (called normal neighborhood) such that $\exp_p : V \to U$ is a diffeomorphism. Given an element in $V \subset T_p\mathscr{M}$ we will be able to associate (uniquely) an element in $U \subset \mathscr{M}$ and vice-versa. The inverse map from U to V is sometimes call logarithmic map. An illustration is given in Figure 1.5.

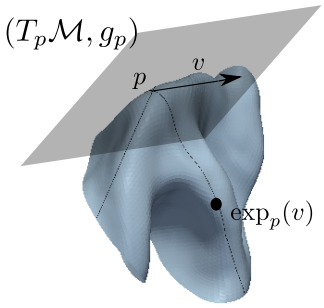


Figure 1.5 An element $\exp_p(v) \in \mathscr{M}$ generated by applying the exponential map $\exp_p$ to the tangent vector $v \in T_p\mathscr{M}$.

The exponential map can be used to define a set of 'special' coordinates on $\mathscr{M}$. In fact, an orthonormal basis $\{E_i\}$ for $T_p\mathscr{M}$ gives an isomorphism $E : \mathbb{R}^n \to T_p\mathscr{M}$ by $E(x_1, \cdot, x_n) = x_i E_i$. If U is a normal neighborhood of p, we can combine this isomorphism with the exponential map to get a coordinate chart $\varphi : U \to \mathbb{R}^n$ defined as

$$\varphi = E^{-1} \circ \exp_p^{-1} : U \to \mathbb{R}^n,$$

which offers a convenient way to parameterize $U \subset \mathscr{M}$ with a subset of $\mathbb{R}^n$. Any such coordinates are called Riemannian normal coordinates centered at p.

1.2.6 Differential operators on Riemannian manifolds

Now we are finally able to introduce generalizations of differential operators to functions or vector fields defined on manifolds. Specifically, we generalize the gradient, which for a function $f : \mathbb{R}^n \to \mathbb{R}$ is $\nabla = (\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n})$, the divergence operator, which for a vector field $(f^1, \ldots, f^n) : \mathbb{R}^n \to \mathbb{R}^n$ is $\mathrm{div} f = \frac{\partial f^1}{\partial x^1} + \ldots + \frac{\partial f^n}{\partial x^n}$ and the Laplacian, which for a function $f : \mathbb{R}^n \to \mathbb{R}$ is $\Delta f := \mathrm{div} \nabla f = \frac{\partial f}{\partial^2 x^1} + \cdots + \frac{\partial f}{\partial^2 x^n}$.

We also extend the definition of Laplacian to vector fields. Note that, for instance, in the planar 2D case a matrix operator for a vector field $u : \mathbb{R}^2 \to \mathbb{R}^2$ can be defined as the isotropic Laplacian operator

$$\begin{bmatrix} \Delta & 0 \\ 0 & \Delta \end{bmatrix},$$

where Δ is the Laplacian operator for real valued functions. The isotropic Laplacian applies the Laplacian operator component-wise to a vector field in $\mathbb{R}^2$, exploiting the fact that, in the Euclidean space $\mathbb{R}^2$ there is a global reference system. The introduction of an analogous operator for vector fields on a manifold is not straightforward for the main reason that nearby vectors, living on different tangent spaces, cannot be compared component-wise, as they are expressed in different local bases. The definition of such coordinate independent operators for vector fields require the additional notions of Riemannian geometry introduced in the previous sections.

Let now $f : \mathscr{M} \to \mathbb{R}$ be a real valued and smooth function on the manifold $\mathscr{M}$. Let $df_p(v)$ be its directional derivative at p in the direction $v \in T_p\mathscr{M}$. The concept of directional derivative can be used to define $(\nabla_{\mathscr{M}} f)(p) \in T_p\mathscr{M}$, the gradient of the function f at p, as the element of the tangent space that satisfies $g_p(\nabla_{\mathscr{M}} f(p), v) = df_p(v)$ for all $v \in T_p\mathscr{M}$, which is a well-known property of the gradient in Euclidean spaces. It can be

shown that $(\nabla_{\mathcal{M}} f)(p) \in T_p\mathcal{M}$ can be expressed in local coordinates $\{x^i\}$ as

$$(\nabla_{\mathcal{M}} f)(p) = g^{ij}(p)\left(\frac{\partial f}{\partial x^i}\right)_p \left(\frac{\partial}{\partial x^j}\right)_p.$$

Let V be a smooth vector field on the closed manifold $\mathcal{M}$, where $V(p) \in T_p\mathcal{M}$. A generalization of the divergence operator $\mathrm{div}_{\mathcal{M}}$ can be introduced by imposing that $\int_{\mathcal{M}} f \mathrm{div}_{\mathcal{M}} V = -\int_{\mathcal{M}} g_p(\nabla_{\mathcal{M}} f, V)$ for all $f \in C^\infty(\mathcal{M})$, this also a well-known property in the Euclidean case. As in the Euclidean case, we can finally define the Laplace-Beltrami operator as $\Delta_{\mathcal{M}} f = \mathrm{div}_{\mathcal{M}} \nabla_{\mathcal{M}} f$. This is related to the second partial derivatives of f on $\mathcal{M}$, i.e. its local curvature. In local coordinates it can be shown that $\Delta_{\mathcal{M}} f(p)$ is of the form

$$(\Delta_{\mathcal{M}} f)(p) = \frac{1}{\sqrt{\det(g_{ij}(p))}}\left(\frac{\partial}{\partial x^i}\right)_p g^{ij}\sqrt{\det(g_{ij}(p))}\left(\frac{\partial f}{\partial x^j}\right)_p,$$

with $\det(g_{ij}(p))$ denoting the determinant of the matrix $(g_{ij}(p))$.

When it comes to vector fields, we can define the Bochner-Laplacian operator, of a smooth section $v \in \mathfrak{X}(\mathcal{M})$, as

$$\Delta_{BL} = \nabla^*\nabla: \quad \mathfrak{X}(\mathcal{M}) \to \mathfrak{X}(\mathcal{M}) \tag{1.21}$$

where ∇^* is the L^2 adjoint of ∇.

The operators $\Delta_{\mathcal{M}} f$ and $\Delta_{BL} v$ define measures of local curvature and thus generalizations of the second derivative for real functions f and vector fields v, respectively. These will play an important role in defining smooth estimates for quantities that are real functions or vector fields supported on a manifold $\mathcal{M}$.

1.3 Geometric Functional Data Analysis

There are three main types of geometric functional data, reflecting how geometry and functional data interplay with each other. We briefly recall them here.

The first setting is that of samples that are real *functions supported on a non-linear manifold* $\mathcal{M} \subset \mathbb{R}^d$, i.e. functional samples $X_i : \mathcal{M} \to \mathbb{R}$ with $i = 1, \ldots, n$. An example of a functional sample in this setting is shown in the first panel of Figure 1.6. Due to the high-dimensionality of the data, and due to the presence of discrete noisy measurements or indirect measurements, it is often necessary to incorporate smoothness in the estimation

of the underlying functional unknowns. Thanks to the geometric operators introduced in Section 1.2, we will be able to achieve this in the context of FPCA.

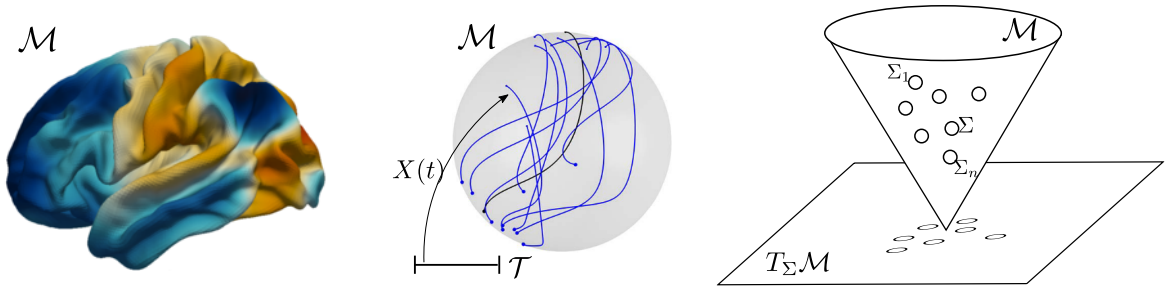


Figure 1.6 From left to right, three different examples of geometric functional data. On the first panel, a functional observation $f : \mathcal{M} \to \mathbb{R}$ supported on the manifold $\mathcal{M}$ representing the geometry of the brain surface. On the second panel, a set of curves $X(t)$ taking values on a manifold $\mathcal{M}$ that is a sphere, i.e. $X(t)|_t \in \mathcal{M}$ for all $t \in \mathcal{T}$. In the third panel, a pictorial representation of the setting of functional objects taking values in a non-linear space $\mathcal{M}$. An example is the space of covariance operators. All three types of geometric functional data are generally referred to as 'functional data on manifolds'.

The second setting is that of *manifold-constrained curves*, where each sample X_i is a curve $\{X_i(t), t \in [0,1]\}$ that takes values on a non-linear manifold $\mathcal{M} \subset \mathbb{R}^d$, specifically, $X_i(t) \in \mathcal{M}$ for all $t \in [0,1]$. An example of such data are flights trajectories, with $\mathcal{M}$ modeling the geometry of the earth, as shown in the second panel of Figure 1.6. In the context of FPCA, this setting has been considered in Dai and Müller (2018), Lin and Yao (2018) and Dai et al. (2018).

Finally, the third setting is that of *manifold-valued functional data*, where the non-linearity is not simply a point-wise non-linearity, as opposed to the second setting. Examples include samples $X_1, \ldots, X_n$ where each sample is constrained to be a monotonically increasing function, or a diffeomorphic function (smooth invertible map with smooth inverse), or a covariance operator; therefore constrained to be symmetric and positive semidefinite. A pictorial representation of this setting is shown on the third panel of Figure 1.6. Examples of manifold-valued shape representations for parametrized curves can be found in Kurtek et al. (2012); Srivastava et al. (2011). Examples of discrete shape representations with non-trivial geometries are: landmark based shape representations (see, e.g. Dryden and Mardia, 2016), skeletal shape representations (Pizer et al., 2013), dihedral angles representations (Eltzner et al., 2018) and projective shape spaces (Mardia and Patrangenaru, 2005).

The main methodological contributions of this thesis can be contextualized in the first and third settings, namely in the setting of functional data supported on manifolds

and manifold-valued functional data. Therefore, a review of the literature on the first setting has been left to Chapter 2, while the third setting is covered in the second half of Chapter 3, for functional data that are covariance operators, and in Chapter 4, for functional data that are diffeomorphic functions. To give a more comprehensive view of the recent works on the analysis of geometric functional data, in the next section, we also review the manifold-constrained curves setting, which is the setting this thesis is only tangentially related to.

1.3.1 Manifold-constrained curves

Let $\mathcal{M}$ denote a d-dimensional, geodesically complete (i.e. the exponential map is defined on the entire tangent space) Riemannian manifold embedded in $\mathbb{R}^{d_0}$, e.g. a two-dimensional sphere in $\mathbb{R}^3$, and let $\mathcal{T} \subset \mathbb{R}$ be a compact interval of the real line. A first branch of the literature on geometric functional data has focused on samples that are random continuous curves with values on $\mathcal{M}$, i.e. random continuous functions from $\mathcal{T}$ to $\mathcal{M}$, an example of which is shown on the second panel of Figure 1.6.

In Dai and Müller (2018), an intrinsic Fréchet mean $\mu_{\mathcal{M}}$ is considered, and this is defined, point-wise, as

$$\mu_{\mathcal{M}}(t) = \underset{p \in \mathcal{M}}{\arg\min}\, \mathbb{E}[d_{\mathcal{M}}(X(t), p)^2],$$

with $d_{\mathcal{M}}$ denoting the geodesic distance on $\mathcal{M}$. Existence and uniqueness of $\mu_{\mathcal{M}}$ is assumed to hold.

Define the log-mapped random functions

$$V(t) = \log_{\mu_{\mathcal{M}}(t)}(X(t)) \in T_{\mu_{\mathcal{M}}(t)}\mathcal{M},$$

which can be shown to have zero mean (Patrangenaru and Bhattacharya, 2003). Thanks to the fact that the manifold is embedded in $\mathbb{R}^{d_0}$, the elements of the tangent spaces can be identified with $\mathbb{R}^{d_0}$. Hence, $V(t) \in \mathbb{R}^{d_0}$ for all $t \in \mathcal{T}$. Dai and Müller (2018) perform classical FPCA analysis on $V(t)$, which for each $t \in \mathcal{T}$ takes values on the tangent space $T_{\mu_{\mathcal{M}}(t)}\mathcal{M}$ and thus belongs to a linear space. Thus, they obtain a KL expansion of V that is

$$V(t) = \sum_{k=1}^{\infty} \zeta_k \psi_k(t).$$

From this they define the truncated approximation of the original process to be

$$X_K(t) = \exp_{\mu_{\mathscr{M}}(t)}\Big(\sum_{k=1}^{K} \zeta_k \psi_k(t)\Big).$$

1.3.2 Estimation

The case of fully observed curves

Suppose we are given n independent fully observed functional samples $X_1, \ldots, X_n$ identically distributed as X, the Fréchet mean is estimated by

$$\hat{\mu}_{\mathscr{M}}(t) = \underset{p \in \mathscr{M}}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} d_{\mathscr{M}}(X_i(t), p).$$

Existence and uniqueness of $\hat{\mu}_{\mathscr{M}}$ are assumed.

The log-mapped data are estimated by $\hat{V}_i(t) = \log_{\hat{\mu}(t)}(X_i(t))$. These are such that $n^{-1}\sum_{i=1}^{n} \hat{V}_i(t) = 0$ (Patrangenaru and Bhattacharya, 2003). The estimates of the PC functions $\{\hat{\psi}_k\}$ are computed by eigen-analysis of the empirical covariance function $\hat{C}_V = n^{-1}\sum_{i=1}^{n} \hat{V}_i(t)\hat{V}_i(s)^T$, for $s, t \in \mathscr{T}$; The estimates of the PC scores are given by $\hat{\zeta}_{ik} = \int_{\mathscr{T}} \hat{V}_i^T(t)\hat{\psi}_k(t)dt$. The estimated truncated expansions take the form

$$\hat{V}_K = \sum_{k=1}^{K} \hat{\zeta}_k \hat{\psi}_k(t), \qquad \hat{X}_K(t) = \exp_{\hat{\mu}_{\mathscr{M}}(t)}\Big(\sum_{k=1}^{K} \hat{\zeta}_k \hat{\psi}_k(t)\Big).$$

Consistency and $\sqrt{n}$ rates of convergence are obtained for the empirical mean, i.e.

$$\sup_{t \in \mathscr{T}} d_{\mathscr{M}}(\hat{\mu}_{\mathscr{M}}(t), \mu_{\mathscr{M}}(t)) = O_p(n^{-1/2}),$$

from which it follows that

$$\sup_{t \in \mathscr{T}} \|\hat{V}_i(t) - V_i(t)\|_2 = O_p(n^{-1/2}),$$

with $\|\cdot\|_2$ denoting the Euclidean distance. Moreover, $\sqrt{n}$ rates of convergence are obtained for the empirical covariance function, i.e.

$$\sup_{t,s \in \mathscr{T}} \|\hat{C}_V(t,s) - C_V(t,s)\|_F = O_p(n^{-1/2}),$$

with $\|\cdot\|_F$ denoting the Frobenius norm. From this, consistency of eigenvalues and eigenfunctions (defined on the tangent space, which is a linear space) follow by classical arguments (see Section 1.1.2).

The use of the discrepancy measures $\|\hat{V}_i(t) - V_i(t)\|_2$ and $\|\hat{C}_V(t,s) - C_V(t,s)\|_F$, is possible thanks to the identification of the elements of the tangent space with the embedding space $\mathbb{R}^{d_0}$. In general, for manifolds not embedded in $\mathbb{R}^{d_0}$, this is not possible as tangent spaces at different points, are generally in different spaces, which are not directly comparable without the introduction of further geometric notions (see discussion in Section 1.2.4).

In Lin and Yao (2018), a very similar setting to Dai and Müller (2018) is considered and the same estimators are proposed. However, they replace the Euclidean discrepancy measure with a fully intrinsic discrepancy measure, based on the idea of parallel transport. This enables statistical analysis on a manifold $\mathcal{M}$ not necessarily embedded in a Euclidean space.

Define two measurable curves $\alpha(t)$ and $\beta(t)$ on $\mathcal{M}$, and two vector fields $U(t)$ and $V(t)$, such that $U(t) \in T_{\alpha(t)}\mathcal{M}$ and $V(t) \in T_{\beta(t)}\mathcal{M}$ for all $t \in \mathcal{T}$. They define the discrepancy vector field

$$U \ominus_\Gamma V := \Gamma_{\alpha,\beta} U - V,$$

with $\Gamma_{\alpha,\beta}$ the parallel transport operator from $T_{\alpha(t)}\mathcal{M}$ to $T_{\beta(t)}\mathcal{M}$, along the geodesics from $\alpha(t)$ to $\beta(t)$. From $\Gamma_{\alpha,\beta}$, we can define an operator $\Phi_{\alpha,\beta}$ such that, for a linear continuous operator $\mathscr{A}$, acting on vector fields along α, $(\Phi_{\alpha,\beta}\mathscr{A})V = \Gamma_{\alpha,\beta}\mathscr{A}\Gamma^*_{\alpha,\beta}V$. Here, $\Gamma^*_{\alpha,\beta}$ denotes the L^2 adjoint operator of $\Gamma_{\alpha,\beta}$. We can define a discrepancy operator, between $\mathscr{A}$, acting on vector fields along α, and $\mathscr{B}$, acting on vector fields along β, as

$$\mathscr{A} \ominus_\Phi \mathscr{B} := \Psi_{\alpha,\beta}\mathscr{A} - \mathscr{B}.$$

The discrepancy between the vector fields U and V is finally measured as $\|U \ominus_\Gamma V\|_\beta$ and the discrepancy between the Hilbert-Schmidt operators $\mathscr{A}$ and $\mathscr{B}$ is measured as $|||\mathscr{A} \ominus_\Phi \mathscr{B}|||_\beta$, with $\|\cdot\|_\beta$ denoting the L^2 norm of the vector field along β and $|||\cdot|||_\beta$ denoting the associated Hilbert-Schmidt norm.

In terms of these discrepancy measures, the $\sqrt{n}$ rates of convergence are recovered for the empirical covariance function, i.e.

$$|||\hat{\mathscr{C}}_V \ominus_\Phi \mathscr{C}_V|||_\mu = O_p(n^{-1/2}),$$

where $\hat{\mathscr{C}}_V$ is 'transported' from being an operator acting on vector fields along $\hat{\mu}(t)$ to an operator acting on vector fields along $\mu(t)$. From this, consistency of eigenvalues and eigenfunctions follow by classical arguments.

Discrete measurements

Suppose now the random curves $X_1, \ldots, X_n$, with values on $\mathscr{M}$, are not directly observable, but instead, each sample X_i is measured at time points t_{ij}, with $j = 1, \ldots, p_i$, and the measurements are contaminated with observational error. Specifically, we observe

$$\{(Y_i(t_{ij}), t_{ij}), i = 1, \ldots, n; j = 1, \ldots, p_i\},$$

with $Y_i(t_{ij}) \in \mathscr{M}$ and $t_{ij} \sim f$, with f a density supported in $\mathscr{T}$ and independent of X_i. The observations are generated by

$$Y_{ij} = \exp_{\mu_{\mathscr{M}}(t_{ij})}\{V_i(t_{ij}) + \varepsilon_i(t_{ij})\},$$

with $V_i(t_{ij}) = \log_{\mu_{\mathscr{M}}(t_{ij})}(X_i(t_{ij}))$ and $\varepsilon_i(t_{ij}) \in T_{\mu_{\mathscr{M}}(t_{ij})}\mathscr{M}$ independent error terms, with isotropic variance σ^2 and $\mathbb{E}[\varepsilon_i(t_{ij})|t_{ij}] = 0$. This is the setting considered in Dai et al. (2018).

Although a Smooth-then-Estimate approach could be adopted to individually reconstruct $X_1, \ldots, X_n$, from the noisy observations, this might be ineffective for very sparsely sampled functions. For this reason, Dai et al. (2018) propose an Estimate-then-Smooth approach. To estimate the mean, a geodesic version of the local polynomial regression (Petersen and Müller, 2019) is applied to the scattered measurements

$$\{(Y_i(t_{ij}), t_{ij}), i = 1, \ldots, n; j = 1, \ldots, p_i\},$$

from which the mean function $\hat{\mu}_{\mathscr{M}}(t)$ is estimated. Once the mean function is estimated, the original data are mapped onto the tangent spaces $T_{\hat{\mu}_{\mathscr{M}}(t_{ij})}\mathscr{M}$, centered at $\hat{\mu}_{\mathscr{M}}(t_{ij})$, leading to the point-wise estimates $\hat{V}_i(t_{ij}) = \log_{\hat{\mu}_{\mathscr{M}}(t_{ij})} Y_{ij}$. Local polynomial regression, for matrix-valued functions, is then applied to the scatterplot

$$\left\{\left(\hat{V}_i(t_{ij})\hat{V}_i(t_{ij})^T, t_{ij}, t_{il}\right), i = 1, \ldots, n; j, l = 1, \ldots, p_i, j \neq l\right\}.$$

Scores prediction, as in the linear case, is tackled by applying, on the tangent space representations, the Best Linear Unbiased Predictor introduced in (Yao et al., 2005).

To the best of our knowledge a Regularized-Estimation approach has not yet been employed in the context of discrete noisy measurements of curves taking values on a manifold $\mathscr{M}$. The setting of indirect measurements has also not been explored, although this does not appear to be relevant to the applications considered in this thesis.

Chapter 2

Functional Principal Component Analysis over manifolds

2.1 Motivation

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 to enhance our present understanding of the brain.

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 it is beneficial to analyze neuroimaging data through the processing of the signals on the cortical surface using surface-constrained techniques.

The goal in this chapter is to introduce a novel Functional Principal Component Analysis (FPCA) 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 used 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 neural activity in the cerebral cortex is generally mapped on a common template cortical surface, to allow multi-subject statistical analysis.

In this work, 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. The resulting FC map for each subject 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 $\mathscr{M} \subset \mathbb{R}^3$ modelling the cortical surface. See Section 1.2 for a formal definition of a Riemannian manifold. Each resting state FC map can be represented by a function $x_i : \mathscr{M} \to \mathbb{R}$. Once we have the correlation maps on the cortical surface we want to study how the phenomena vary from subject to subject. A statistical technique for this study is Functional Principal Component Analysis (PCA). It is natural to contextualize this task in the framework of Geometric Functional Data Analysis for functions supported on a non-linear manifold.

In Section 2.2, we formalize the setting considered in this chapter and give a brief overview of the literature on functional data supported on non-linear manifold domains. In Section 2.3 we introduce a novel FPCA model and propose an algorithm for its resolution. We then give some simulation results in Section 2.4, indicating the performance of our methodology, as compared to other methods in literature. We then return to the FC maps

example in Section 2.5, to consider how the surface based FPCA analysis might be used in this case and draw some concluding remarks in Section 2.6. Further simulations are provided in Appendix A.

2.2 Setting

Consider the space of square integrable functions on $\mathcal{M}$: $L^2(\mathcal{M}) = \{f : \mathcal{M} \to \mathbb{R} : \int_{\mathcal{M}} |f(p)|^2 dp < \infty\}$ with the inner product $\langle f, g \rangle_{\mathcal{M}} = \int_{\mathcal{M}} f(p)g(p)dp$ and norm $\|f\|^2_{\mathcal{M}} = \int_{\mathcal{M}} |f(p)|^2 dp$. We let X be a random variable with values in $L^2(\mathcal{M})$, mean $\mu = \mathbb{E}[X]$ and finite second moment, i.e. $\int_{\mathcal{M}} \mathbb{E}[X^2] < \infty$, and assume that its covariance function $K(p,q) = \mathbb{E}[(X(p) - \mu(p))(X(q) - \mu(q))]$ is square integrable.

The function space $L^2(\mathcal{M})$ is a linear space and the theoretical setup introduced in Section 1.1.1 readily applies to this setting. Indeed, we can perform FPCA on n fully observed smooth samples from X, here denoted with $x_1, \ldots, x_n$, by applying the estimators in Section 1.1.2, for which the same rates of convergence hold.

In practice, for each sample x_i, only noisy evaluations $x_i(p_j)$ on a fixed discrete grid of points $p_1, \ldots, p_s \subset \mathcal{M}$ are given. In this setting, we want take take advantage of the smoothing properties of the underlying functional samples to obtain estimators with better finite sample properties. As mentioned in Chapter 1, there are three approaches to this problem: the Smooth-then-Estimate, the Estimate-then-Smooth and the Regularized-Estimate approach.

The Smooth-then-Estimate approach consists of individually smoothing the samples $x_1, \ldots, x_n$ and then performing FPCA. We will compare the methodology presented in this chapter with this approach, demonstrating that smoothing the samples individually generally leads to poorer performances. The Estimate-then-Smooth approach consists of computing the discrete estimators

$$\bar{x}(p_j) = \frac{1}{n}\sum_{i=1}^{n} x_i(p_j), \qquad \hat{C}_X(p_j, p_l) = \frac{1}{n}\sum_{i=1}^{n} \big(x_i(p_j) - \bar{x}(p_j)\big)\big(x_i(p_l) - \bar{x}(p_l)\big),$$

followed by a smoothing on $\bar{x}(p_j)$ and $\hat{C}_X(p_j, p_l)$. However, due to the potentially high number of locations $p_1, \ldots, p_s$, as it is the case in some neuroimaging applications, it might not be possible to store in memory the entire object $\hat{C}_X(p_j, p_l)$, limiting the applicability of the Estimate-then-Smooth approach, as presented for 1D functional data. This motivates the introduction of a Regularized-Estimate approach to FPCA, where instead of estimating the covariance $\hat{C}_X$, we estimate directly, in a regularized fashion, its eigenfunctions.

The first two approaches to FPCA rely on a smoothing procedure for noisy measurements of a function $x : \mathcal{M} \to \mathbb{R}$, at locations $p_1, \ldots, p_s \subset \mathcal{M}$, where the natural concept of neighborhood on the domain $\mathcal{M}$ is induced by the 2D geodesic distance. This is also the case for the Regularized-Estimate approach as it will be clear in the next section.

With the aim of defining a Regularized-Estimate FPCA model, in the next section, we introduce a generalization of the smoothing splines formulation (1.11). In the simpler setting where the aim is smoothing a single function $x : \mathcal{M} \to \mathbb{R}$, observed at locations $\{p_j\}$, this takes the form

$$\hat{x} = \underset{f:\mathcal{M}\to\mathbb{R}}{\arg\min} \sum_{j=1}^{s} \Big(x(p_j) - f(p_j)\Big)^2 + \lambda \int_{\mathcal{M}} (\Delta_{\mathcal{M}} f)^2 dp,$$

where $\Delta_{\mathcal{M}}$ is the Laplace-Beltrami operator, measuring the curvature of the function f on the manifold. Further details on this approach will follow in the next section. This formulation overcomes the limitations of the kernel regression and heat kernel approaches to smoothing on a non-linear manifold, which we briefly review here.

A kernel regression estimator, for the estimation of $x : \mathcal{M} \to \mathbb{R}$ from noisy evaluations $\{x(p_j)\}$, has been first proposed in Pelletier (2006). Let $K : \mathbb{R}_+ \to \mathbb{R}_+$ be a 1D positive and continuous kernel, the proposed kernel regression estimator takes the form

$$x(p) = \frac{1}{s} \sum_{j=1}^{s} x(p_j) \frac{1}{h^d} \frac{1}{\theta_{p_j}(p)} K\left(\frac{d_{\mathcal{M}}(p, p_j)}{h}\right), \qquad p \in \mathcal{M}$$

with $\frac{1}{\theta_{p'}(p)}$ a function normalizing the densities $K\left(\frac{d_{\mathcal{M}}(\cdot, p)}{h}\right)$ to sum to 1, h a positive constant tuning parameter and with $d_{\mathcal{M}}$ denoting the geodesic distance on $\mathcal{M}$. This approach becomes infeasible for a generic manifold due to the necessity to compute the normalizing function $\theta_{p'}(p)$, which can in general be defined only on a neighborhood of p. In the special cases where $\mathcal{M}$ is a sphere or $\mathcal{M}$ is the space of symmetric positive-definite matrices we can find an explicit computation of $\theta_{p'}(p)$ in Pelletier (2006) and Chevallier et al. (2017), respectively. An alternative formulation has been proposed in Kim and Park (2013), where the kernels are defined on the tangent space of each location $\{p_j\}$ and are applied to the tangent space representations of the neighboring locations. The applicability of this approach to generic non-linear domains is also limited, due to the local nature of the tangent space representations.

Alternative approaches to smoothing are based on the construction of heat kernels, which are solutions to the heat equation on a manifold $\mathcal{M}$. The heat kernel smoothing

of a non-smooth function x, is given by $(K_\eta * x)(p) = \int_{\mathscr{M}} K_\eta(p,q)x(q)dq$, where η is the smoothing parameter and K_η is the heat kernel, which can also be expressed in terms of the spectrum of the Laplace-Beltrami operator $\{(\lambda_l, \psi_l)\}$ as

$$K_\eta(p,q) = \sum_{l=1}^{\infty} = e^{-\lambda_l \eta}\psi_l(p)\psi_l(q).$$

It can be shown that for η small and for q close to p we have that

$$K_\eta(p,q) \approx \frac{1}{(2\pi\eta)^{\frac{1}{2}}} exp\left(-\frac{d^2_{\mathscr{M}}(p,q)}{2\eta^2}\right). \tag{2.1}$$

In Chung et al. (2005), the authors rely on this approximation to define a smoother for the discrete measurements $\{x(p_j)\}$ as $\hat{x}(p) = \sum_j \tilde{K}_\eta(p,p_j)x(p_j)$, with $\tilde{K}_\eta$ the approximation (2.1), truncated to zero for p and q that are not 'close' to each other. This is called the Iterated Heat Kernel (IHK) smoother as the desired level of smoothing can be reached after k iterations, thanks to the following property:

$$K^k_\eta * f = \underbrace{K_\eta * \ldots \times K_\eta}_{k \text{ times}} * f = K_{\sqrt{k}\eta}.$$

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). Instead, in Seo et al. (2010), the function x is expanded on the eigenfunctions of the heat kernel as $x = \sum_l \beta_l \psi_l$. Its smoothed version is given by $(K_\eta * x)(p) = \sum_l e^{-\lambda_l \eta}\beta_l \psi_l(p)$. The coefficients $\{b_l\}$ are estimated by least squares. Note that in both cases we need dense measurements of x, as the approximation (2.1) holds only for p and q close to each other and the least square estimate of the coefficients β_l is accurate only for densely observed functions.

2.3 Smooth FPCA over two-dimensional manifolds

2.3.1 Model

Without loss of generality, we now assume that the random function X has zero-mean. Suppose the sample of n functions $x_i : \mathscr{M} \to \mathbb{R}$ is observed at a fixed set of points $p_1, \ldots, p_s$ in $\mathscr{M}$ (this will be relaxed later). Define the $n \times s$ matrix $\mathbf{X} = (x_i(p_j))$.

Let $f : \mathscr{M} \to \mathbb{R}$ be a real valued and twice differentiable function and let $\mathbf{u} = \{u_i\}_{i=1,\ldots,n}$ be a n-dimensional real column vector. We propose to estimate the first PC function

$\hat{f}: \mathcal{M} \to \mathbb{R}$ and the associated PC scores vector $\hat{\mathbf{u}}$, by solving the equation

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{s} (x_i(p_j) - u_i f(p_j))^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f, \tag{2.2}$$

where the Laplace-Beltrami operator is integrated over the manifold $\mathcal{M}$, enabling a global roughness penalty on f. The empirical term encourages f to capture the strongest mode of variation, as it is the empirical counterpart of the objective function $\mathbb{E} \int_{\mathcal{M}} \{X - \langle X, f \rangle f\}^2$, minimized by the first PC function (see Section 1.1, for details). 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 ($\mathbf{u} \to c\mathbf{u}, f \to \frac{1}{c} f$), with c a constant, and the transformation ($\mathbf{X} \to c\mathbf{X}, \mathbf{u} \to c\mathbf{u}$). Then the objective function in (2.2) 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 the model introduced as Smooth Manifold FPCA (SM-FPCA).

2.3.2 Iterative algorithm

Here we present the numerical algorithm for the resolution of the model introduced above. Our approach for the minimization of the functional (2.2) can be summarized in the following two steps:

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

Let $\mathbf{f}_s$ be the vector of length s such that $\mathbf{f}_s = (f(p_1), \ldots, f(p_s))^T$. The expression in (2.2) can be rewritten as

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\arg\min} \|\mathbf{X} - \mathbf{u}\mathbf{f}_s^T\|_F^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f, \tag{2.3}$$

with $\|\cdot\|_F$ the Frobenius matrix norm. 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 (2.3). In particular we set the constraint $\|\mathbf{u}\|_2 = 1$, with $\|\cdot\|_2$ the Euclidean norm, as this allows us to leave the infinite-dimensional optimization in f unconstrained.

Our proposal for the minimization of the criterion (2.3) 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 (2.3) is

$$\mathbf{u} = \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{f}_s\|_2^2 + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f}, \tag{2.4}$$

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

$$\mathbf{u} = \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{X}\mathbf{f}_s\|_2}. \tag{2.5}$$

Step 2 Estimation of f given $\mathbf{u}$. For a given $\mathbf{u}$, solving (2.3) with respect to f is equivalent to finding the function f that minimizes

$$J_{\lambda,\mathbf{u}}(f) = \mathbf{f}_s^T \mathbf{f}_s + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f - 2\mathbf{f}_s^T \mathbf{X}^T \mathbf{u}. \tag{2.6}$$

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 $p_1, \ldots, p_s$. The problem in Step 2 is not trivial, consisting in an infinite-dimensional minimization problem. Let z_j denote the jth element of the vector $\mathbf{X}^T\mathbf{u}$, then minimizing the functional in (2.6) is equivalent to minimizing

$$\sum_{j=1}^{s} \Big(z_j - f(p_j)\Big)^2 + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \tag{2.7}$$

This problem involves estimating a smooth field f defined on a manifold, starting from noisy observations z_j at points p_j. As already mentioned, in the case of real functions defined on the real line, adopting a penalty of the form $\lambda \int f''$, the minimization problem turns out to have a finite-dimensional closed form solution that is a cubic spline [Green and Silverman (1994)]. For real functions defined on an Euclidean space, cubic splines are generalized by thin-plate splines. In this case, for an opportune smoothing penalty,

the solution of the minimization problem can be expressed in terms of a finite linear combination of radial basis functions [Duchon (1977)].

However, the case of real functions defined on a non-Euclidean domain $\mathscr{M}$ is more involved. In the special case where $\mathscr{M}$ is a sphere or a sphere-like surface, that is $\mathscr{M} = \{\sigma(v) = \rho(v)v : v \in S\}$ where $S \subset \mathbb{R}^3$ is the unit sphere centered at the origin, this smoothing problem has been considered, among others, by Wahba (1981) and Alfeld et al. (1996). Moreover, the functional (2.7) is considered, among others, by Ettinger et al. (2016) and Dassi et al. (2015). Here $\mathscr{M}$ is respectively a manifold homeomorphic to an open ended cylinder and a manifold homeomorphic to a sphere. In the latter 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. (2016) and Dassi et al. (2015), we construct here a FE space directly on the triangulated surface $\mathscr{M}_{\mathscr{T}}$ that approximates the manifold $\mathscr{M}$, i.e. we use surface FE, avoiding any flattening step and thereby allowing the formulation to be applicable to any manifold topology.

2.3.3 Surface Finite Element discretization

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

$$\hat{f} = \underset{f \in H^2(\mathscr{M})}{\arg\min}\, J_{\lambda,\mathbf{u}}(f). \tag{2.8}$$

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

$$\sum_{j=1}^{s} \varphi(p_j)\hat{f}(p_j) + \lambda \int_{\mathscr{M}} \Delta_{\mathscr{M}}\varphi \Delta_{\mathscr{M}}\hat{f} = \sum_{j=1}^{s} \varphi(p_j) \sum_{i=1}^{n} x_i(p_j)u_i \tag{2.9}$$

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

The key idea is to minimize $J_{\lambda,\mathbf{u}}(f)$ by differentiating this functional with respect to f. This leads to (2.9), that characterizes the estimate $\hat{f}$ as the solution of a linear fourth-order

problem. A more general version of Proposition 2 will be introduced in Chapter 3 in an inverse problems setting.

Consider now a triangulated surface $\mathscr{M}_{\mathscr{T}}$, union of the finite set of triangles $\mathscr{T}$, giving an approximated representation of the manifold $\mathscr{M}$. Figure 2.1 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 $\mathscr{M}_{\mathscr{T}}$ that are linear affine where restricted to any triangle τ in $\mathscr{T}$, i.e.

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

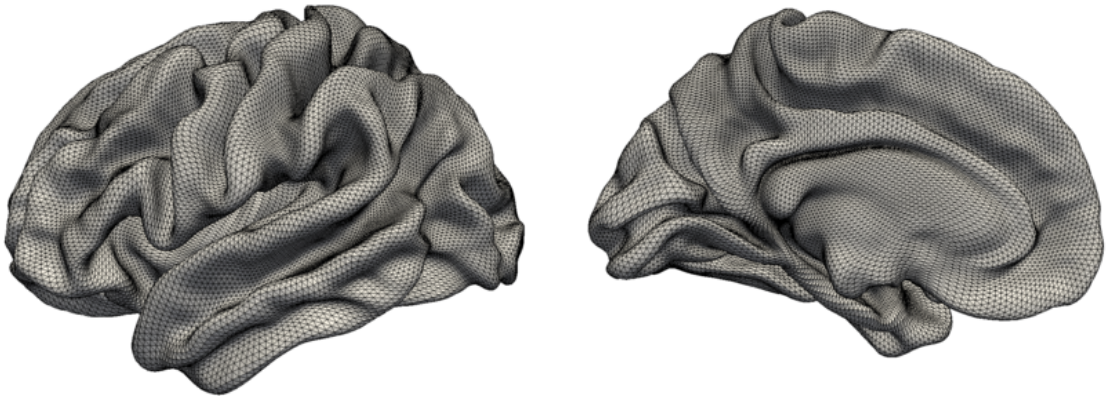


Figure 2.1 The triangulated surface approximating the left hemisphere of the template brain. The mesh is composed by 32K nodes and by 64K triangles

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

$$f(p) = \sum_{k=1}^{K} f(\xi_k)\psi_k(p) = \mathbf{f}^T\psi(p) \tag{2.10}$$

for each $p \in \mathscr{M}_{\mathscr{T}}$. The surface finite element space provides a finite dimensional subspace of $H^1(\mathscr{M})$ [Dziuk (1988)]. To use this finite element space to discretize the infinite-dimensional problem (2.9), that is well posed in $H^2(\mathscr{M})$, we first need a reformulation of (2.9) that involves only first-order derivatives. This can be obtained by introducing an auxiliary function g that plays the role of $\Delta_{\mathscr{M}} f$, splitting the equation (2.9) 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 Section 3.8, in the more general setting of

Chapter 3. The discrete estimators $\hat{f}_h, \hat{g}_h \in V$ are then obtained by solving

$$\begin{cases} \int_{\mathcal{M}_\mathcal{T}} \nabla_{\mathcal{M}_\mathcal{T}} \hat{f}_h \nabla_{\mathcal{M}_\mathcal{T}} \varphi_h - \int_{\mathcal{M}_\mathcal{T}} \hat{g}_h \varphi_h = 0 \\ \lambda \int_{\mathcal{M}_\mathcal{T}} \nabla_{\mathcal{M}_\mathcal{T}} \hat{g}_h \nabla_{\mathcal{M}_\mathcal{T}} v_h + \sum_{j=1}^{s} \hat{f}_h(p_j) v_h(p_j) = \sum_{j=1}^{s} v_h(p_j) \sum_{i=1}^{n} x_i(p_j) u_i \end{cases} \tag{2.11}$$

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 $\mathbf{R}_0 = \int_{\mathcal{M}_\mathcal{T}} (\psi \psi^T)$ and $\mathbf{R}_1 = \int_{\mathcal{M}_\mathcal{T}} (\nabla_{\mathcal{M}_\mathcal{T}} \psi)(\nabla_{\mathcal{M}_\mathcal{T}} \psi)^T$. Then, exploiting the representation (2.10) of functions in V we can rewrite (2.11) as a linear system. Specifically the Finite Element solution $\hat{f}_h(p)$ of the discrete counterpart (2.11) is given by $\hat{f}_h(p) = \psi(p)^T \hat{\mathbf{f}}$ where $\hat{\mathbf{f}}$ is the solution of

$$\begin{bmatrix} \Psi^T \Psi & \lambda \mathbf{R}_1 \\ \lambda \mathbf{R}_1 & -\lambda \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \Psi^T \mathbf{X}^T \mathbf{u} \\ \mathbf{0} \end{bmatrix} \tag{2.12}$$

Solving (2.12) leads to

$$\hat{\mathbf{f}} = (\Psi^T \Psi + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1)^{-1} \Psi^T \mathbf{X}^T \mathbf{u}. \tag{2.13}$$

Although this last formula is a compact expression of the solution, it is preferable to compute the solution from the linear system (2.12) due to the sparsity property of the matrix in the left-hand side. As an example, in the simulations and the application shown in Sections 2.4-2.5, respectively less then 1% and less then 0.1% of the elements in the matrix in the left hand side of (2.12) 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 are sampled on the common set of points $p_1, \ldots, p_s \in \mathcal{M}$. Suppose moreover, $p_1, \ldots, p_s \in \mathcal{M}$ coincide with the vertices of the triangulated surface $\mathcal{M}_\mathcal{T}$. In this particular case, an alternative approach could consist of interpreting the points $p_1, \ldots, p_s \in \mathcal{M}_\mathcal{T}$ as the nodes of a graph linked by the edges of the triangulation and considering the model (2.2) with a discrete smoothness operator term instead of the Laplace-Beltrami operator (see e.g. Belkin and Niyogi (2002) for the choice of the penalization term and Cai et al. (2011) for an application to matrix decomposition). However, thanks to its functional nature, the formulation (2.2) can be easily extended to the case of missing data or sparsely sampled functional data. Specifically, suppose now that each function x_i is observable on a set of points $p_1^i, \ldots, p_{s_i}^i$, then the natural extension of the model (2.2) becomes

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{s_i} (x_i(p_j^i) - u_i f(p_j^i))^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \tag{2.14}$$

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

Step 1 For a given f, the unitary-norm vector $\mathbf{u}$ minimizing (2.14) is given by

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

Step 2 For a given $\mathbf{u}$, the function f minimizing (2.14) is given by
$f = \mathbf{f}^T \psi$ 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

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

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

2.3.4 SM-FPCA Algorithm

The algorithm for the resolution of the model SM-FPCA (2.2) 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: $\mathbf{X} = \mathbf{UDV}^T$

(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:

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

The problems (2.2)-(2.14) are non-convex minimization problems 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.

2.3.5 Parameters selection

The SM-FPCA model has 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 of a K-fold cross validation. The data matrix $\mathbf{X}$ is partitioned by rows into K roughly equal groups. For each group of data $k = 1, \ldots, K$ the dataset can be split into a validation set $\mathbf{X}^k$, composed of the elements of the kth 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 (2.4), where $\int_{\mathcal{M}} \Delta^2_{\mathcal{M}}$ 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:

$$CV(\lambda) = \sum_{k=1}^{K} \frac{\sum_{i=1}^{n} \sum_{j=1}^{s} \left(x_i(p_j) - u_i^k f^{-k}(p_j) \right)^2}{np}. \tag{2.15}$$

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

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

The GCV score represents the average misfit of the regression model with a leave-one-out cross-validation strategy on the observations' vector $\mathbf{X}^T \mathbf{u}$. However, excluding the ith element from the vector $\mathbf{X}^T \mathbf{u}$ can be interpreted as removing ith column from the data-matrix $\mathbf{X}$. Thus, in terms of the data-matrix, this strategy can be interpreted as a leave-one-column-out cross-validation strategy, as opposed to the K-fold, where the data matrix $\mathbf{X}$ is partitioned by rows. The GCV approach is generally faster then the K-fold approach. However, K-fold 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.

2.3.6 Total explained variance

Another parameter that must be chosen is the number of PCs that satisfactorily reduces the dimension of the data. A classical approach consists of selecting this parameter on the basis of cumulative 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. It is nevertheless possible to define an index of explained variance as follows. 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. Since in our estimation procedure the PC scores are normalized to have unitary norm, the variance of the PCs is captured by the PC functions. It is thus necessary to consider here the unnormalized PC scores, obtained by multiplying each score vector by the norm of the associated PC function. 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. (2006) propose to remove linear dependence between correlated PC scores vectors, by regression projection. Thus they compute the QR decomposition of $\hat{\mathbf{U}}$ as $\hat{\mathbf{U}} = \mathbf{QR}$ and define the *adjusted total variance* as $\sum_{j=1}^{k} \mathbf{R}_{jj}^2$, where $\mathbf{R}_{jj}$ represents the variance explained by the jth PC that is not already explained by the previous $j-1$ components.

2.4 Simulation studies

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

We consider as domain of the functional observations a triangulated surface $\mathscr{M}_{\mathscr{T}}$ with 642 nodes that approximate the brainstem. On this triangulated surface we generate the orthonormal functions $\{v_l\}_{l=1,2,3}$, consisting of three eigenfunctions of the Laplace-Beltrami operator, as shown in Figure 2.2. These functions represent the first three PC functions. We then generate $n = 50$ smooth functions $x_1, \dots, x_{50}$ on $\mathscr{M}_{\mathscr{T}}$ by

$$x_i = u_{i1}v_1 + u_{i2}v_2 + u_{i3}v_3 \quad i = 1, \dots, n, \tag{2.16}$$

where u_{i1}, u_{i2}, u_{i3} are independent random variables that represent the scores and are distributed as $u_{il} \sim \mathscr{N}(0, \sigma_l^2)$, with $\sigma_1 = 5$, $\sigma_2 = 3$ and $\sigma_3 = 1$. The smooth functions x_i are then 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 we add 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 thus interested in recovering the smooth PC functions $\{v_l\}_{l=1,2,3}$ from these noisy observations over $\mathscr{M}_{\mathscr{T}}$. We compare the proposed SM-FPCA technique to two alternative approaches.

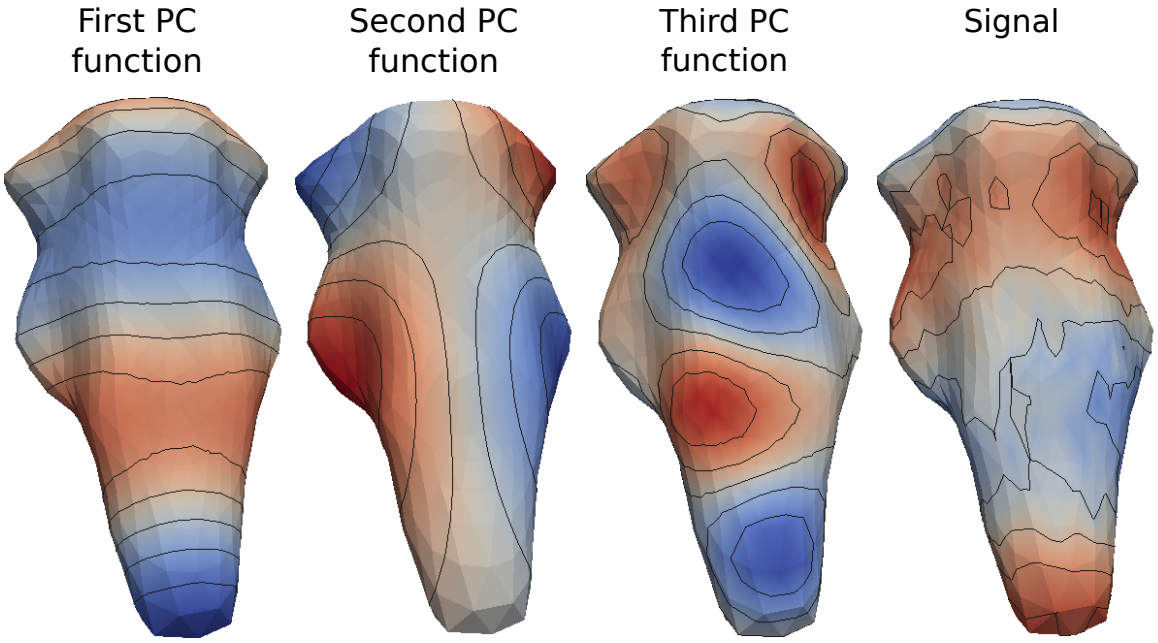


Figure 2.2 From left to right, a plot of the true first, second and third PC functions and a plot of a noisy observation on the brainstem, generated from these three PC functions.

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 $\mathscr{M}_{\mathscr{T}}$. Finally they are normalized to have unitary norm in $L^2(\mathscr{M}_{\mathscr{T}})$.

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, $j = 1,\dots,s$. In this case the smoothing technique applied is IHK smoothing (see Section 2.2 for details). 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. The number of iterations is selected via the F-test criterion outlined in Chung et al. (2005). We 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 Figure 2.3. 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. In

Figure 2.4 we show the plots with the empirical variances explained, where in the case of SM-FPCA, the explained variance has been computed as detailed in the Section 2.3.5.

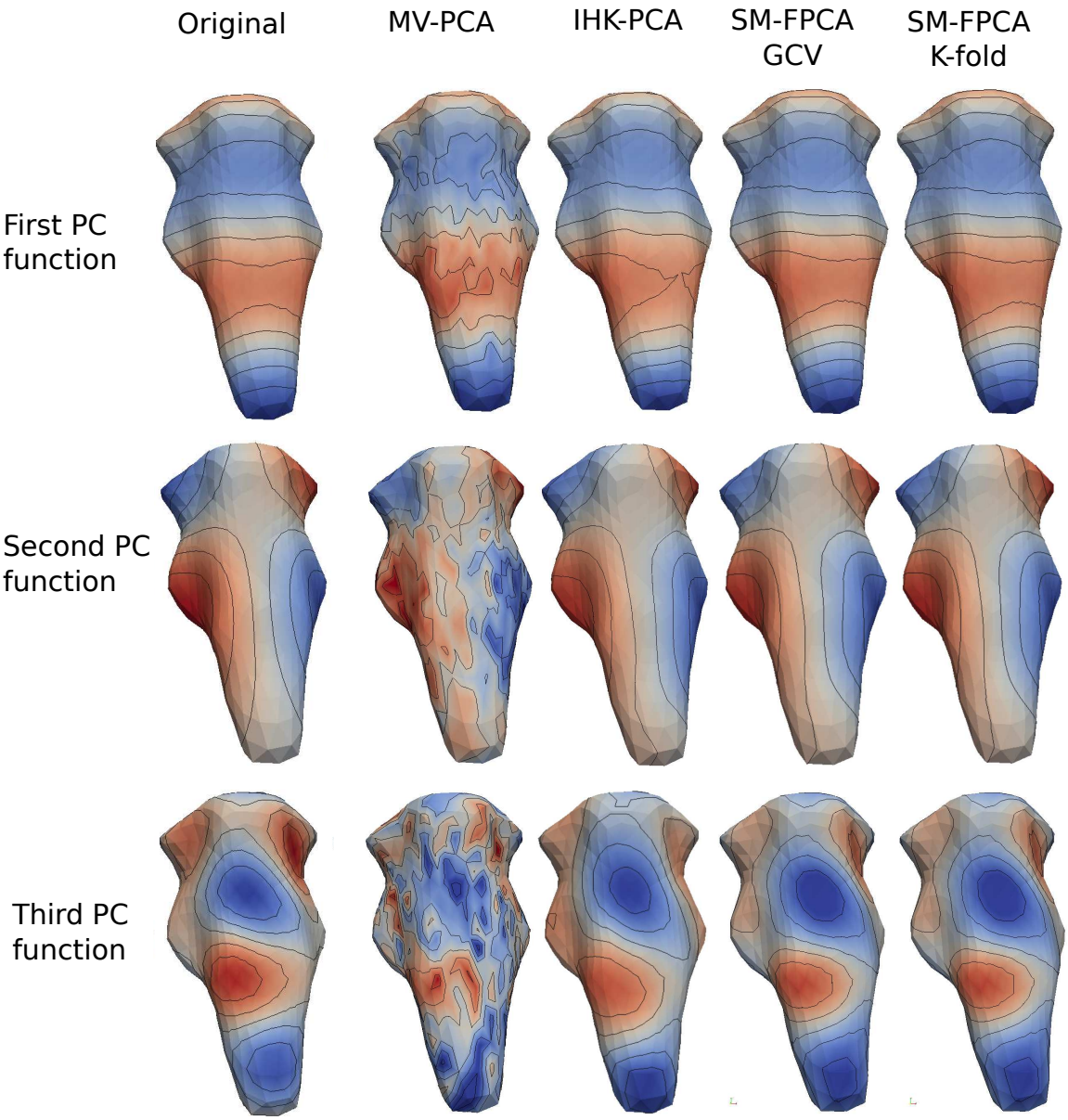


Figure 2.3 From left to right, contours of the original PC functions and their estimates respectively with MV-PCA, IHK-PCA, SM-FPCA GCV and SM-FPCA K-fold. From a visual inspection, MV-PCA shows unsatisfactory results, while a better estimation is achieved by IHK-PCA and SM-FPCA. In particular SM-FPCA is able to better capture details that IHK-PCA ignores. This is apparent for instance in the third PC function reconstruction, in the top-left and top-right corners.

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.

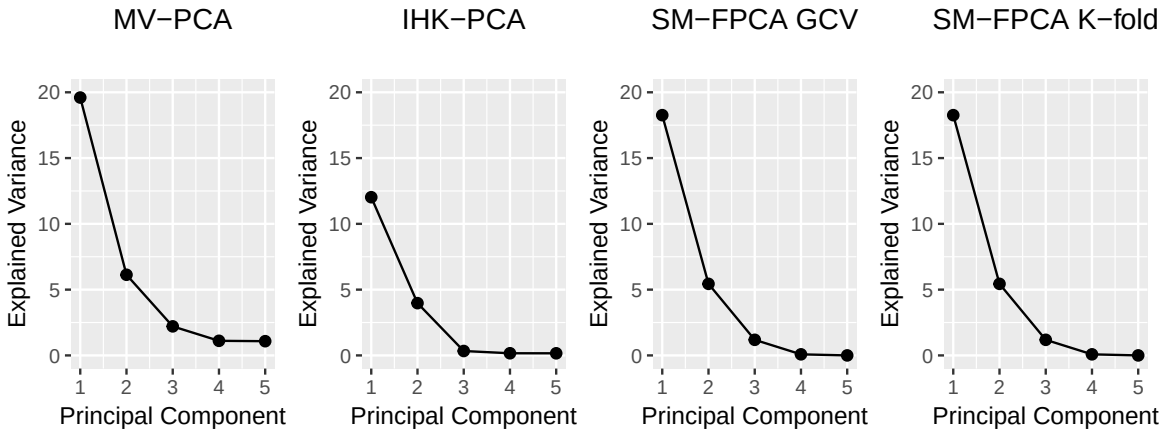


Figure 2.4 From left to right, plot of the empirical variances explained by the first 5 PCs computed with MV-PCA, IHK-PCA, SM-FPCA GCV and SM-FPCA K-fold.

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 the two subspaces are. For this purpose we construct the $s \times 3$ matrices $\mathbb{V} = (v_i(p_j))$ and $\hat{\mathbb{V}} = (\hat{v}_i(p_j))$, where $\hat{v}_i$ is the ith 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 2.5, which compares the MV-PCA, IHK-PCA and SM-FPCA algorithms with respect to the reconstruction's errors of the PC functions $\{v_l\}_{l=1,2,3}$, the PC scores $\{\mathbf{u}_l\}_{l=1,2,3}$ where $\mathbf{u}_l = (u_{il})$, the reconstructed signals $x_i = u_{i1}v_1 + u_{i2}v_2 + u_{i3}v_3$ for $i = 1,\dots,50$ and the principal 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, corresponding scores vectors, signals and subspace reconstruction.

2.5 Application

The data set which we consider in this work arises from the Human Connectome Project Consortium [HCP, Van Essen et al. (2012)], which is collecting data such as structural scans, resting-state and task-based functional MRI scans, and diffusion-weighted MRI

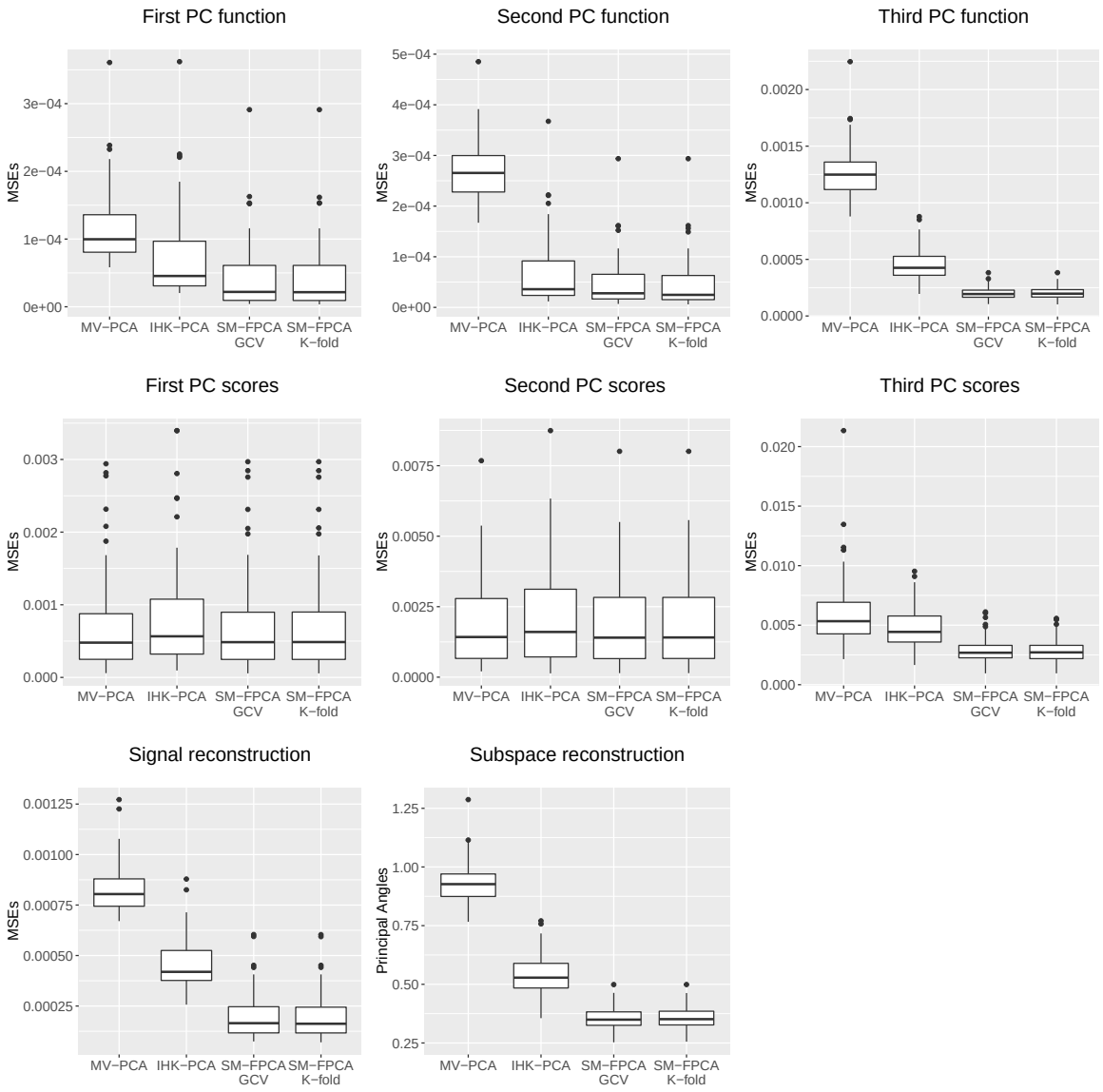


Figure 2.5 Boxplots summarizing the performance of IHK-PCA and SM-FPCA. For the SM-FPCA both GCV and K-fold have been applied for the selection of the smoothing parameter.

scans from a large number of healthy volunteers to help elucidate 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 extracted cortical surfaces 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 2.1 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. This, of course, raises questions about the implications of anatomical alignment, and a small simulation study in Appendix A investigates this issue. 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. Summarizing, the data considered are fMRI filtered time-series on a common triangulated template mesh.

As already mentioned in Section 2.1, 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. This is the area whose behaviour, as compared to the rest of the cortical surface, is of interest for the investigator. Within each subject, a cross-sectional average of all the time-series in the ROI is 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 subject-specific 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. (2016), where a group-average boundary map of the cortical surface is derived from resting state fMRI (Figure 2.6). 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 reasonable ROIs in individual subjects. The parcel that served as ROI in the following analysis is highlighted in red in Figure 2.6. For the chosen ROI, a snapshot of the RSFC map of one subject is shown in Figure 2.7.

The mean RSFC map is shown in Figure 2.8. 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.

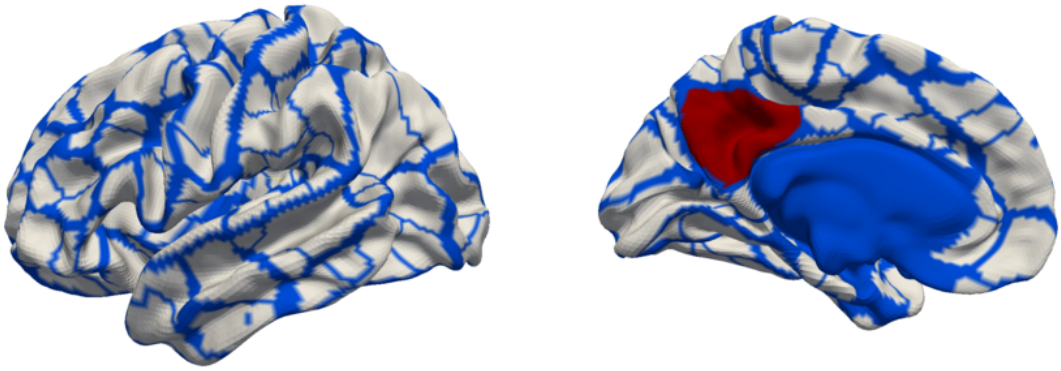


Figure 2.6 Parcellation of the cortical surface derived in Gordon et al. (2016). In red the Region of Interest chosen for the computation of the RSFC maps. This region is localized on an area of the cerebral cortex called precuneus. The blue colours indicate the parcellated regions, with the major blue area being the join between the two brain hemispheres, which does not lie on the manifold surface and which is therefore excluded from the cortical surface analysis.

We wish now to understand which are the main modes of variation of these RSFC maps among the different subjects, by applying a PCA.

The first three PC functions, estimated with SM-FPCA, are shown in Figures 2.9-2.10-2.11 as compared to the PC functions derived from MV-PCA and IHK-PCA. The choice of 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, and the spatial information is completely ignored by this model. In fact, even recent attempts to model the subject variability from resting state fMRI leads to the conclusion that spatial mismatches, introduce 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 to 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 [Marron et al. (2015)]. 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

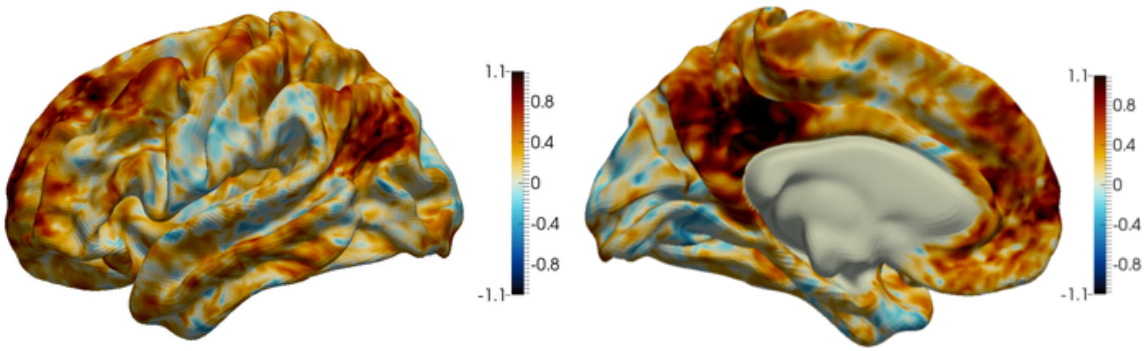


Figure 2.7 A snapshot of the RSFC map of one subject.

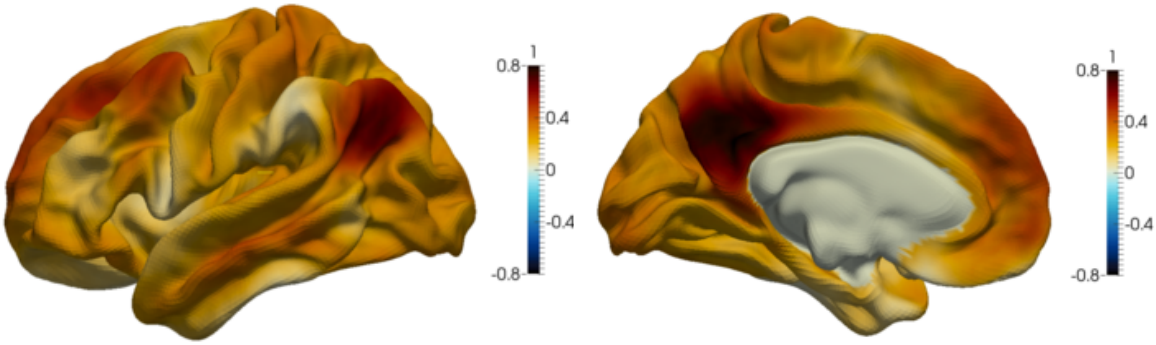


Figure 2.8 The mean RSFC map computed over 491 subject. As expected, high correlation values are visible inside the ROI.

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.

Both IHK-PCA and SM-FPCA returns smooth PC functions. A visual inspection of the estimated PC functions though highlights that IHK-PCA completely smooths out sharper changes in the modes of variations, missing some localized features that are apparent in MV-PCA and are also very well captured by the proposed SM-FPCA. Comparing for instance the estimated third PC functions, in the top views of Figure 2.11, one can see for both MV-PCA and SM-PCA corresponding localized areas with very high values (in red) and very low values (in blue) that are instead missing in the IHK-PCA estimate. By

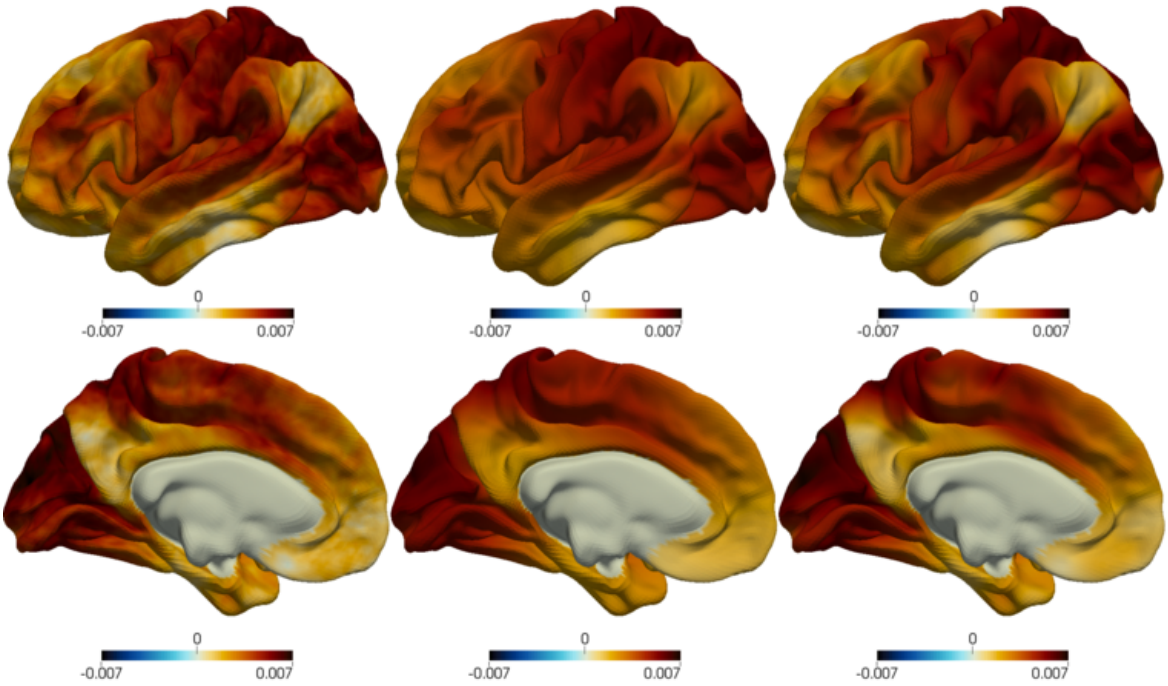


Figure 2.9 From left to right, two views of the first PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.

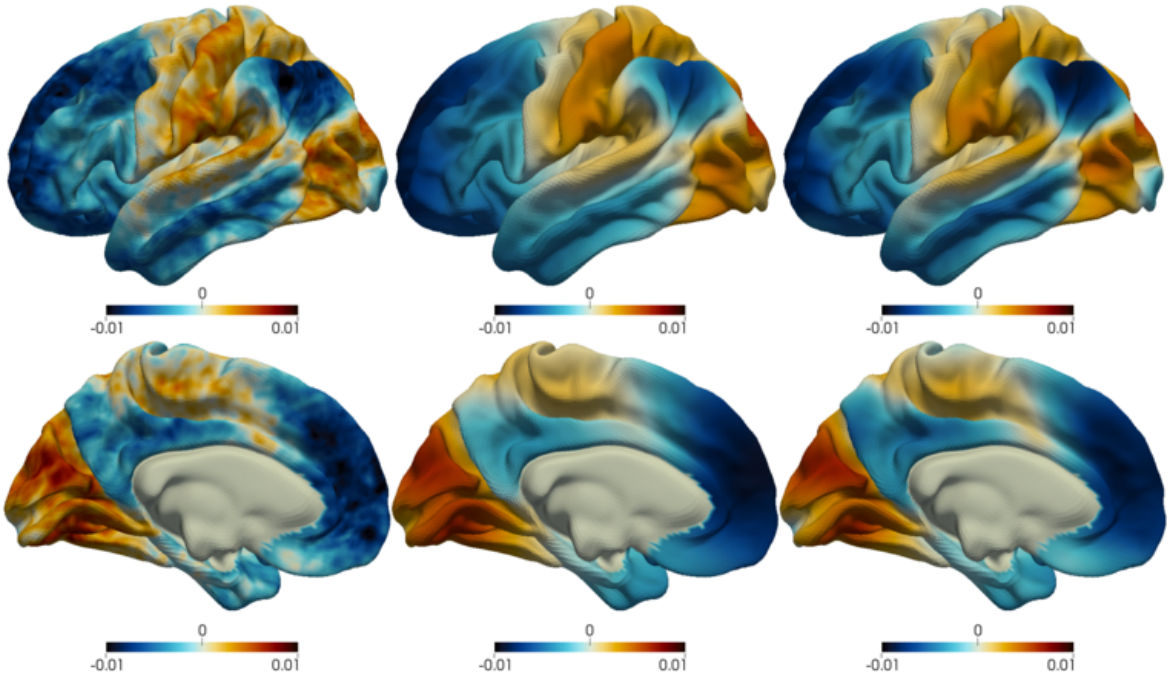


Figure 2.10 From left to right, two views of the second PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.

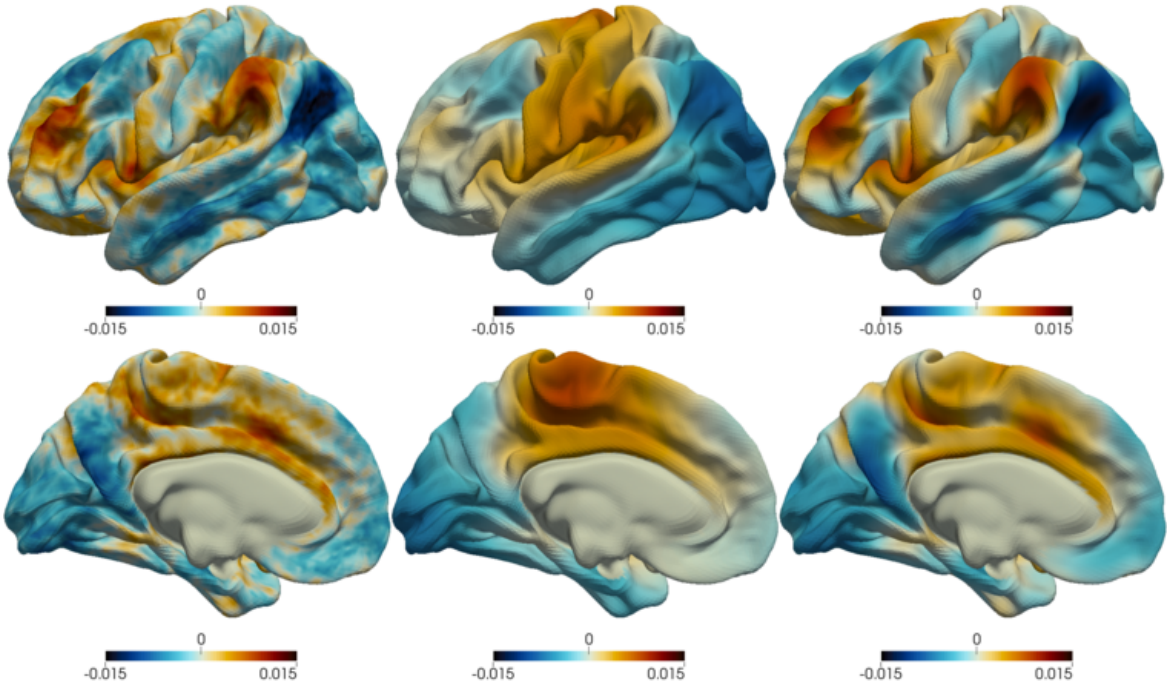


Figure 2.11 From left to right, two views of the third PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.

contrary, the pre-smoothing approach appears to introduce some artifacts: looking at the bottom views in Figure 2.11, one can for instance notice that IHK-PCA estimated third PC function has high values in the higher part of the plot, that do not have match neither on the MV-PCA nor on the SM-FPCA estimate.

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 2.12 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]. However, it should be noted that variability can be both somewhat localised as well as more spatially smooth, indicating that even in resting state data, brain regions have differential response which is not simply a result of noise in the data.

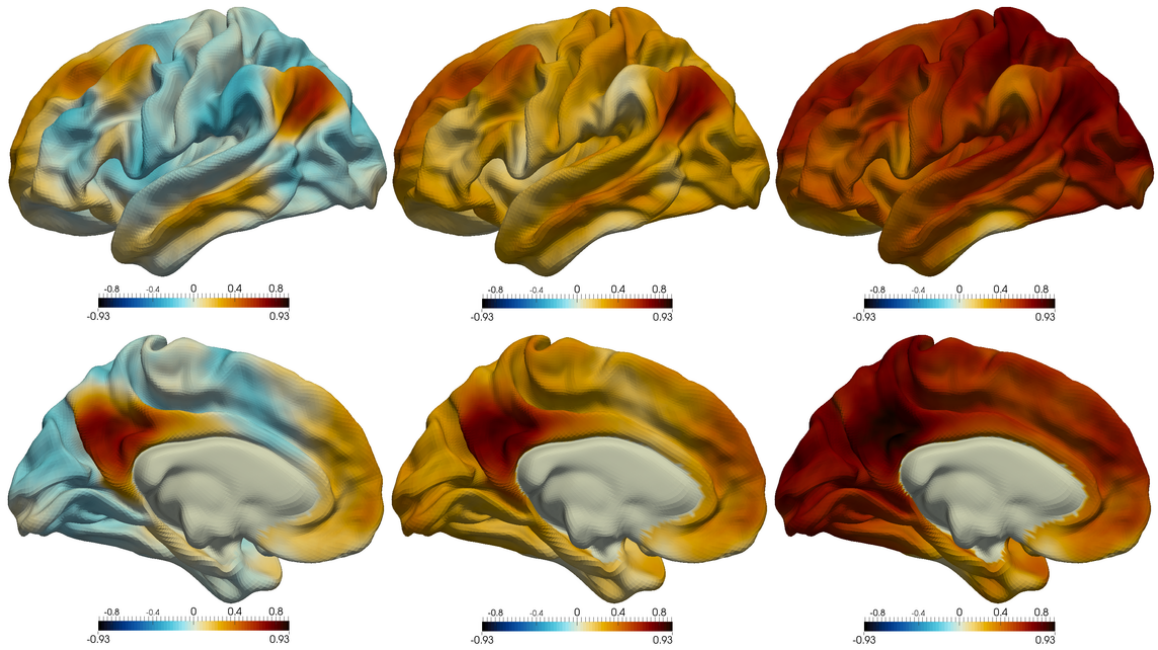


Figure 2.12 From left to right, two views of $\mu - 2\sigma f$, μ, $\mu + 2\sigma f$, where μ denotes the mean RSFC map, σ denotes the standard deviation of the first PC scores vector and f denotes the first PC function.

2.6 Discussion

In this chapter we introduced 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 considered 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 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 functions. It could also be imagined that in more complex study designs (such as patient versus control studies) these PC functions, along with the associated scores, could be used to investigate diverse difference between groups or covariate effects.

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 dimensions, related to the number of nodes of the triangular mesh, and the number of point-wise observations available for each functional sample as well as 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×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.

Chapter 3

Functional Principal Component Analysis in the inverse problem setting

3.1 Motivation

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, 2004), medical imaging (Arridge, 1999; Lustig et al., 2008) and machine learning (De Vito et al., 2005).

We recall that solving a linear inverse problem means finding an unknown x, for instance a function or a surface, from a noisy indirect observation y, which is a solution to the model

$$y = \mathscr{K} x + \varepsilon, \tag{3.1}$$

where y and ε are elements of a Hilbert space. The map $\mathscr{K}$ (known) is called the forward operator. The term ε models observational error. In Chapter 1, Section 1.1.4, we give a brief review of the possible approaches to the reconstruction of x from y.

In this chapter, we look at the inverse problem from a Functional Data Analysis (FDA) perspective. 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

$$y_i = \mathscr{K}_i x_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{3.2}$$

with $\mathscr{K}_1, \ldots, \mathscr{K}_n$ a collection of sample specific known forward operators.

Problem 3.2 has been classically dealt with a Reconstruct-then-Estimate approach, i.e. by reconstructing each observation independently. In other words, the underlying statistical model of the data is ignored, and such a problem is formulated as n separate Problem 3.1s. 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.

In imaging sciences, it is sometimes of interest to find an optimal representation and perform statistics on the second order information associated to the functional samples, i.e. their covariance operators. This is, for instance, the case in a number of areas of neuroimaging, particularly those investigating functional connectivity. Therefore, in this chapter, we also establish a framework for reconstructing and optimally representing indirectly observed samples $\mathscr{C}_1, \ldots, \mathscr{C}_n$, that are covariance operators, expressing the second order properties of the underlying unobserved functions. The indirect observations are covariance operators generated by the model

$$\mathscr{S}_i = \mathscr{K}_i \circ \mathscr{C}_i \circ \mathscr{K}_i^* + \mathscr{E}_i, \qquad i = 1, \ldots, n, \tag{3.3}$$

where $\mathscr{K}_i^*$ denotes the adjoint operator and the term $\mathscr{E}_i$ models observational error. The term $\mathscr{K}_i \circ \mathscr{C}_i \circ \mathscr{K}_i^*$ represents the covariance operator of $\mathscr{K}_i X^{(i)}$, with $X^{(i)}$ an underlying random function whose covariance operator is $\mathscr{C}_i$.

As opposed to more classical linear inverse problems formulations, Problem 3.3 introduces the following additional difficulties:

- We are in a setting where each sample is an object that is a covariance operator and it is important to take advantage of the information from all the samples in the reconstruction and representation of each of them.

- The elements $\{\mathscr{C}_i\}$ and $\{\mathscr{S}_i\}$ live on non-Euclidean spaces, as they belong to the positive semidefinite cone and it is important to account for this constraint in the formulation of the associated estimators.

- In an inverse problem setting it is fundamental to be able to introduce spatial regularization, however it is not obvious how to feasibly construct a regularizing term for covariance operators reflecting, for instance, smoothness assumptions on the underlying functional images.

We tackle Problem 3.3 by generalizing the concept of functional Principal Component Analysis (fPCA) to indirectly observed covariance operators. We do this in a Regularized-Estimate fashion to mitigate the error terms amplification due to the ill-posedness of the inverse problem. The aim is to optimally represent and understand the variation associated with samples that are indirectly observed covariance operators.

3.1.1 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.

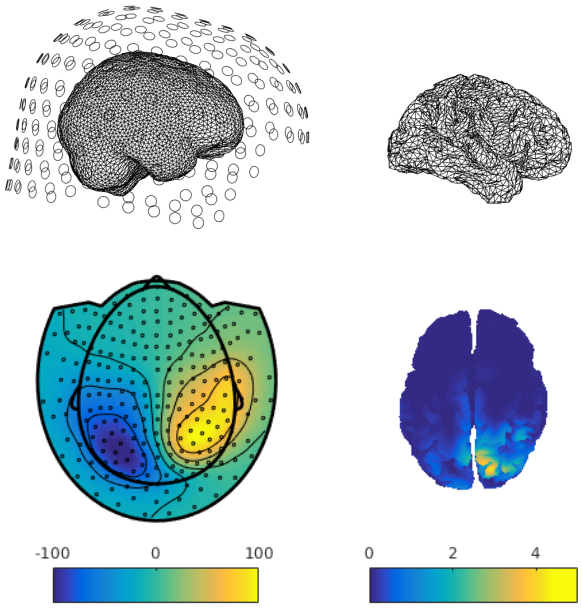


Figure 3.1 On the top left, head model of a subject and superimposition of the 248 MEG sensors positioned around the head, called 'sensors space'. On the top right, brain model of the same subject represented by a triangular mesh of $8K$ nodes, which represents the 'brain space'. On the bottom left, 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.

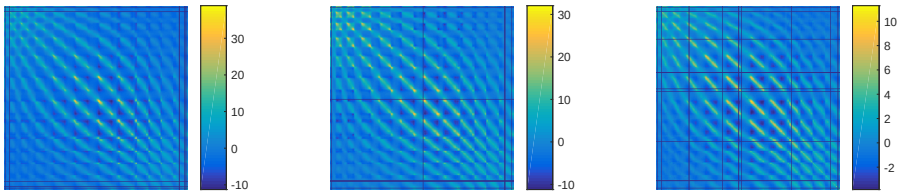


Figure 3.2 Covariance matrices of the signal detected by the MEG sensors from three different subjects of the Human Connection Project. The size of the matrices is 248×248. The dark blue bands represent missing data, which are due to the exclusion of some channels after a quality check of the signal.

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. Some of these imaging modalities produce indirect measurements, where, as shown in Figure 3.1 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 images 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 chapter for convenience, however it's 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 soft-field methods like EEG, MEG and Functional Near-Infrared Spectroscopy, 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 3.1, we show an example of a signal on the brain space and the associated signal on the sensors space.

From a practical prospective, 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. In Figure 3.2, we show the covariance matrices, on the sensor space, computed from the MEG signals of three different subjects.

The remainder of this chapter is organized as follows. In Section 3.2 we give a formal description of the problem. We then introduce a model for indirectly observed smooth functional images in Section 3.3 and present the more general model, for observations that are indirectly observed covariances, in Section 3.4. In Section 3.5, we perform simulations to access the validity of the estimation framework. In Section 3.6 we apply the proposed models to the MEG data and we give some concluding remarks in Section 3.7. The proofs are presented in Section 3.8.

3.2 Mathematical description of the problem

To set the notation of this chapter, we now introduce the problem using our driving application as an example. To this purpose, let $\mathscr{M}$ 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 3.1. We denote with $L^2(\mathscr{M})$ the space of square integrable functions on $\mathscr{M}$. Define X to be a random function with values in a Hilbert functional space $\mathscr{F} \subset L^2(\mathscr{M})$ 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 $\mathscr{C}_X$ is defined as $\mathscr{C}_X g = \int_{\mathscr{M}} C_X(v, v') g(v) dv$, for all $g \in L^2(\mathscr{M})$. Recall from Chapter 1 that Mercer's Lemma guarantees the existence of a non-increasing sequence $\{\kappa_r\}$ of eigenvalues of $\mathscr{C}_X$

and an orthonormal sequence of corresponding PC functions $\{\psi_r\}$, such that

$$C_X(v,v') = \sum_{r=1}^{\infty} \kappa_r \psi_r(v)\psi_r(v'), \qquad \forall\, v,v' \in \mathcal{M}.$$

The random function X can be expanded as $X = \mu + \sum_{r=1}^{\infty} \zeta_r \psi_r$, where the PC scores $\{\zeta_r\}$ are given by $\zeta_r = \int_{\mathcal{M}} \{X(v) - \mu(v)\}\psi_r(v)\,dv$. Recall also that, for any fixed $M \in \mathbb{N}$, the first M PC functions of X satisfy

$$(\psi_m)_{m=1}^{M} = \underset{(\{\phi_m\}_{m=1}^{M} : \langle \phi_m, \phi_l \rangle = \delta_{ml})}{\arg\min} \mathbb{E} \int_{\mathcal{M}} \left\{ X - \mu - \sum_{m=1}^{M} \langle X - \mu, \phi_m \rangle \phi_m \right\}^2, \tag{3.4}$$

where δ_{ml} is the Kronecker delta.

3.2.1 The case of indirectly observed functions

In the case of indirect observations, the signal is detectable only through s sensors on the sensors space. Let $\{K_i\}$ be a collection of $s \times p$ real matrices, representing the subject specific forward operators relating the signal at p pre-defined points $\{v_j : j = 1,\dots,p\}$ on the cortical surface $\mathcal{M}$ with the signal captured by the s sensors. Moreover, define the evaluation operator $\Psi : \mathcal{F} \to \mathbb{R}^p$ 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 $(f(v_1),\dots,f(v_p))^T$. The operators Ψ and $\{K_i\}$ are known. However, in the described problem the random function X can be observed only through indirect measurements $\{\mathbf{y}_i \in \mathbb{R}^s : i = 1,\dots,n\}$ generated from the model

$$\begin{cases} x_i &= \mu + \sum_{r=1}^{\infty} \zeta_{i,r} \psi_r \\ \mathbf{y}_i &= K_i \Psi x_i + \boldsymbol{\varepsilon}_i, \qquad i = 1,\dots,n \end{cases} \tag{3.5}$$

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 $\{\boldsymbol{\varepsilon}_i\}$ 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 (3.5) is an implementation of the idealized Problem 3.2. In Figure 3.3 we give an illustration of the introduced setting.

Note that it would not be necessary to define an evaluation operator if the forward operators were defined to be functionals $\{\mathcal{K}_i : \mathcal{F} \to \mathbb{R}^s\}$, 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 3.6 for details) for a pre-specified set of points $\{v_j\} \subset \mathscr{M}$ and it is thus convenient to take this into account in the model, through the introduction of an evaluation operator Ψ.

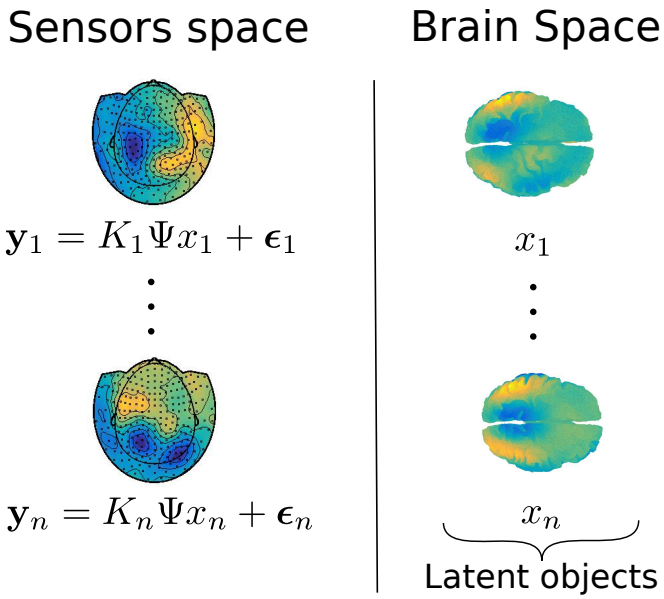


Figure 3.3 Illustration of the setting introduced with Model 3.5.

In the case of single subject studies, the surface $\mathscr{M}$ is the subject's reconstructed cortical surface, an example of which is shown on the right panel of Figure 3.1. In this case, it is natural to assume that there is one common forward operator K for all the detected signals. In the more general case of multi-subject studies, $\mathscr{M}$ is assumed to be a template cortical surface. We are thus assuming that the individual cortical surfaces have been registered to the template $\mathscr{M}$, which means that there is a smooth and one-to-one correspondence between the points on each individual brain surface and the template surface $\mathscr{M}$, where the PC functions are defined.

As mentioned in Section 1.1.4, a Reconstruct-then-Estimate approach is suboptimal in this setting. For this reason, different alternatives have been proposed in the literature. In the simplified setting of a fixed forward operator $K := K_1 = \ldots = K_n$, Amini and Wainwright (2012) propose to estimate the space spanned by the first R PC functions in a Regularized-Estimate fashion. The PC functions $\{\psi_r\}$ are modelled as elements of a Reproducing Kernel Hilbert Space (RKHS). On the sensors space, they define a smoothing matrix S^{-1} translating the smoothness assumption on the PC functions to the sensors space. The

PC loadings on the sensors space $\{\mathbf{z}_r\} \subset \mathbb{R}^s$ are in practice estimated as eigenvectors of a regularized version of the empirical covariance $\hat{C}_Y = \frac{1}{n}\sum_i \mathbf{y}_i \mathbf{y}_i^T$, namely

$$\hat{C}_Y - \lambda S^{-1},$$

with λ a weighting coefficient. The PC functions of X are estimated as the minimum norm functions $\{\psi_r\}$ satisfying

$$(K\Psi)(\hat{\psi}_r) = \hat{\mathbf{z}}_\mathbf{r}.$$

The estimated PC functions $\hat{\psi}_r$ are such that their images on the sensors space $(K\Psi)(\hat{\psi}_r)$ interpolate $\hat{\mathbf{z}}_\mathbf{r}$ and have minimum norm in the RKHS where the PC functions are defined.

An extension to the case of subject specific forward operators has been proposed in Katsevich et al. (2015). The authors formulate a least square estimator for the discretized covariance function $(C_X(v_j, v_l))_{j,l}$, which in the notation of this chapter takes the form

$$\hat{C}_X = \underset{C \in \mathbb{R}^{p\times p}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \|(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T - K_i C K_i^T - \sigma^2 I_s\|_F^2,$$

with $\bar{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^n \mathbf{y}_i$. The term $\sigma^2 I_s$, with $\sigma > 0$ a constant and I_s the identity matrix of size s, captures a diagonal structure, due for instance, to observational error. Intuitively, the 'covariance' C_X (note that C_X is not constrained to be positive semi-definite) is such that its projections on the sensors space $K_i C_X K_i^T$ match as closely as possible the covariances in the sensors space $(\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T$.

In Dobriban et al. (2017), in the context of optimal prediction, the point-wise evaluations of the PC functions $\{\psi_r\}$, on $v_1 \ldots, v_p$, are estimated from the eigenvectors of the empirical covariance of the backprojected data, i.e.

$$\frac{1}{n}\sum_{i=1}^{n} K_i^T \mathbf{y}_i \mathbf{y}_i^T K_i.$$

The approach in Amini and Wainwright (2012) cannot be immediately extended to the case of subject-specific forward operators and, to make use of the RKHS machinery, would require the definition of a kernel on the non-linear domain $\mathscr{M}$, which is not a trivial task. The approaches in Dobriban et al. (2017); Katsevich et al. (2015) both lack of a regularization step, fundamental in the inverse problem setting considered here. Also, in Katsevich et al. (2015) the estimated eigenvectors could potentially be associated to negative eigenvalues, as the estimated covariance is not constrained to be positive semi-definite. This motivates a novel estimator proposed in Section 3.3, where we formulate an

extension of the Regularized-Estimate fPCA model introduced in Chapter 2 to the inverse problem setting described with model (3.5).

3.2.2 The case of indirectly observed covariance operators

A natural generalization of the setting in the previous section is considering observations that have group specific covariance operators. In detail, suppose now we are given a set of n covariance functions $\{C_i : i = 1,\dots,n\}$, representing the associated covariance operators $\{\mathscr{C}_i : i = 1,\dots,n\}$ on the brain space. In our driving application, each covariance function $C_i : \mathscr{M} \times \mathscr{M} \to \mathbb{R}$ describes the functional connectivity of the ith individual or the functional connectivity of the same individual at the ith time-point.

Here we consider the problem of defining and estimating a set of modes of variation from $\{C_i\}$, 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 modes of variation of $\{C_i\}$ from the set of indirectly observed covariance matrices, computed from the signal on the sensors space, and thus given by $\{S_i \in \mathbb{R}^{s\times s}, i = 1,\dots,n\}$ with

$$S_i = K_i \mathbb{C}_i K_i^T + E_i^T E_i, \qquad i = 1,\dots,n, \tag{3.6}$$

where $\mathbb{C}_i = (C_i(v_j, v_l))_{jl}$ and $\{v_j : j = 1,\dots,p\}$ are the sampling points associated to the operator Ψ. 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 σ. Model (3.6) could be regarded as an implementation of the idealized Problem 3.3, where the covariance operators are represented by the associated covariance functions. An illustration of the setting introduced can be found in Figure 3.4.

The case of statistical samples that are covariance functions has not been extensively covered in the literature. In fact, PCs are generally 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, on the space of covariance functions, a linear subspace, or part of it, is likely to fall outside the non-Euclidean cone of positive semidefinite operators. In the statistical literature,

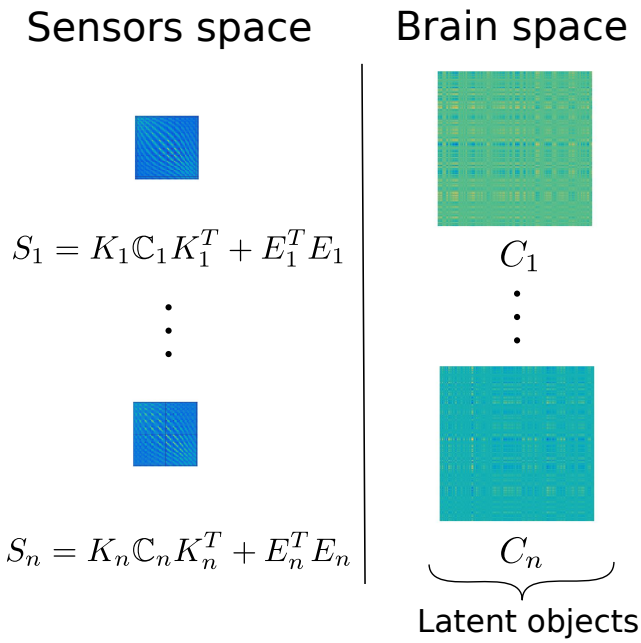


Figure 3.4 Illustration of the setting introduced with Model 3.6.

this non-Euclidean structure is accounted for by introducing a distance in the space of covariance matrices or covariance operators (see e.g. Dryden et al., 2009; Pigoli et al., 2014).

In the discrete setting, the analysis of responses that are directly observable positive-definite matrices has been considered, among others, in Zhu et al. (2009). Suppose we observe $(C_i, \mathbf{x}_i) \in \mathbb{R}^{s\times s} \times \mathbb{R}^q$, for $i = 1, \dots, n$, with C_i symmetric positive-definite matrices, and $\mathbf{x}_i$ regressors associated to the ith observation. In Zhu et al. (2009) a semi-parametric model is considered, where the parametrization is formulated on unconstrained representations of $\{C_i\}$ (e.g. matrix logarithms). In Yuan et al. (2012), such a class of models is generalized to a fully non-parametric setting, by defining local polynomial regression between the regressors x_i and the unconstrained representations of C_i. An extension of this setting to responses in general Riemannian symmetric spaces has been proposed in Cornea et al. (2017). In Lin et al. (2017), an extrinsic approach is instead adopted, where a kernel estimate of the relation between C_i and $\mathbf{x}_i$ is defined ignoring the symmetric positive-definiteness constraints on $\{C_i\}$, and subsequently the unconstrained estimate is projected onto the space of valid estimates.

In the discrete setting, Dryden et al. (2009) introduce a PC model for directly observed covariance matrices. An extension to directly observed covariance operators has been pro-

posed in Pigoli et al. (2014). Such models cannot deal with indirectly observed covariances, we thus propose a novel approach for this problem in Section 3.4.

3.3 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 $\{\mathbf{y}_i\}$, defined in (3.5). The main difficulty in defining an estimation model for C_X, directly from the samples $\{\mathbf{y}_i\}$, is the necessity to include an equivalent regularizing term for the function C_X, to improve the finite sample properties of the estimator. The model proposed here, tackles this problem by working on the estimation of the elements of the spectral decomposition of $\mathscr{C}_X$, or in other terms on the PC functions.

3.3.1 Model

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

$$(\hat{\mathbf{z}}, \hat{f}) = \underset{\mathbf{z}\in\mathbb{R}^n, f\in H^2(\mathscr{M})}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - z_i K_i \Psi f\|^2 + \lambda \mathbf{z}^T \mathbf{z} \int_{\mathscr{M}} \Delta^2_{\mathscr{M}} f, \tag{3.7}$$

where the Laplace-Beltrami operator $\Delta_{\mathscr{M}}$, integrated over the manifold $\mathscr{M}$, 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 $\mathbf{y}_i$. The parameter λ 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 $\{\mathbf{y}_i - \hat{z}_i K_i \Psi \hat{f}\}$, and so on for the subsequent PCs. The proposed model can be interpreted as a regularized least square estimation of the first PC function ψ_1 in (3.5), with the terms $\{z_i\}$ playing the role of estimates of the variables $\{\zeta_{i,1}\}$.

In the simplified case of a single forward operator $K = K_1 = \ldots = K_n$, the minimization problem (3.7) can be reformulated in a more classical form. In fact, fixing f in (3.7) and minimizing $\mathbf{z}$ gives

$$z_i = \frac{\mathbf{y}_i^T K \Psi f}{\|K\Psi f\|^2 + \lambda \int_{\mathscr{M}} \Delta^2_{\mathscr{M}} f}, \qquad i = 1, \ldots, n, \tag{3.8}$$

which can then be used to show that the minimization problem (3.7) is equivalent to the generalized eigenproblem

$$\frac{(\Psi f)^T K^T \mathbb{Y}^T \mathbb{Y} K(\Psi f)}{\|K\Psi f\|^2 + \lambda \int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f}, \tag{3.9}$$

with $\mathbb{Y}$ a $n \times s$ real matrix, where the ith row of $\mathbb{Y}$ is the observation $\mathbf{y}_i^T$. This reformulation gives further insights on the interpretation of $\hat{f}$ in (3.7). In fact, $\hat{f}$ is such that $K\Psi\hat{f}$ maximizes $\frac{1}{n}\mathbb{Y}^T\mathbb{Y}$, i.e. the point-wise estimate of the covariance matrix in the sensors space. The term $\mathbf{z}^T\mathbf{z}$ in (3.7), places the regularization term $\lambda \int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f$ in the denominator of the equivalent formulation (3.9). Thus, $\hat{f}$ is regularized by the choice of norm in the denominator of (3.9), in a similar fashion to the classic functional principal component formulation of Silverman (1996) in equation (1.14).

3.3.2 Algorithm

Here we propose a minimization approach for the objective function in (3.7). Similarly to what we did in Section 2.3.2, we approach the problem by alternating the minimization of $\mathbf{z}$ and f in an iterative algorithm. In (3.7), a normalization constraint must be considered to make the representation unique, as in fact multiplying $\mathbf{z}$ by a constant and dividing f by the same constant does not change the objective function. We optimize in $\mathbf{z}$ under the constraint $\|\mathbf{z}\|_2 = 1$, which leads to a normalized version of the estimator (3.8)

$$z_i = \frac{\mathbf{y}_i^T K_i \Psi f}{\sqrt{\sum_{i=1}^n \mathbf{y}_i^T K_i \Psi f}}, \qquad i = 1, \dots, n. \tag{3.10}$$

For a given $\mathbf{z}$, solving (3.7) with respect to f will turn out to be equivalent to solving an inverse problem, which we discretize adopting a Finite Elements approach, similar to the one in Chapter 2, that we recall here. Note also that for $K_1, \dots, K_n = I_p$, the identity matrix of size p, that is the case of discrete and noisy measurements of the samples, we obtain the same estimators as in Chapter 2.

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 of a set of globally continuous functions over $\mathcal{M}_{\mathcal{T}}$ that are affine where restricted to any triangle τ in $\mathcal{T}$, i.e.

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

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 $\mathscr{M}_\mathscr{T}$. Such basis functions are Lagrangian, meaning that $\phi_i(\xi_j)=1$ if $i=j$ and $\phi_i(\xi_j)=0$ otherwise. Setting $\mathbf{c}=(f(\xi_1),\ldots,f(\xi_\kappa))^T$ and $\boldsymbol{\phi}=(\phi_1,\ldots,\phi_\kappa)^T$, every function $f\in V$ has the form

$$f(v)=\sum_{k=1}^{\kappa} f(\xi_k)\phi_k(v)=\mathbf{c}^T\boldsymbol{\phi}(v) \tag{3.11}$$

for all $v\in\mathscr{M}_\mathscr{T}$. To ease the notation, we assume that the p points $\{v_j\}$ associated with the evaluator function Ψ coincide with the nodes of the triangular mesh $\xi_1,\ldots,\xi_\kappa$, and thus we have that the coefficients $\mathbf{c}$ are such that $\mathbf{c}=\Psi f$ for any $f\in V$. Consequently, we are assuming the forward operators $\{K_i\}$ to be $s\times\kappa$ matrices, relating the κ points on the ith 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 ith subject.

Let now M and A be the mass and stiffness $\kappa\times\kappa$ matrices defined as $(M)_{jl}=\int_{\mathscr{M}_\mathscr{T}}\phi_j\phi_l$ and $(A)_{jl}=\int_{\mathscr{M}_\mathscr{T}}\nabla_{\mathscr{M}_\mathscr{T}}\phi_j\cdot\nabla_{\mathscr{M}_\mathscr{T}}\phi_l$, where $\nabla_\mathscr{M}$ is the gradient operator on the manifold $\mathscr{M}$. Than, the solution of (3.7), in the discrete space V, is given by the following proposition.

Proposition 3. *The Surface Finite Element solution $\hat{f}_h\in V$ of model (3.7), for a given unitary norm vector $\mathbf{z}$, is $\hat{f}_h=\hat{\mathbf{c}}^T\boldsymbol{\phi}$ where $\hat{\mathbf{c}}$ is the solution of*

$$\hat{\mathbf{c}}=(\sum_{i=1}^{n} z_i^2K_i^TK_i+\lambda AM^{-1}A)^{-1}\sum_{i=1}^{n} z_iK_i^T\mathbf{y}_i. \tag{3.12}$$

Equation (3.12) has the form of a penalized regression, where the penalty operator resulting from the discretization procedure is $AM^{-1}A$.

Algorithm 2 Inverse fPCA Algorithm

1: Initialization:

(a) Computation of M and A

(b) Initialize $\mathbf{z}$, the scores vector associated to the first PC function

2: PC function's estimation:

Compute $\mathbf{c}$ such that

$$\Big(\sum_{i=1}^{n} z_i^2 K_i^T K_i + \lambda A M^{-1} A\Big)\mathbf{c} = \sum_{i=1}^{n} z_i K_i^T \mathbf{y}_i$$

$$f_h \leftarrow \mathbf{c}^T \boldsymbol{\phi}$$

3: Scores estimation:

$$z_i \leftarrow \frac{\mathbf{y}_i^T K_i \Psi f_h}{\sqrt{\sum_{i=1}^{n} \mathbf{y}_i^T K_i \Psi f_h}}, \qquad i = 1, \ldots, n$$

4: Repeat Steps 2–3 until convergence

The sparsity of the linear system (3.12), 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 $\tilde{M}^{-1}$, where $\tilde{M}$ is the diagonal matrix such that $\tilde{M}_{jj} = \sum_l M_{jl}$. The penalty operator $A\tilde{M}^{-1}A$ 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 (3.12) can be rewritten as the sparse overdetermined system

$$\begin{bmatrix} K \\ \sqrt{\lambda}\tilde{M}^{-1/2}A \end{bmatrix} \mathbf{c} = \begin{bmatrix} Y^T \mathbf{z} \\ \mathbf{0} \end{bmatrix}, \tag{3.13}$$

to be interpreted in a least-square sense. A sparse QR solver can be finally applied to efficiently solve the linear system (3.13).

In Algorithm 2 we summarize the main algorithmic steps to compute the PC functions and associated PC scores for indirectly observed functions.

3.3.3 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 2, and moving the normalization step from z_i to f_h, we obtain the iterations

$$(K^T K + \lambda A M^{-1} A)\mathbf{c} = K^T \sum_{i=1}^{n} (\mathbf{y}_i \mathbf{y}_i^T) K \Psi f_h$$
$$f_h \leftarrow \mathbf{c}^T \boldsymbol{\phi}; \; f_h \rightarrow \frac{f_h}{\|f_h\|}.$$

The obtained algorithm depends on the data only through $\sum_{i=1}^{n} (\mathbf{y}_i \mathbf{y}_i^T)$ 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 $\{\mathbf{y}_i\}$ are not available, but we are given only the associated $s \times s$ covariance matrix S on the sensors space, computed from $\{\mathbf{y}_i\}$.

3.4 Principal components of indirectly observed covariance functions

In this section we extend the proposed methodology to samples that are indirectly observed covariance operators. Specifically, we formulate models that allow us to reconstruct and optimally represent a set of latent covariance functions.

3.4.1 Representation models for covariance operators

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 3.2. Recall moreover that we denote with $\mathscr{M}$ the brain surface template and with $\{K_i \in \mathbb{R}^{s \times p}\}$ the set of subject specific forward operators, relating the signal at the p pre-specified points $\{v_j\}$ on the cortical surface $\mathscr{M}$ with the signal detected on the s sensors.

The aim of this section is to introduce models for the reconstruction and representation of 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 covari-

ance functions $\{C_i\}$ through formula (3.6) that we recall here being

$$S_i = K_i \mathbb{C}_i K_i^T + E_i^T E_i, \qquad i = 1, \dots, n,$$

with $\mathbb{C}_i = (C_i(v_j, v_l))_{jl}$ and $\{v_j\}$ the sampling points associated to the operator Ψ.

A subject-specific model

Let $S_i^{1/2}$ be a square-root decomposition of S_i, which is a decomposition such that $S_i = (S_i^{1/2})^T S_i^{1/2}$, for all $i = 1, \dots, n$. This could be given, for instance, by $S_i^{1/2} = V_i D_i^{1/2}$ where $S_i = V_i D_i V_i^T$ is the spectral decomposition of S_i and $D_i^{1/2}$ denotes the diagonal matrix whose entries are the square root of the (non-negative) entries of D_i. Each square-root decomposition $S_i^{1/2}$ can be interpreted as a data-matrix whose empirical covariance is S_i.

In the most general setting, each covariance matrix S_i is an indirect observation of an underlying covariance function C_i, which can be expressed in terms of its spectral decomposition as

$$C_i(v, v') = \sum_{r=1}^{\infty} \kappa_{ir} \psi_{ir}(v) \psi_{ir}(v'), \qquad \forall v, v' \in \mathcal{M},$$

where, for each $i = 1, \dots, n$, we have a sequence of non-increasing variances $\kappa_{i1} \geq \kappa_{i2} \geq \cdots \geq 0$ and $\{\psi_{ir}\}_r$ the set of orthonormal eigenfunctions of the associated covariance operator $\mathscr{C}_i$.

Introduce now $\{\hat{f}_i \in H^2(\mathcal{M})\}$ and $\{\hat{\mathbf{z}}_i \in \mathbb{R}^s\}$, obtained by applying Model (3.7) to each sample independently, i.e.

$$\left\{(\hat{\mathbf{z}}_i, \hat{f}_i)\right\}_i = \underset{\{\mathbf{z}_i\} \subset \mathbb{R}^s, \{f_i\} \subset H^2(\mathcal{M})}{\arg\min} \| S_i^{1/2} - \mathbf{z}_i (K_i \Psi f_i)^T \|_F^2 + \lambda \|\mathbf{z}_i\|^2 \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f_i, \qquad i = 1, \dots, n. \quad (3.14)$$

Each estimate $\hat{f}_i$, from the proposed model (3.14), can be interpreted as a regularized estimate of the leading PC function of $S_i^{1/2}$ and thus of the eigenfunction ψ_{i1}. The subsequent eigenfunctions can be estimated by classical deflation methods, i.e. by removing the estimated components $\hat{\mathbf{z}}_i (K_i \Psi \hat{f}_i)^T$ from $S_i^{1/2}$ and reapplying model (3.14). This leads to a set of estimates $\{\hat{f}_{ir}\}$ and $\{\hat{\mathbf{z}}_{ir}\}$.

The unregularized version of model (3.14) is equivalent to a Singular Value Decomposition applied to each matrix $S_i^{1/2}$ independently, which would lead to a set orthogonal estimates $\{\hat{\mathbf{z}}_{ir}\}_r \subset \mathbb{R}^s$, for each $i = 1, \dots, n$. Hence, despite not imposing any explicit orthogonality constraints in the regularized model, the scores vectors tend in practice to be

orthogonal. The estimated PC components could also be orthogonalized post-estimation by means of a QR decomposition.

An approximate representation of $S_i = (S_i^{1/2})^T S_i^{1/2}$ is thus given by

$$S_i = K_i \sum_r \left\{ \|\hat{\mathbf{z}}_{ir}\|_2^2 (\Psi \hat{f}_{ir})(\Psi \hat{f}_{ir})^T \right\} K_i^T, \tag{3.15}$$

and the associated approximate representation of C_i, in terms of $\hat{\mathbf{z}}_{ir}$ and $\hat{f}_{ir}$, is

$$C_i = \sum_r \|\hat{\mathbf{z}}_{ir}\|_2^2 \hat{f}_{ir} \otimes \hat{f}_{ir},$$

where $\hat{\kappa}_{ir} := \|\hat{\mathbf{z}}_{ir}\|_2^2$ is an estimate of the variance κ_{ir} and $\hat{f}_{ir}$ is an estimate of ψ_{ir}. The regularizing term in (3.14) introduces spatial coherence on the estimated $\hat{f}_{ir}$ and thus on the estimated eigenfunctions of $\{C_i\}$, fundamental in an inverse problems setting.

On the space of reconstructed covariances $\{C_i\}$, we can then define a distance and perform PCA on vectorizations of $(C_i(v_j, v_l))_{jl}$, as done in Dryden et al. (2009) and Pigoli et al. (2014). However, this is prohibitive due to the high-dimensionality of $(C_i(v_j, v_l))_{jl}$.

In the terminology introduced in this thesis, Model (3.14) follows a Reconstruct-then-Estimate approach, as the covariance functions $\{C_i\}$ are reconstructed individually and PCA is performed on the reconstructed functions. In the next section, we introduce a Regularized-Estimate approach.

A population model

Let $\{\hat{\mathbf{z}}_i\}_{i=1}^n \subset \mathbb{R}^s$ and $\hat{f} \in H^2(\mathcal{M})$ be given by the following model:

$$(\{\hat{\mathbf{z}}_i\}, \hat{f}) = \underset{\{\mathbf{z}_i\} \subset \mathbb{R}^s, f \in H^2(\mathcal{M})}{\arg\min} \sum_{i=1}^n \|S_i^{1/2} - \mathbf{z}_i (K_i \Psi f)^T\|_F^2 + \lambda \sum_{i=1}^n \|\mathbf{z}_i\|^2 \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \tag{3.16}$$

The newly defined model, as opposed to model (3.14), has now a subject specific s-dimensional vector $\mathbf{z}_i$ and a term f that is common to all samples. As in the previous model, the subsequent components can be estimated by deflation methods, leading to a set of estimates $\hat{f}_r$ and $\hat{\mathbf{z}}_{ir}$.

The empirical term in Model (3.16) suggests an approximate representation of S_i that is

$$C_i = \sum_r \|\hat{\mathbf{z}}_{ir}\|_2^2 \hat{f}_r \otimes \hat{f}_r, \tag{3.17}$$

where each underlying covariance function C_i is approximated by the product of a subject specific constant $\|\hat{\mathbf{z}}_{ir}\|_2^2$ and a component $\hat{f}_r \otimes \hat{f}_r$ common to all the observations. The regularizing term in (3.16) introduces spatial coherence on the estimated functions $\{\hat{f}_r\}$.

The covariance operators $\{\mathscr{C}_i\}$ are said to be commuting if $\mathscr{C}_i\mathscr{C}_{i'} = \mathscr{C}_{i'}\mathscr{C}_i$ for all $i, i' = 1, \ldots, n$. This property can be equivalently characterized as

$$C_i(v, v') = \sum_{r=1}^{\infty} \kappa_{ir}\psi_r(v)\psi_r(v'), \qquad \forall v, v' \in \mathscr{M}, \tag{3.18}$$

with $\{\kappa_{ir}\}_r$ subject-specific variances and $\{\psi_r\}$ a set of common orthonormal functions. Thus, a collection of commuting covariance operators is such that its covariance operators can be simultaneously diagonalized by a basis $\{\psi_r\}$. In this case, the functions $\{\hat{f}_r\}$ can be regarded as estimates of $\{\psi_r\}$ and $\{\hat{\kappa}_{ir} := \|\hat{\mathbf{z}}_{ir}\|_2^2\}$ estimates of $\{\kappa_{ir}\}$.

On the one hand, Model (3.16) constrains the estimated covariances to be of the form $C_i = \sum_r \hat{\kappa}_{ir}\hat{f}_r \otimes \hat{f}_r$ and not of the more general form $C_i = \sum_r \hat{\kappa}_{ir}\hat{f}_{ir} \otimes \hat{f}_{ir}$. On the other hand, such a model takes advantage of all the n samples to estimate the components $\{\hat{f}_r \otimes \hat{f}_r\}$, so it could be regarded as a Regularized-Estimate model. The associated variables $\{\hat{\kappa}_{ir}\}$ give a convenient approximate description of the ith covariance, as they are comparable across samples, as opposed to the one computed from Model (3.15). In fact, the ith covariance function can be represented by the variance vector $(\hat{\kappa}_{i1}, \ldots, \hat{\kappa}_{iR})^T$, for a suitable truncation level R, where each entry is associated to the rank-one component $\hat{f}_r \otimes \hat{f}_r$. For each r, a scatter plot of the variances $\{\kappa_{ir}\}_i$, helps understand what is the average contribution of the rth components and what is its variability across samples. Model (3.17) could also be interpreted as a common PCA model (Benko et al., 2009; Flury, 1984), as $\{\hat{f}_r\}$ are the estimated regularized eigenfunctions of the pooled covariance $C = \sum_{i=1}^n C_i$.

Potentially, PCA could be performed on the descriptors $(\hat{\kappa}_{i1}, \ldots, \hat{\kappa}_{iR})^T$ to find rank-R components that maximize the variance of linear combinations of $\{\hat{\kappa}_{ir}\}$ (i.e. the variance of the variances). However, results would be difficult to interpret, as they would involve variations that are rank-R covariance functions around the rank-R mean covariance function.

3.4.2 Algorithm

The minimization in (3.14), for each fixed i, is a particular case of the one in (3.7) (see Section 3.3.3), so we focus on the minimization problem in (3.16) which is approached in an iterative fashion. We set $\sum_{i=1}^n \|\mathbf{z}_i\|^2 = 1$ in the estimation procedure. This leads to the

estimates of $\{\mathbf{z}_i\}$, given f, that are

$$\mathbf{z}_i = \frac{\tilde{\mathbf{z}}_i}{\sqrt{\sum_{i=1}^{n} \|\tilde{\mathbf{z}}_i\|^2}}, \qquad i = 1, \dots, n,$$

with

$$\tilde{\mathbf{z}}_i = S_i^{1/2} K_i \Psi f_h, \qquad i = 1, \dots, n.$$

The estimate of f given $\{\mathbf{z}_i\}$, in the discrete space V introduced in Section 3.3.2, is given by the following proposition.

Proposition 4. *The Surface Finite Element solution $\hat{f}_h \in V$ of model (3.16), given the vectors $\{\mathbf{z}_i\}$, is $\hat{f}_h = \hat{\mathbf{c}}^T \boldsymbol{\phi}$ where $\hat{\mathbf{c}}$ is the solution of*

$$\Big(\sum_{i=1}^{n} \|\mathbf{z}_i\|^2 K_i^T K_i + \lambda A M^{-1} A \Big) \hat{\mathbf{c}} = \sum_{i=1}^{n} K_i^T S_i^{T/2} \mathbf{z}_i. \tag{3.19}$$

Algorithm 3 contains a summary of the estimation procedure. From a practical point of view, the choice to define the representation basis to be a collection of rank one (i.e. separable) covariance functions, of the type $F = \hat{f} \otimes \hat{f}$, is mainly driven by the following reasons. Firstly, rank-one covariance functions are easier to interpret 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 (3.14), and this is fundamental in a setting of indirectly observed covariance functions. Finally, due to their size, it might not be possible to store the full covariance functions on the brain space, instead Model 3.17 allows an efficient joint representation of such covariance functions in terms of their rank-one components.

3.5 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 3.6. 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 $\mathscr{M}$ 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.

Algorithm 3 Inverse Covariance fPCA Algorithm

1: Square-root decompositions

(a) Compute the representations $S_1^{1/2},\dots,S_n^{1/2}$ from $S_1,\dots,S_n$ as

$$S_i^{1/2} = V_i D_i^{1/2},$$

with $S_i = V_i D_i V_i^T$ the spectral decomposition of S_i.

2: Initialization:

(a) Computation of M and A

(b) Initialize $\{\mathbf{z}_i\}_{i=1}^n$, the scores of the first PC

3: PC function's estimation from model (3.14):
Compute $\mathbf{c}$ such that

$$\Big(\sum_{i=1}^n \|\mathbf{z}_i\|^2 K_i^T K_i + \lambda A M^{-1} A\Big)\mathbf{c} = \sum_{i=1}^n K_i^T S_i^{T/2}\mathbf{z}_i$$

$$f_h \leftarrow \mathbf{c}^T \boldsymbol{\phi}$$

4: Scores estimation from model (3.14):

$$\mathbf{z}_i \leftarrow S_i^{1/2} K_i \Psi f_h, \qquad i = 1,\dots,n$$

$$\mathbf{z}_i \leftarrow \frac{\mathbf{z}_i}{\sqrt{\sum_{i=1}^n \|\mathbf{z}_i\|^2}}, \qquad i = 1,\dots,n$$

5: Repeat Step 3-4 until convergence

3.5.1 Indirectly observed functions

We consider $\mathscr{M}_{\mathscr{T}}$ to be a triangular mesh, with 8K nodes, representing the cortical surface geometry of a subject, as shown on the left panel of Figure 3.1. Each of the nodes will represent the discrete set of locations $\{v_j\}$ associated to the sampling operator Ψ. 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 241×24K matrix.

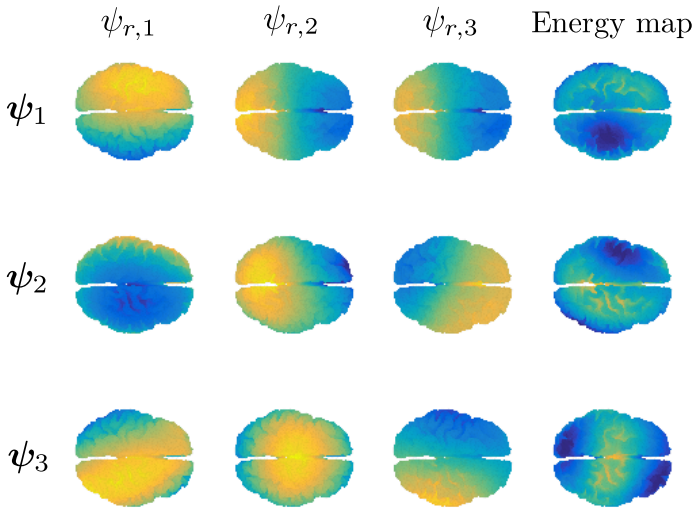


Figure 3.5 From top to bottom the components and the energy maps of the PC functions $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$.

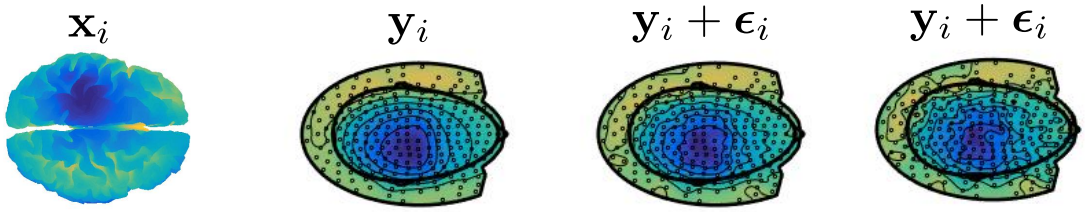


Figure 3.6 From left to right, the energy map of a generated function $\mathbf{x}_i$, the associated signal $\mathbf{y}_i$ on the sensors space with respectively no additional error, Gaussian error of standard deviation $\sigma = 5$ and Gaussian error of standard deviation $\sigma = 10$.

We first want to assess the performances of the proposed model in the case of indirect functional observations belonging to a linear space. To this purpose, we produce synthetic data following the generative model (3.5). Specifically, on $\mathscr{M}_{\mathscr{T}}$, we construct the three L^2 orthonormal vector-valued functions $\{\boldsymbol{\psi}_r = (\psi_{r,1}, \psi_{r,2}, \psi_{r,3}) : r = 1,2,3\}$, with $\boldsymbol{\psi}_r : \mathscr{M}_{\mathscr{T}} \to \mathbb{R}^3$. These represent the PC functions to be estimated. In Figure 3.5 we show the three components of $\{\boldsymbol{\psi}_r\}$ and the associated energy maps $\{\|\boldsymbol{\psi}_r\|^2 : r = 1,2,3\}$, with $\|\cdot\|$ denoting the Euclidean norm in $\mathbb{R}^3$. We then generate $n = 50$ smooth vector-valued functions $\{\mathbf{x}_i\}$ on $\mathscr{M}_{\mathscr{T}}$ by

$$\mathbf{x}_i = z_{i1}\boldsymbol{\psi}_1 + z_{i2}\boldsymbol{\psi}_2 + z_{i3}\boldsymbol{\psi}_3 \qquad i = 1,\dots,n,$$

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

In the following, we compare the PC model (3.7) to an alternative approach. In fact, the individual functions $\{\mathbf{x}_i\}$ could be estimated from $\{\mathbf{y}_i\}$ by use of classical inverse problem estimators. Here, we adopt the estimates $\{\hat{\mathbf{x}}_i\}$ defined as

$$\hat{\mathbf{x}}_i = \underset{\substack{\mathbf{f}=(f_1,f_2,f_3):\\ f_1,f_2,f_3\in H^2(\mathscr{M})}}{\arg\min} \sum_{i=1}^{n} \|\mathbf{y}_i - K\Psi\mathbf{f}\|^2 + \lambda \int_{\mathscr{M}} \|\Delta_{\mathscr{M}}\mathbf{f}\|^2, \qquad i = 1,\dots,n, \tag{3.20}$$

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

$$\Delta_{\mathscr{M}}\mathbf{f} = \begin{bmatrix} \Delta_{\mathscr{M}} & 0 & 0 \\ 0 & \Delta_{\mathscr{M}} & 0 \\ 0 & 0 & \Delta_{\mathscr{M}} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix},$$

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 models introduced in Sections 3.3-3.4. The constant λ is chosen by K-fold cross-validation, with $K = 2$. Once we obtain the estimates $\{\hat{\mathbf{x}}_i\}$ we can compute the estimated PC functions $\{\boldsymbol{\psi}_r\}$ by applying classical multivariate PC analysis on the reconstructed objects $\hat{\mathbf{x}}_i$.

The estimates are compared to the proposed PC function model, as described in Algorithm 2, with 15 iterations. The smoothing coefficient λ 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 rth PC function is then measured with $\sum_{l=1}^{3} \|\nabla_{\mathscr{M}}(\psi_{r,l} - \hat{\psi}_{r,l})\|^2$. The operator $\nabla_{\mathscr{M}}$ is the gradient operator on the manifold $\mathscr{M}$ 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 3.7, for two different signal to noise ratios, where the Gaussian noise has standard deviation $\sigma = 5$ and $\sigma = 10$. In Figure 3.6 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 3.8.

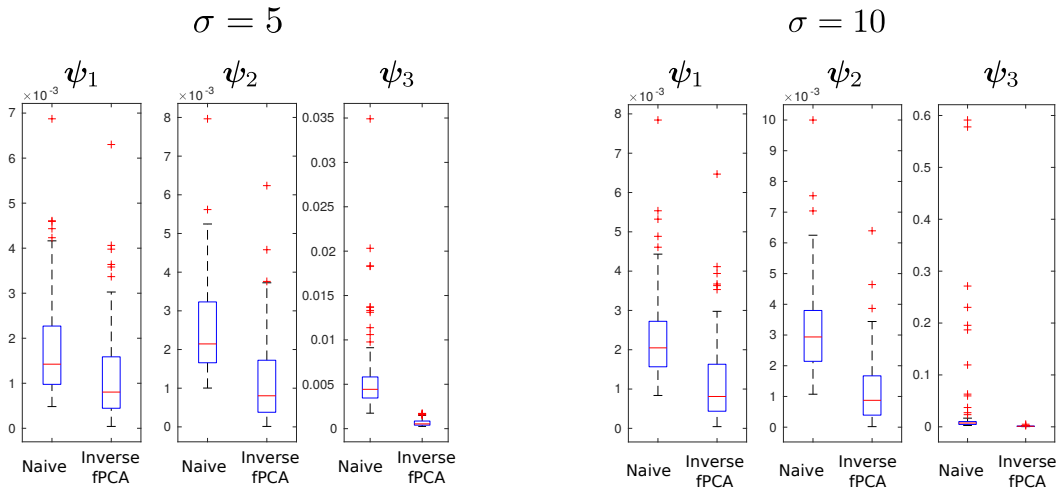


Figure 3.7 On the left, a summary of the results in a medium signal-to-noise ratio regime. On the right, a summary of the results in a low signal-to-noise ratio regime. Each boxplot compares the reconstruction errors obtained by applying the two steps naive method with those obtained by applying Algorithm 2.

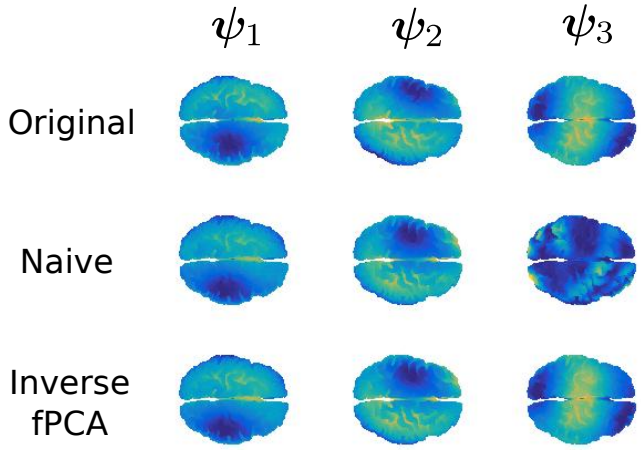


Figure 3.8 On the first row the energy maps of the true three PC components to be estimated, on the second row the estimations given by the two steps naive method, and on the third row the reconstructions obtained by applying Algorithm 2.

3.5.2 Indirectly observed covariance functions

In this section, we consider $\mathscr{M}_{\mathscr{T}}$ to be a $8K$ nodes triangular mesh, this time representing a template geometry of the cortical surface, which is shown in Figure 3.9. 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 ith subject. The $8K$ nodes of each subject's triangular mesh are in correspondence with the $8K$ nodes of the template mesh $\mathscr{M}_{\mathscr{T}}$. This allows the model to be defined on the template $\mathscr{M}_{\mathscr{T}}$.

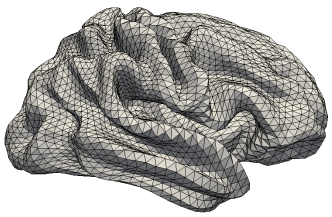


Figure 3.9 The template triangular mesh $\mathscr{M}_{\mathscr{T}}$ composed of 8K nodes.

As in the previous Section, we construct three functions, L^2 orthonormal in $\mathscr{M}_{\mathscr{T}}$ $\{\boldsymbol{\psi}_r = (\psi_{r,1}, \psi_{r,2}, \psi_{r,3}) : r = 1,2,3\}$. The energy maps of $\{\boldsymbol{\psi}_r\}$ are shown in Figure 3.10. We generate synthetic data from model (3.6) as follows:

$$C_i = \sum_{r=1}^{3} z_{ir}^2 \boldsymbol{\psi}_r \otimes \boldsymbol{\psi}_r = \sum_{r=1}^{3} z_{ir}^2 \begin{bmatrix} \psi_{r,1} \otimes \psi_{r,1} & \psi_{r,1} \otimes \psi_{r,2} & \psi_{r,1} \otimes \psi_{r,3} \\ \psi_{r,2} \otimes \psi_{r,1} & \psi_{r,2} \otimes \psi_{r,2} & \psi_{r,2} \otimes \psi_{r,3} \\ \psi_{r,3} \otimes \psi_{r,1} & \psi_{r,3} \otimes \psi_{r,2} & \psi_{r,3} \otimes \psi_{r,3} \end{bmatrix},$$

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 $\mathbb{C}_i \in \mathbb{R}^{24K \times 24K}$, from which the correspondent covariance matrices on the sensors space are defined as

$$S_i = K_i \mathbb{C}_i K_i^T + E_i^T E_i, \qquad i = 1, \ldots, n.$$

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

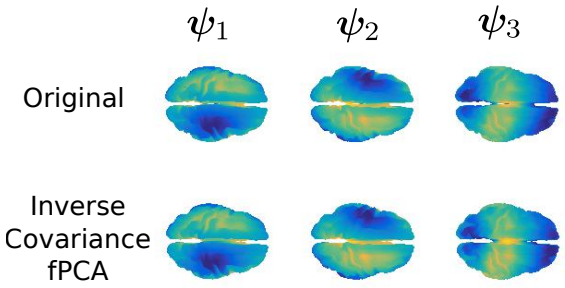


Figure 3.10 On the top row, the energy maps of ψ_1, ψ_2 and ψ_3. On the bottom row the energy maps of the estimates $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying Algorithm 3.

independent sample from a Gaussian distribution with mean zero and standard deviation 5. We then apply Algorithm 3 with, 15 iterations, feeding in input $\{S_i\}$. The results are shown in Figure 3.10, in terms of energy maps of the reconstructed functions $\{\hat{\boldsymbol{\psi}}_r\}$. These are a close approximation of the underlying functions $\{\boldsymbol{\psi}_r\}$. The fidelity measure $\sum_{l=1}^{3} \|\nabla_{\mathcal{M}}(\psi_{r,l} - \hat{\psi}_{r,l})\|^2$ of such estimates is 1.6×10^{-3}, 1.4×10^{-3} and 1×10^{-2}, for $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2$ and $\boldsymbol{\psi}_3$ respectively, which is comparable in terms 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 λ that under-smoothes the solution, due to very similar associated signals on the sensors space of the under-smoothed solution and the real solution. However, the cross-validation is only a possible approach to the choice of the penalization constant, and other options have been proposed in the inverse problems literature, (see, e.g., Vogel, 2002). These, however, might involve visual inspection.

3.6 Application

In this section, we apply the developed models to the publicly available HCP Young Adult dataset (Van Essen et al., 2012). In the following, we briefly review the pre-processing pipeline, applied to such data by the HCP, to ultimately facilitate their use.

3.6.1 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 3.1, 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 3.9. 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 248×24K 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.

3.6.2 Analysis

Here we apply the models proposed in this chapter 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 $S_1, \dots, S_{20}$. The aim is understanding the main modes of variation of the functional connectivity of the subject on its brain space. Thus, Algorithm 3 is applied to $S_1, \dots, S_{20}$ to find the common modes of variation, with 20 iterations and K-fold cross-validation, with $K = 2$. The energy maps of the estimated $\hat{\boldsymbol{\psi}}_1$, $\hat{\boldsymbol{\psi}}_2$ and $\hat{\boldsymbol{\psi}}_3$ resulting from the analysis are shown in Figure 3.11. These are associated to the first three common modes of variation $\hat{\boldsymbol{\psi}}_1 \otimes \hat{\boldsymbol{\psi}}_1$, $\hat{\boldsymbol{\psi}}_2 \otimes \hat{\boldsymbol{\psi}}_2$ and $\hat{\boldsymbol{\psi}}_3 \otimes \hat{\boldsymbol{\psi}}_3$, which are in fact difficult to visualize and interpret, so instead it is useful to focus on the energy maps of $\hat{\boldsymbol{\psi}}_1$, $\hat{\boldsymbol{\psi}}_2$ and $\hat{\boldsymbol{\psi}}_3$. In fact, high intensity areas, in yellow, give a good representation of which areas present high average connectivity. From the plots of the associated variances $\{\hat{\kappa}_{ir}\}$ we can see that these areas are also the ones that show high variability in connectivity across time. These are points where care should be taken in establishing static functional connectivity assessments.

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 $S_1, \dots, S_{40}$. The template geometry in Figure 3.9 is used as a model of the brain space. Algorithm 3 is then applied to $S_1, \dots, S_{40}$ to find the common modes of variation on the template brain, associated to $S_1, \dots, S_{40}$. 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 $\hat{\boldsymbol{\psi}}_1$, $\hat{\boldsymbol{\psi}}_2$ and $\hat{\boldsymbol{\psi}}_3$, associated to the first three common modes of variation $\hat{\boldsymbol{\psi}}_1 \otimes \hat{\boldsymbol{\psi}}_1$, $\hat{\boldsymbol{\psi}}_2 \otimes \hat{\boldsymbol{\psi}}_2$ and $\hat{\boldsymbol{\psi}}_3 \otimes \hat{\boldsymbol{\psi}}_3$, are shown in Figure 3.12. High intensity areas, in yellow, indicate which areas present high average connectivity, and from the plots of the associated variances we can see that these are also ordered by amount of variability in connectivity among subjects.

The presented methodology 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 analyzing connectivity networks. In fact, it is of interest to note that the modes of variation in functional connectivity, for the single and multi subject settings, respectively in Figure 3.11 and 3.12, are similar

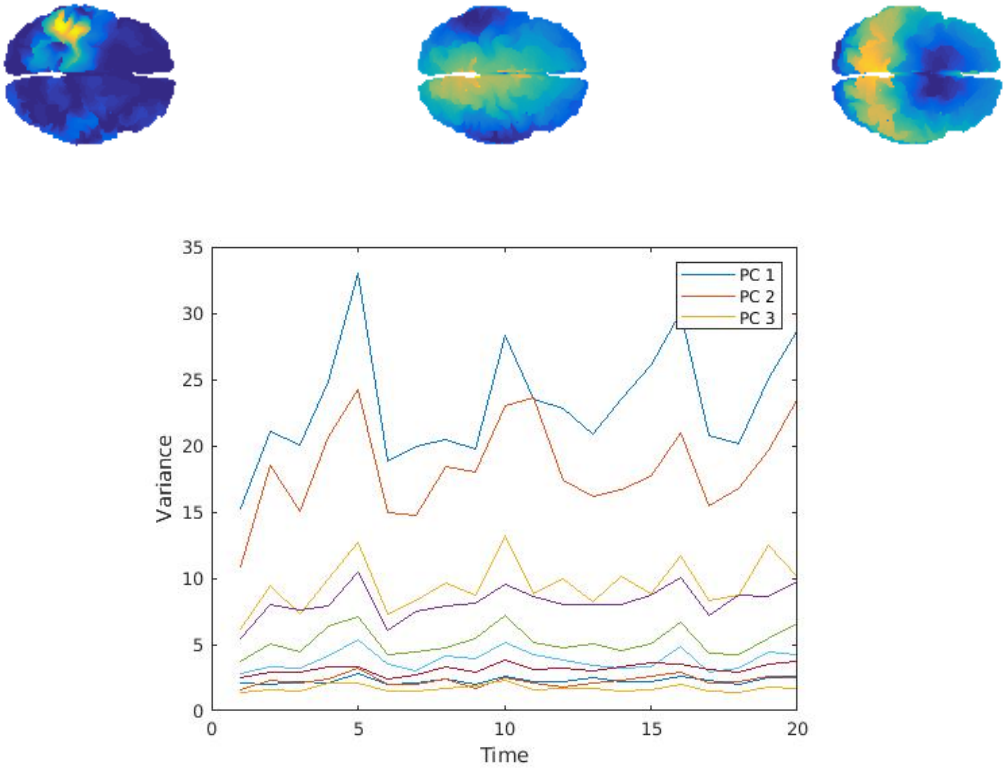


Figure 3.11 On the top, energy maps of the estimated $\hat{\boldsymbol{\psi}}_1$, $\hat{\boldsymbol{\psi}}_2$ and $\hat{\boldsymbol{\psi}}_3$ obtained by applying Algorithm 3 to the covariance matrices computed from the MEG resting state data of a single subject on 20 consecutive time intervals. On the bottom, a plot of the variances associated to the first 10 rank-one components, including $\hat{\kappa}_{i1}$, $\hat{\kappa}_{i2}$ and $\hat{\kappa}_{i3}$.

in the first mode and largely overlapping with the Inferior parietal lobule, although only on the left hemisphere. The second mode of variation for the single subject setting and the third mode for the multi-subject setting, respectively in Figure 3.11 and 3.12, seem highlighting a dynamic behavior of the Posterior cingulate cortex. Both the Posterior cingulate cortex and the Inferior parietal lobule are part of the default network (Buckner et al., 2008), i.e. the brain regions known to have highly correlated activity when the subject is not performing any specific task. This might suggest that the default network, or a subpart of it, is not only a brain area with high functional connectivity levels, in resting state conditions, but it is also the brain region that shows among the highest levels of spontaneous variability in connectivity.

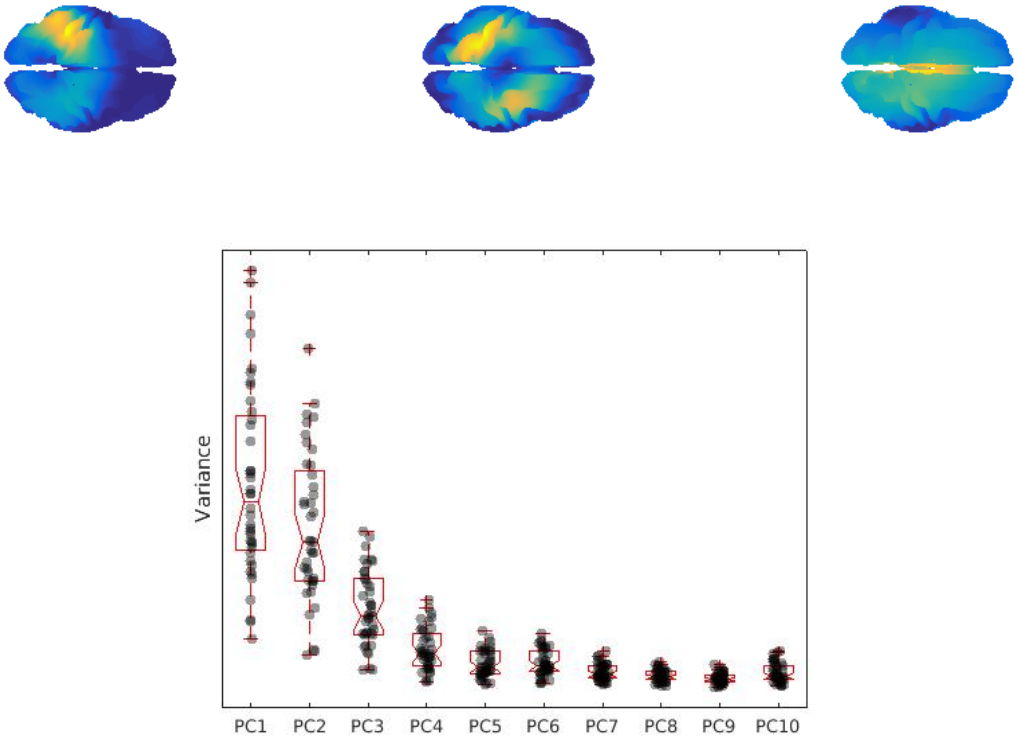


Figure 3.12 On the top, energy maps of the estimated $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying Algorithm 3 to the covariance matrices computed from the MEG resting state data of 40 different subjects. On the bottom, a plot of the variances $\hat{\kappa}_{ir}$ associated to the first 10 rank-one components.

3.7 Discussion

In this chapter, we introduce a general framework for the representation and analysis of covariance operators in an inverse problem context. We first introduce a model for indirectly observed functional images in an unconstrained space, which outperforms the naive approach of solving the inverse problem individually for each sample. This model plays an important role in the case of samples that are indirectly observed covariance functions, and thus constrained to be positive semidefinite. We deal with the non-linearity introduced by such constraints by working with unconstrained representations, 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 set-

tings, 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 population approach we adopt in this chapter. 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.

3.8 Proofs

Here we present the proofs of Proposition 3 and Proposition 4. The proof of Proposition 2 and the subsequent FE discretization follow from that of Proposition 3 as a special case, with $K_1, \dots, K_n = I_p$, the identity matrix of size p.

Proof of Proposition 3. We want to find a minimizer $\hat{f} \in H^2(\mathcal{M})$, given $\mathbf{z}$ with $\|\mathbf{z}\| = 1$, of the objective function in (3.7):

$$\begin{aligned}
&\sum_{i=1}^{n} \|\mathbf{y}_i - z_i K_i \Psi f\|^2 + \lambda \mathbf{z}^T \mathbf{z} \int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f \\
&\propto (\Psi f)^T (\sum_{i=1}^{n} z_i^2 K_i^T K_i) \Psi f - 2(\Psi f)^T (\sum_{i=1}^{n} z_i K_i^T \mathbf{y}_i) + \lambda \int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f. \qquad (3.21)
\end{aligned}$$

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

$$(\Psi\varphi)^T (\sum_{i=1}^{n} z_i^2 K_i^T K_i) \Psi \hat{f} + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}} \varphi \Delta_{\mathcal{M}} \hat{f} = (\Psi\varphi)^T (\sum_{i=1}^{n} z_i K_i^T \mathbf{y}_i) \qquad (3.22)$$

for every $\varphi \in H^2(\mathcal{M})$ (see Braess, 2007, Chapter 2). Moreover, such minimizer is unique if $A(\varphi, f) = (\Psi\varphi)^T (\sum_{i=1}^{n} z_i^2 K_i^T K_i) \Psi f + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}} \varphi \Delta_{\mathcal{M}} f$ is positive definite. Given that for a closed manifold $\mathcal{M}$, $\int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f = 0$ iff f is a constant function (Dziuk and Elliott, 2013), such condition is equivalent to assuming 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 reformulate equation (3.22) in a form that involves only first-order derivatives by integration by parts against a test function. We then look for a solution in

the discrete space $V \subset H^1(\mathcal{M})$, i.e. finding $\hat{f}, g \in V$

$$\begin{cases} (\Psi\varphi)^T(\sum_{i=1}^n z_i^2 K_i^T K_i)\Psi\hat{f} + \lambda \int_{\mathcal{M}} \nabla_{\mathcal{M}}\varphi \cdot \nabla_{\mathcal{M}} h = (\Psi\varphi)^T(\sum_{i=1}^n z_i K_i^T \mathbf{y}_i) \\ \int_{\mathcal{M}} \nabla_{\mathcal{M}}\hat{f} \cdot \nabla_{\mathcal{M}} v - \int_{\mathcal{M}} hv = 0 \end{cases} \tag{3.23}$$

for all $h, v \in V$. The operator $\nabla_{\mathcal{M}}$ is the gradient operator on the manifold $\mathcal{M}$. The gradient operator $\nabla_{\mathcal{M}}$ is such that $(\nabla_{\mathcal{M}} v)(p)$, 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.

We recall here the definition of the $\kappa \times \kappa$ matrices to be $(M)_{jl} = \int_{\mathcal{M}_\mathcal{T}} \phi_j \phi_l$ and $(A)_{jl} = \int_{\mathcal{M}_\mathcal{T}} \nabla_{\mathcal{M}_\mathcal{T}} \phi_j \cdot \nabla_{\mathcal{M}_\mathcal{T}} \phi_l$. Note that requiring (3.23) to hold for all $h, v \in V$ is equivalent to requiring that (3.23) holds for all h, v that are basis elements of V, thus exploiting the basis expansion formula (3.11) we can characterize (3.23) with the solution of the linear system

$$\begin{bmatrix} \sum_{i=1}^n z_i^2 K_i^T K_i & \lambda A \\ A & -M \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \\ \hat{\mathbf{q}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n z_i K_i^T \mathbf{y}_i \\ \mathbf{0} \end{bmatrix}, \tag{3.24}$$

where $\hat{\mathbf{c}}$ and $\hat{\mathbf{q}}$ are the basis coefficients of $f \in V$ and $g \in V$, respectively. Solving (3.24) in $\hat{\mathbf{c}}$ leads to

$$(\sum_{i=1}^n z_i^2 K_i^T K_i + \lambda A M^{-1} A)\hat{\mathbf{c}} = \sum_{i=1}^n z_i K_i^T \mathbf{y}_i. \tag{3.25}$$

□

Proof of Proposition 4. We want to find a minimizer $\hat{f} \in H^2(\mathcal{M})$, given $\{\mathbf{z}_i\}$ with $\sum_{i=1}^n \|\mathbf{z}_i\|^2 = 1$, of the objective function in (3.14):

$$\begin{aligned} &\sum_{i=1}^n \|S_i^{1/2} - \mathbf{z}_i (K_i \Psi f)^T\|^2 + \lambda \sum_{i=1}^n \|\mathbf{z}_i\|^2 \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f \\ &\propto (\Psi f)^T (\sum_{i=1}^n \|\mathbf{z}_i\|^2 K_i^T K_i)\Psi f - 2(\Psi f)^T \sum_{i=1}^n K_i^T S_i^{T/2} \mathbf{z}_i. \end{aligned} \tag{3.26}$$

Comparing (3.26) with (3.21) it is evident that by following the same steps of the Proof of Proposition 3 we obtain the desired result, which is

$$\hat{\mathbf{c}} = \Big(\sum_{i=1}^n \|\mathbf{z}_i\|^2 K_i^T K_i + \lambda A M^{-1} A \Big)^{-1} \sum_{i=1}^n K_i^T S_i^{T/2} \mathbf{z}_i.$$

□

Chapter 4

Functions on Surfaces

4.1 Motivation

Advances in medical imaging acquisition are constantly increasing the complexity of data representing anatomical objects. In particular, some of these imaging modalities offer a richer representation of anatomical manifolds, as a geometric object coupled with a function defined on the geometric object itself, i.e. a Function on a Surface (FoS). In this work we focus on Functions on Surfaces (FoSs) that are real functions located on domains that are two-dimensional manifolds, where the domains themselves are subject to variability from sample to sample, as shown in Figure 1. In the applied mathematics literature, these are also known with the name of Functional Shapes (Charon and Trouvé, 2014). However, as it will be clear from the methodological section of this chapter, the proposed framework can be extended to deal with more complex situations, such as vector-valued functions describing features arising from multi-modal imaging techniques or the RGB representation of colors, as done in Yao et al. (2017), with the purpose of inferring the underlying geometry. Further extensions could also include situations where the functions have an inherent time component. For simplicity of exposition, we will concentrate on univariate FoS data in this work.

The aim of the present chapter is the introduction of a comprehensive statistical framework for the analysis of FoSs. To this end, a statistical model is introduced, with the main aim of jointly representing the *geometric variability* and *functional variability* of the data. Suppose there is an underlying true one-to-one correspondence between the points on the geometries of the observed FoSs. By geometric variability we mean variations on the shape of the domains, i.e. variations of the point positions from one FoS to another. By functional variability we mean variations on the amplitude of the functions

of the observed FoSs, at the points in correspondence. For instance, it is evident that the three FoSs in Figure 1, show both geometric variability and functional variability. In order to quantify these two types of variability, we introduce estimators of the underlying unknown quantities within the proposed statistical model.

In this context, estimating functional variability is challenging because, in such high-dimensional settings, there is need to incorporate prior information, like smoothness, on complex domains. Estimating geometric variability is challenging because the space where the geometric objects live is non-Euclidean, and this invalidates classical linear models, which could lead to predictions that do not belong to the original space, for instance, self-intersecting objects. The formulation of estimators constrained to lie in the deformation space is therefore required. Moreover, as it is clear from their definitions, the study of geometric variability cannot be performed independently from the study of functional variability, as the results of the latter generally depend on the former. This motivates the introduction of a novel diffeomorphic registration algorithm for functional data whose domain is a two-dimensional manifold, which enables the exploitation of the functional information to achieve a better registration.

Often times practitioners have approached the analysis of FoSs in two completely separate steps. In the first step, the surfaces are registered to a template surface, and the functions are transported on the template through such estimated registration maps. In the second step, the analysis of the functions is performed on the template surface, independently of the previous step. This approach has two main drawbacks. Firstly, the complete separation of the two steps precludes any study aimed at understanding how the geometric variability relates to the functional variability. Secondly, for each subject, there is an infinite number of registrations that bring the template to match the target surface. However, for different registration maps, the registered functions exhibit different functional variability. In other words, the registration step is responsible for separating the variability due to geometric differences from the variability due to differences in the functions, and this strongly influences the subsequent analysis on the functions. Thus, the two steps should not be performed independently. Indeed, in the one-dimensional analogue situation, it has been seen that considerably more information can be gleaned from a joint approach than a step-wise approach (see Marron et al., 2015, and references therein).

Many ideas from the literature on image registration (see e.g. Dupuis et al., 1998; Thirion, 1995, 1998) and the literature on landmarked shapes (see e.g. Bookstein, 1992, 1997; Dryden and Mardia, 2016) have been recently extended to the more general setting of

surfaces, without a functional component, both from the applied mathematics prospective (Younes, 2010) and from the statistical prospective (Patrangenaru and Ellingson, 2015). It is also natural to contextualize FoSs in the Functional Data Analysis (FDA) framework. However, FDA is generally performed in controlled environments, where data are assumed to be smooth functions on a fixed interval of the real line (Ramsay and Silverman, 2005), or more generally, smooth functions on a fixed domain. The setting considered here represents a new challenge for this branch of statistics.

More recently, a joint mathematical model for geometric and functional variability has been proposed in Charlier et al. (2017). The approach consists of generalizing the notion of deformation to a notion of metamorphosis, introduced for 2D images in Trouvé and Younes (2005). A metamorphosis includes both a geometric deformation term and an additive functional term. This enables the representation of any FoS as a metamorphosis of a template FoS. The geometric deformation and the functional additive term, to explain a given FoS, can be weighted by two different parameters in the model. In contrast, our approach takes a statistical perspective on the problem of analyzing a set of FoSs, and aims to offer a methodological toolset that can be feasibly applied to the analysis of the brain surfaces shown in Figure 1.

4.1.1 Motivating application

The motivating application of the proposed model is the study of a collection of FoSs derived from Magnetic Resonance Imaging (MRI). A 2D surface representing the geometry of the cerebral cortex, the outermost layer of the brain, can be extracted from 3D MRI data thanks to fully automated surface-extraction algorithms (Glasser et al., 2013). The cerebral cortex is a highly convoluted thin sheet of 2 to 4 millimeters of thickness which consists of neuronal cell bodies and it is the source of large parts of our neuronal activity. An illustration of the surface-extraction step is shown in Figure 4.1.

Thanks to complementary imaging techniques, like functional MRI, a function can be associated to the estimated cerebral cortex (see, e.g., Hagler et al., 2006), resulting in a FoS. Such functions can be vector-valued functions, where each component represents a feature of the cerebral cortex, extracted from a different imaging technique. However, in this work, the function we consider is the map of thickness measurements of the cerebral cortex. In fact, thanks to the recent improvements of the resolution of MRI scans it is now possible to have an accurate estimation of this thickness map (Lerch and Evans, 2005). Details on the cerebral cortex surface reconstruction and the cortical thickness estimation

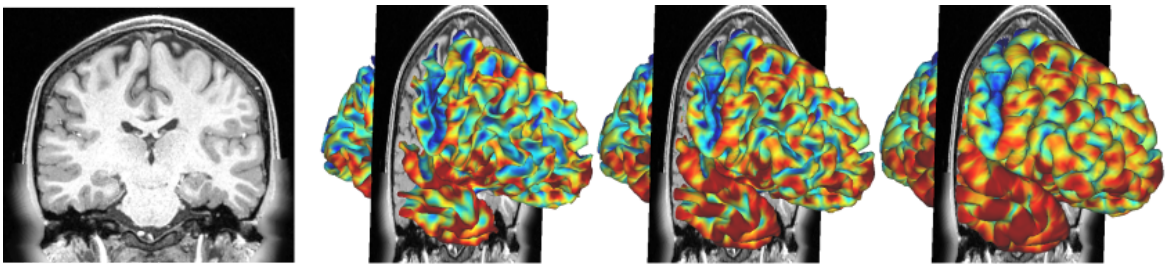


Figure 4.1 From left to right, in the first panel we have a section of a structural MRI (T1-weighted) of a single subject. In the next three panels we have, respectively, the estimated inner, mid-thickness and outer surfaces of the cerebral cortex. The inner and outer surfaces enclose the cerebral cortex, the mid-thickness surface interpolates its middle points and it is used to represent its geometry. These images have been produced using Connectome Workbench (Glasser et al., 2013).

are covered in Section 4.5. In Figure 1 we show the FoSs representing cerebral cortex geometry and thickness of three different individuals.

Almost all studies of these kind presume a preprocessing registration step and so do not consider the inherent variability effects that might be induced by the registration step on the functional measurements. Indeed, this issue goes beyond neuroimaging, as the same techniques are often used in a wide variety of medical imaging settings (Audette et al., 2000), as well as computer vision applications (Zaetz and Kurtek, 2015).

The rest of the chapter is organized as follows. In Section 4.2 we introduce a generative statistical model which allows for both geometric and functional variability. In Section 4.3, we propose the statistical estimators of the underlying unknown quantities of the generative model. We perform a simulation study on synthetic data in Section 4.4, to investigate our estimation procedure. We then apply the framework introduced to study the relation between geometry and thickness of the human cerebral cortex in Section 4.5 and draw some concluding remarks in Section 4.6. Moreover, in the appendices, we present further details of the proposed methodology and an additional simulation study.

4.2 Model for Functions on Surfaces

4.2.1 Definitions

A set of FoSs, such as the ones in Figure 1, can be mathematically formulated as a collection of pairs $\{(\mathcal{M}_i, Y_i) : i = 1, \ldots, n\}$. The collection $\{\mathcal{M}_i : i = 1, \ldots, n\}$ is a set of topologically equivalent smooth two-dimensional manifolds, embedded in $\mathbb{R}^3$, representing the ge-

ometry of the data. The functional aspect of the data is represented by the collection $\{Y_i : i = 1,\dots,n\}$, where Y_i is an element of the function space $L^2(\mathscr{M}_i)$, i.e. the Hilbert space of square integrable functions on $\mathscr{M}_i$ with respect to the area measure.

Here, we propose a statistical generative model for FoSs, modelled in terms of mathematically more tractable objects. To this end, we define a *deformation operator* φ, such that $\varphi_v : \mathbb{R}^3 \to \mathbb{R}^3$ is parametrized by the elements of a Hilbert space $\{v : v \in \mathscr{V}\}$. Moreover, we assume φ_v is a homeomorphism of $\mathbb{R}^3$ for all $v \in \mathscr{V}$ and that $\varphi_0(x) = x$ for all $x \in \mathbb{R}^3$. For each $v \in \mathscr{V}$, $\varphi_v : \mathbb{R}^3 \to \mathbb{R}^3$ represents a deformation of the space $\mathbb{R}^3$, which means that when φ_v is applied to a point $x \in \mathbb{R}^3$ this is relocated to the location $\varphi_v(x) \in \mathbb{R}^3$. In addition, φ_v being a homeomorphism of $\mathbb{R}^3$ implies that, for a fixed $v \in \mathscr{V}$, there is a one-to-one correspondence between each element $x \in \mathbb{R}^3$ and the relocated element $\varphi_v(x) \in \mathbb{R}^3$.

Moreover, we introduce $\mathscr{M}_0$, a smooth two-dimensional manifold topologically equivalent to $\{\mathscr{M}_i\}$, which represents a fixed template geometric object. Given a FoS, the geometric template together with the deformation operator offers an alternative representation of the geometry of the FoS in hand as: $\varphi_v \circ \mathscr{M}_0$, for a particular choice of $v \in \mathscr{V}$. Here, $\varphi_v \circ \mathscr{M}_0$ is the geometric object obtained by deforming $\mathscr{M}_0$ through the map φ_v, and specifically, by relocating each point $x \in \mathscr{M}_0$ to the new location $\varphi_v(x)$, to resemble the target manifold. For this reason, we will informally say that the element $v \in \mathscr{V}$ encodes the geometry, or the shape, of a FoS, as in fact v defines the deformation φ_v, which defines the geometry $\varphi_v \circ \mathscr{M}_0$. The choice of the deformation operator is driven by the particular problem in hand. We first introduce the generative model and subsequently discuss different choices of this operator.

4.2.2 The model

Let now $\{v_i : i = 1,\dots,n\}$ be a set of random samples of a zero-mean and finite second moment $\mathscr{V}$-valued random function V and $\{Z_i : i = 1,\dots,n\}$ be a set of random smooth samples of a zero-mean and finite second moment random real function Z with values in $L^2(\mathscr{M}_0)$. We assume the following generative model for the ith observation $(\mathscr{M}_i, Y_i)$:

$$\begin{cases} \mathscr{M}_i &= \varphi_{v_i} \circ \mathscr{M}_0, \\ X_i &= \mu + \delta Z_i, \\ Y_i &= X_i \circ \varphi_{v_i}^{-1}, \end{cases} \tag{4.1}$$

where $\mu \in L^2(\mathscr{M}_0)$ is a fixed function, modelling the common function behavior between the different samples, and δ is a coefficient representing the magnitude of the function variations around the mean μ. In addition, we assume the objects in Model 4.1 are subject to a discretization error, which is considered in the estimation process. This formulation generalizes an often used model for the one-dimensional functional registration problem (see, e.g. Tang and Müller (2008)).

Model 4.1 achieves the goal of representing FoSs as a collection of more tractable objects, decomposing the generation of the ith FoS into three main steps. In the first step, the geometry $\mathscr{M}_i$ of the ith object is generated by the deformation φ_{v_i} applied to the template $\mathscr{M}_0$, where v_i is a random sample of V. In the second step, a random function X_i, on the template, is generated as the sum of the fixed function μ and a stochastic term δZ_i. In the third step the generated function X_i is transported on the manifold $\mathscr{M}_i$, defining Y_i. This is done through the equation $Y_i = X_i \circ \varphi_{v_i}^{-1}$, which means that for all $x \in \mathscr{M}_0$, $Y_i(\varphi_{v_i}(x)) = X_i(x)$, or informally that the functional value $X_i(x)$ is 'transported' with the deformation to the location $\varphi_{v_i}(x) \in \mathscr{M}_i$.

We now describe the FoSs generation process from Model 4.1, for different choices of the deformation operator:

- *Shift operator*: Let $\mathcal{V} = \mathbb{R}^3$, we define φ_v to be such that $\varphi_v(x) = x + v$ for all $v \in \mathcal{V}$, $x \in \mathbb{R}^3$. Clearly, in this case, $\{\mathscr{M}_i = \varphi_{v_i} \circ \mathscr{M}_0\}$ in Model 4.1 would generate a collection of surfaces shifted in the directions specified by $\{v_i\}$.

- *Identity operator*: Let $\mathcal{V}$ be the space of smooth functions $v : \mathscr{M}_0 \to \mathbb{R}^3$ and let $\varphi_v(x) = x + v(x)$ for all $x \in \mathscr{M}_0$. In this case, $\{\mathscr{M}_i\}$ would be a collection of smoothly deformed versions of the template $\mathscr{M}_0$. Note however, that the maps being only smooth and not homeomorphic, it cannot be guaranteed that every choice of $v \in \mathcal{V}$ preserves the topology of $\mathscr{M}_0$. Nevertheless, this choice might still represent a valid option in a small deformations setting.

  To solve this problem, we could think of restricting $\mathcal{V}$ to contain only smooth and homeomorphic functions, however, in this way, the linearity of the space $\mathcal{V}$ is lost, and this is a property of fundamental importance to the subsequent analysis, given that we want to apply linear statistics on the random function V, which takes values on $\mathcal{V}$.

- *Diffeomorphic operator*: Let $\mathcal{V}$ be a Sobolev space of sufficiently smooth vector fields from $\mathbb{R}^3$ to $\mathbb{R}^3$ vanishing, with their derivatives, at infinity. Let φ be a diffeomorphic

deformation operator, i.e. an operator such that φ_v is a diffeomorphism of $\mathbb{R}^3$ for all $v \in \mathcal{V}$. Then, for different choices of v, Model 4.1 would generate a collection of surfaces that are diffeomorphic (and thus homeomorphic) deformations of the template $\mathcal{M}_0$. More importantly, these deformations are parametrized by the linear space $\mathcal{V}$, where linear statistics can be applied. For this choice, an illustration of the generative process is shown in Figure 4.2. The diffeomorphic deformation operator can be defined by means of an Ordinary Differential Equation (ODE). Details of this are described in Section 4.2.4.

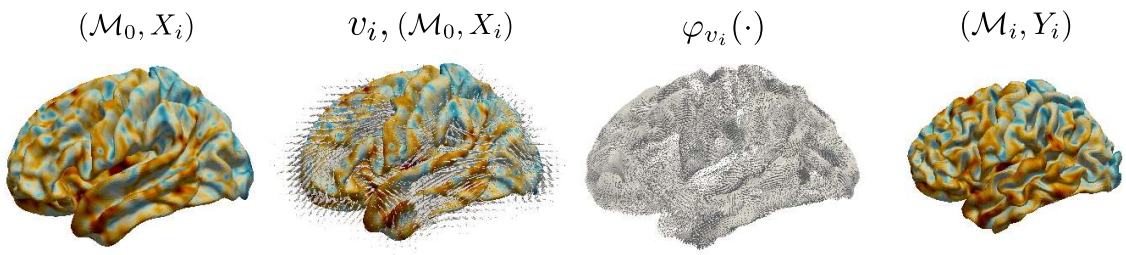


Figure 4.2 An illustration of the generation of a FoS through Model 4.1 with φ the diffeomorphic deformation operator. From left to right, in the first panel we have a functional sample X_i on the geometric template $\mathcal{M}_0$. In the second panel we have a vector field $v_i \in \mathcal{V}$, a sample of the random function V, evaluated on a uniform grid in $\mathbb{R}^3$. This is shown together with $(\mathcal{M}_0, X_i)$. In the third panel we have the diffeomorphic deformation φ_{v_i}, obtained from v_i as described in Section 4.2.4, here displayed as the set of vectors $\{\varphi_{v_i}(\xi_k)\} \subset \mathbb{R}^3$ with $\{\xi_k\}$ the nodes of the triangulated surface representing the template $\mathcal{M}_0$. In the fourth panel, we have the FoS $(\mathcal{M}_i, Y_i)$ obtained by applying the deformation φ_{v_i} to $\mathcal{M}_0$ and 'transporting' the functional values with it.

More complicated generative models could be built from Model 4.1. For example, the functions $\{v_i\}$ and $\{X_i\}$, representing respectively geometries and functions, could be modelled in terms of conditional expectation of different sources of information on the subjects such as age, disease status or other subject-specific explanatory variables, as done, in the case of functional data located on 1D domains, in Hadjipantelis et al. (2015). However, Model 4.1 is the simplest model enabling a comprehensive study of the relation between geometric and functional variability.

4.2.3 Geometric and Functional variability

Here we formalize the geometric and functional variability relationship. Recalling the definition of geometric and functional variability, given in the introduction, we can notice that in Model 4.1 we have that $\{v_i\}$ describe the geometric variability in the data, while

$\{X_i\}$ describe the functional variability in the data. The key idea of this work is to formalize geometric and functional variability by means of functional Principal Component Analysis (fPCA), so that geometric variability can be represented in terms of the Principal Components (PCs) of the random function V, generating the samples $\{v_i\}$, and functional variability can be represented in terms of the PCs of the random function $X = \mu + \delta Z$ generating the random samples $\{X_i\}$. Under the hypothesis that a finite number of PCs is sufficient to represent V and X, we can than use classical multivariate statistics, such as multivariate regression or canonical correlation analysis, to model the relation between the PCs of V and the PCs of X, ultimately formalizing the concept of geometric and functional variability being related. This should also further clarify the choice to introduce a deformation operator φ. In fact, as already mentioned, the deformation operator allows us to parametrize the space of deformations through the linear space $\mathcal{V}$, and thus linear fPCA can be applied on the $\mathcal{V}$-valued random variable V.

More formally, under typical assumptions on V, thanks to fPCA, V can be expanded in terms of the orthonormal sequence of eigenfunctions $\{\psi_j^G\}$ of the covariance operator of V, as

$$V = \sum_{j=1}^{\infty} a_j^G \psi_j^G,$$

where $a_1^G, a_2^G, \ldots$ are uncorrelated real random variables, with variances in decreasing order $\kappa_1^G, \kappa_2^G, \ldots$. The collection $\{\psi_j^G\}$ defines the strongest modes of variation of the random function V and these are called PC functions. We refer to ψ_j^G as the jth mode of geometric variation, or alternatively the jth geometric PC function. This represents variation of the type $c\psi_j^G$ around the mean of V, with $c \in \mathbb{R}$. The PC function ψ_j^G is thus associated to the geometric deformations $\varphi_{c\psi_j^G}$ of $\mathbb{R}^3$, that applied to the geometric template, correspond to the geometries described by $\varphi_{c\psi_j^G} \circ \mathcal{M}_0$. In practice, we visualize the jth mode of geometric variation by visualizing the associated geometries for some specific choice of c, e.g. $\varphi_{\pm\sqrt{\kappa_j^G}\psi_j^G} \circ \mathcal{M}_0$. An example of this visualization is given in Figure 4.15. PCA has been previously used in a similar fashion in Vaillant et al. (2004) and Tward et al. (2017), with φ the diffeomorphic deformation operator, to represent anatomical geometries.

With analogous considerations, the random variable X can be expanded, using the associated orthonormal eigenfunctions $\{\psi_j^F\}$, of the covariance operator of X, as

$$X = \mu + \sum_{j=1}^{\infty} a_j^F \psi_j^F,$$

where the real random variables $a_1^F, a_2^F, \ldots$ are uncorrelated with variances, in decreasing order, $\kappa_1^F, \kappa_2^F, \ldots$. We refer to ψ_j^F as the jth mode of functional variation.

FPCA basis expansions have the fundamental property of separating the discrete set of stochastic terms from the functional terms. Hence, the relation between the geometry and the functional terms can be formalized in terms of the random variables $\{a_j^G\}$ and $\{a_j^F\}$. We assume that only a finite number of the PC functions are necessary to describe the phenomenon in hand and denote with $\mathbf{a}^G$ the associated K^G-dimensional random vector $(a_1^G, \ldots, a_{K^G}^G)$ and with $\mathbf{a}^F$ the K^F-dimensional random vector $(a_1^F, \ldots, a_{K^F}^F)$.

Different multivariate statistical models can be applied at this stage, to formalize the geometric and functional variability relation in terms of the relation between the random vectors $(a_1^G, \ldots, a_{K^G}^G)$ and $(a_1^F, \ldots, a_{K^F}^F)$. A first possible formalization of the geometric and functional variability relation is

$$\mathbb{E}[X|V] = \mu + \sum_{j=1}^{K^F} \mathbb{E}[a_j^F|\mathbf{a}^G]\psi_j^F. \tag{4.2}$$

Under linear assumptions on the dependency, the conditional expectation term can be modelled as

$$\mathbb{E}[a_j^F|\mathbf{a}^G] = \boldsymbol{\beta}_j'\mathbf{a}^G,$$

with $\boldsymbol{\beta}_j$ the K^G-dimension vector of the regression coefficients of the jth functional mode of variation.

The model above describes how the main modes of geometric variation explain each mode of functional variation, implying that we expect the geometry to influence the functions. This might be the case of neurodegenerative disease, where we expect the functional activity (the function) to adapt to the disease progression (the geometry). However, the reverse roles of geometry on functions is also plausible in some cases. For instance, through a comparative study between taxi drivers and bus drivers, it has been shown that the different functional activation patterns influence the growth of the gray matter volume, and thus the brain geometry (Maguire et al., 2006). Moreover, given that in model (4.2) each mode of functional variations is explained thorugh a linear combination of the modes of geometric variability, the interpretability of the overall model strongly relies on the interpretability of the singular functional main modes of variations.

A second possible formalization of the geometric and functional variability relationship might consist of simply examining the maximal directions of correlation between geometry and function. This is equivalent to performing a Canonical Correlation Analysis

(CCA). A CCA analysis on the coefficients of the fPCA basis expansion is equivalent to finding a new basis expansions for V and X as a linear combination of the respective fPCA basis. However, the elements of the new basis are ordered in a way that maximizes the correlation between their coefficients, i.e. the interdependency between geometry and function, representing how the geometric variability associates with the functional variability and vice versa.

In this work, we focus on studying the relationship between the pair of random functions (V, X) generating the samples $\{(v_i, X_i) : i = 1, \ldots, n\}$ describing respectively the brain geometry and a functional variable of interest. However, in practice, multiple descriptors are available for each individual. For example, besides the cortical thickness maps, we could consider cortical myelin maps, cortical curvature maps, or different types of data such as behavioral and demographic measures, leading to high-dimensional data of the type

$$\left\{\left(v_i, X_i^1, \ldots, X_i^{p_X}, W_i^1, \ldots, W_i^{p_W}\right) : i = 1, \ldots, n\right\},$$

with p_X functional variables $X_i^1, \ldots, X_i^{p_X}$ and p_W univariate descriptors $W_i^1, \ldots, W_i^{p_W}$. A comprehensive analysis in this setting requires more sophisticated tools than CCA to analyze and decompose the variation. A natural candidate is Joint and Individual Variation Explained (JIVE) (Lock et al., 2013), where the variation is subdivided into a low-rank joint variation component, a low-rank individual variation component, and a residual noise component of the n-row datasets

$$\mathbb{V}, \quad \mathbb{X}^1, \quad \cdots \quad \mathbb{X}^{p_X}, \quad \mathbb{W},$$

formed by filling each row with the ith observation of the associated variables. Intuitively, the joint variation component presents a common scores matrix across datasets and a dataset-specific loadings matrix. Instead, the individual variation component allows for both dataset-specific scores and loadings matrices. A more refined analysis could try to model joint variation within subsets of the available variables.

4.2.4 The diffeomorphic deformation operator

The deformation operator, introduced in Section 4.2.1, has to be chosen in such a way that it is flexible enough to represent the observed surfaces, as a deformation of the template surface. Clearly, the shift operator is not sufficient to capture the variations in geometry of the FoSs in Figure 1, in terms of deformation of the template. However, this operator

should only include 'sensible' deformations, in the sense that the deformation operator should have its image contained in the set of diffeomorphic deformations from $\mathbb{R}^3$ to $\mathbb{R}^3$. This choice is driven by the fact that diffeomorphic deformations are smooth deformations that preserve the topological properties of the shapes and that avoid two separate points on the template collapsing to one point on the observed surface.

For this reason, we rely on the idea of constructing diffeomorphic deformations as flows of an ODE (Dupuis et al., 1998), which can be parameterized by a Hilbert function space. Specifically, let now $\mathcal{V}$ be a Sobolev space of sufficiently smooth vector fields from $\mathbb{R}^3$ to $\mathbb{R}^3$ vanishing, with their derivatives, at infinity. Let $v : [0,1] \times \mathbb{R}^3 \to \mathbb{R}^3$ be a *time dependent* vector field in $L^2([0,1], \mathcal{V})$, the space of vector fields with finite (squared) norm $\int_0^1 \|v_t\|_{\mathcal{V}}^2 dt$. Then, for a given v, the solution $\phi_v : [0,1] \times \mathbb{R}^3 \to \mathbb{R}^3$ of the ODE

$$\frac{\partial \phi_v}{\partial t}(t,x) = v_t \circ \phi_v(t,x) \qquad t \in [0,1], x \in \mathbb{R}^3. \tag{4.3}$$

with initial conditions $\phi_v(0,x) = x$, is a smooth diffeomorphic map in $\mathrm{Diff}(\mathbb{R}^3)$, at each fixed time t (see, e.g., Younes, 2010). The ODE (4.3) is intuitively defining the solution ϕ_v to be a function such that, for all $t \in [0,1]$, the time-derivative $\frac{\partial \phi_v}{\partial t}(t,x)$ (i.e. the velocity field at time t) is given by the vector field $v_t \circ \phi_v(t,x)$. In other terms ϕ_v represents the 'flow' described by the velocity vector field $\{v_t : t \in [0,1]\}$.

Note that we use φ and ϕ to represent two different object and their relation is defined as follows.

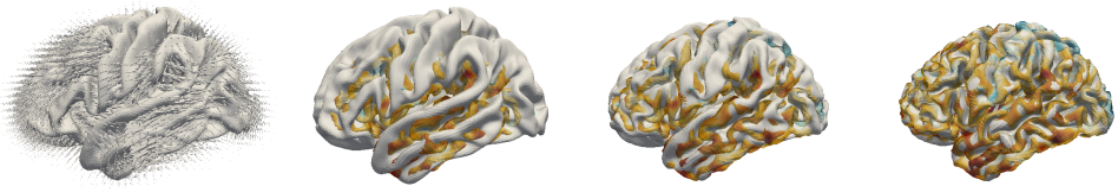


Figure 4.3 From left to right, in the first panel we have an initial vector field $v_0 \in \mathcal{V}$ and in gray the template $\mathcal{M}_0$. In the consecutive panels, we show the solution ϕ_v of the ODE at the times $t = 0, 0.5, 1$ (which are diffeomorphic deformations of $\mathbb{R}^3$), as deformations of the template $\mathcal{M}_0$. In this specific case, the initial vector field v_0 has been chosen in such a way that the surface $\varphi_{v_0} \circ \mathcal{M}_0$ is a close approximation of a target surface, i.e. the colored surface in the figure.

Given an initial vector field v_0, we define $\{v_t : t \in [0,1]\}$ to be the time-variant vector field which minimizes the quantity $\int_0^t \|v_t\|_{\mathcal{V}}^2 dt$. For this choice, the vector field $\{v_t : t \in [0,1]\}$ can be derived from v_0 through the resolution of the EPDiff equation (Miller et al.,

2006). Finally, the deformation operator can be defined to be $\varphi_{v_0}(x) = \phi_v(1,x)$, where $v_0 \in \mathcal{V}$ is the initial vector field generating $\{v_t : t \in [0,1]\}$, through the EPDiff equation, and ϕ_v is the solution of the ODE (4.3). The choice to define $\varphi_{v_0}(x)$ with the solution of the ODE (4.3) at time $t = 1$ is arbitrary, in fact any other choice of a fixed $t > 0$ would have been equivalent, given that the $\phi_v(t,x)$ is guaranteed to be a diffeomorphism of $\mathbb{R}^3$ for any $t > 0$.

A summary of the main elements necessary to define φ is given by the following

$$\underbrace{v_0 \xrightarrow[\text{EPDiff}]{} \{v_t : t \in [0,1]\} \xrightarrow[\text{ODE (4.3)}]{} \phi_v \to \varphi_{v_0} := \phi_v(1,\cdot)}_{\varphi_{v_0}:\mathbb{R}^3 \to \mathbb{R}^3}. \tag{4.4}$$

In Figure 4.3 we show the solution of the ODE (4.3) for a given initial vector field v_0. We emphasize that the ODE (4.3) is not used here to model the phenomenon in hand, but it is just a convenient tool to generate a diffeomorphism of $\mathbb{R}^3$ from a smooth vector field v_0 belonging to the linear space $\mathcal{V}$.

4.3 Estimation framework

The arguments made in the previous section are formalized in terms of quantities derived from the underlying unknown random variables modelling the data generation. However, in practice, only a set of observed noisy FoSs is available, and those quantities have to be estimated from the data. In this section, we mostly work with the set of idealized FoSs $\{(\mathcal{M}_i, Y_i) : i = 1,\dots,n\}$. Instead, when the specific computer representation is of importance to the proposed algorithms, we work with the associated collection of pairs denoted with $\{(\mathcal{M}_i^{\mathcal{T}}, Y_i^{\mathcal{T}}) : i = 1,\dots,n\}$, each composed by a triangulated surface $\mathcal{M}_i^{\mathcal{T}} \subset \mathbb{R}^3$, approximating the underlying smooth two-dimensional manifold $\mathcal{M}_i \subset \mathbb{R}^3$, and a real piecewise linear function $Y_i^{\mathcal{T}} \in L^2(\mathcal{M}_i^{\mathcal{T}})$ representing a noisy approximation of the underlying smooth function $Y_i \in L^2(\mathcal{M}_i)$.

In this section, we outline the estimation procedures applied to the data to recover the different quantities in Model 4.1. A flow chart summarizing the main steps is shown in Figure 4.4. The implementation details are covered in the appendices.

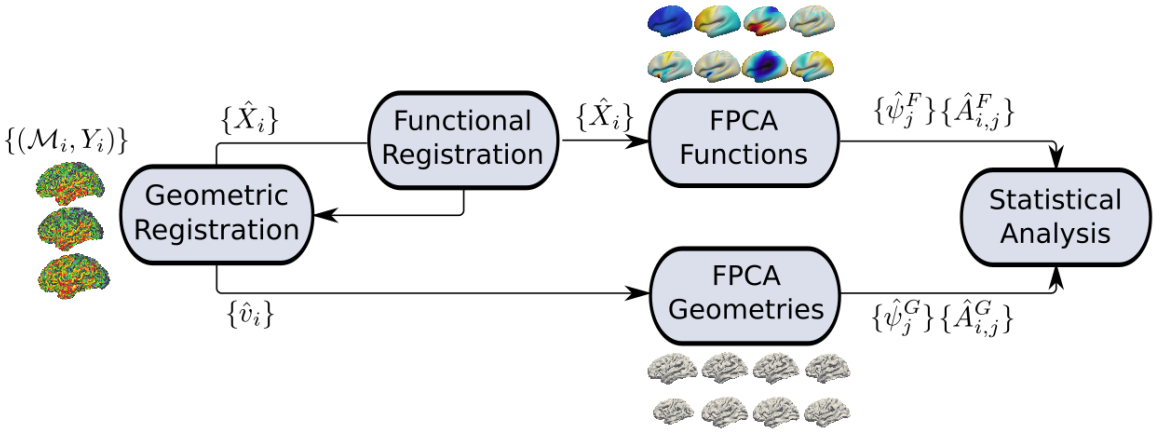


Figure 4.4 A flow chart summarizing the main steps of the estimation procedure proposed in Section 4.3. In the Geometric Registration step, as described in Section 4.3.1, the FoSs are registered to a template $\mathcal{M}_0$, i.e. for each FoS we estimate a vector field $\hat{v}_i$ representing the geometric object $\mathcal{M}_i$ as the deformed template $\varphi_{\hat{v}_i} \circ \mathcal{M}_0$. Each function $\hat{X}_i$ is obtained by transporting Y_i on the template through the estimated registration. In the Functional Registration step, as described in Section 4.3.2, the functional information is used to achieve a more accurate registration leading to a corrected version of the functions $\{\hat{X}_i\}$ and the vector fields $\{\hat{v}_i\}$. In the FPCA Functions and FPCA Geometries steps, as described in Section 4.3.3, fPCA is performed on the functions $\{\hat{X}_i\}$ and the vector fields $\{\hat{v}_i\}$ respectively, leading to the estimation of a set of PC functions and scores representing functional variability and geometric variability. Finally, as described in Section 4.3.4, classical statistical analysis is performed on the PC scores $\hat{A}_{i,j}^F$ and $\hat{A}_{i,j}^G$ to study the relation between functional and geometric variability.

4.3.1 Geometric Registration and Linear representation of shapes

In practice the computation of a diffeomorphic deformation between the template $\mathcal{M}_0$ and the surface $\mathcal{M}_i$ is achieved by solving a minimization problem of the form

$$\hat{v}_i = \underset{v_i \in \mathcal{V}}{\arg\min}\, D^2(\varphi_{v_i} \circ \mathcal{M}_0, \mathcal{M}_i) + \lambda \| v_i \|_{\mathcal{V}}^2, \tag{4.5}$$

where $D^2(\varphi_{v_i} \circ \mathcal{M}_0, \mathcal{M}_i)$, the shape similarity function, is a measure of the amount of mismatching between the deformed template surface and the target surface. The constant λ is a weighting parameter between the data-fidelity term and the term $\| v_i \|_{\mathcal{V}}^2$, which could be regarded as a measure of the amount of deformation induced by φ_{v_i}. The functions $\{\hat{v}_i\}$ are an estimation of $\{v_i\}$ in Model 4.1. In Figure 4.3 we show an example of a vector field in $\mathcal{V}$, estimated by solving (4.5), with the aim of representing a target surface as a deformation of a template.

The procedure described is also referred to as the registration step, as in fact the estimated map $\varphi_{\hat{v}_i}$, up to approximation error, defines a one-to-one smooth correspondence between the points of the target manifold $\mathscr{M}_i$ and the template $\mathscr{M}_0$. Thus, the function $\hat{X}_i \in L^2(\mathscr{M}_0)$, obtained by registering Y_i to the template, can be defined as the element $\hat{X}_i$ such that $\hat{X}_i(x) = Y_i(\varphi_{\hat{v}_i}(x))$ for all $x \in \mathscr{M}_0$. The registered maps $\{\hat{X}_i\}$ can be regarded as a first approximation of $\{X_i\}$ in Model 4.1. In practice, there might be a small approximation error between $\varphi_{v_i} \circ \mathscr{M}_0$ and $\mathscr{M}_i$, which might contrast with the definition $\hat{X}_i(x) = Y_i(\varphi_{\hat{v}_i}(x))$ for all $x \in \mathscr{M}_0$, as $\varphi_{\hat{v}_i}(x)$ might not exactly belong to $\mathscr{M}_i$. However, we assume that $\varphi_{\hat{v}_i}(x)$ is close enough to $\mathscr{M}_i$, for all $x \in \mathscr{M}_0$, and in practice define $\hat{X}_i(x) = Y_i(y)$ with $y \in \mathscr{M}_i$ the nearest neighbor of $\varphi_{\hat{v}_i}(x)$.

The implementation of the registration algorithm (4.5) requires the definition of a shape similarity function D. As already mentioned, the geometry of a FoS is in practice encoded as a triangulated mesh, we thus define the similarity function D between triangulated surfaces. We should differentiate between two possible settings at this point. In the first setting, we suppose that a correspondence between the points of the triangulated surfaces $\mathscr{M}_0^{\mathscr{T}}$ and $\mathscr{M}_i^{\mathscr{T}}$ is known for all $i = 1, \ldots, n$. In other terms, we suppose that $\mathscr{M}_0^{\mathscr{T}}$ and $\mathscr{M}_i^{\mathscr{T}}$ have already been registered and thus there is a set of landmarks $\{x_l, y_l : x_l \in \mathscr{M}_0^{\mathscr{T}}, y_l \in \mathscr{M}_i^{\mathscr{T}}\}$ in correspondence between them. In this case, a simple mismatching functional is given by the Euclidean distance between the correspondent landmarks i.e.

$$D^2(\varphi_{v_i} \circ \mathscr{M}_0^{\mathscr{T}}, \mathscr{M}_i^{\mathscr{T}}) = \sum_l \|\varphi_{v_i}(x_l) - y_l\|^2_{\mathbb{R}^3}. \tag{4.6}$$

This choice has been adopted for instance in Joshi and Miller (2000).

This situation is frequent in neuroimaging, a field that has developed their own ad hoc registration algorithms and where diffeomorphic constraints are explicitly imposed without the necessity to use a diffeomorphic deformation operator. In this case the estimates $\{\hat{X}_i\}$ are already provided, given that $\mathscr{M}_0^{\mathscr{T}}$ and $\mathscr{M}_i^{\mathscr{T}}$ have already been registered, nevertheless, the framework introduced here is still of relevance, in fact, we still need to estimate $\{\hat{v}_i\} \subset \mathscr{V}$, in equation (4.5), to represent the given registration maps (i.e. deformation maps), and thus the geometries, in terms of elements of a linear space, which is a fundamental property to the subsequent analysis.

In the second setting, we suppose that a registration step has not been performed yet. In this situation, registration and linear representation can be performed jointly by choosing an appropriate shape similarity function D not based on landmarks, but for instance, proximity. An example of such a similarity function is proposed in Vaillant and

Glaunès (2005) and Vaillant et al. (2007), and is defined as follows. Let $K_{\mathcal{Z}} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ be a Gaussian isotropic kernel of variance $\sigma^2_{\mathcal{Z}}$, i.e. $K_{\mathcal{Z}}(x,y) = \exp(-\|x-y\|_2^2/(2\sigma^2_{\mathcal{Z}}))\mathrm{Id}_{3\times 3}$, with $\mathrm{Id}_{3\times 3}$ denoting a 3×3 identity matrix. Indeed, such kernel can be any symmetric positive definite kernel, however it is common to choose a Gaussian kernel. Denote with $c(l)$ and $\eta(l)$, respectively, the center point and the normal vector of the lth triangle of the mesh $\varphi_{v_i} \circ \mathscr{M}_0^{\mathcal{T}}$. Denote with $c_i(q)$ and $\eta_i(q)$, respectively, the center point and the normal vector of the qth triangle of the mesh $\mathscr{M}_i^{\mathcal{T}}$. Moreover, let the triangles of the mesh $\varphi_{v_i} \circ \mathscr{M}_0^{\mathcal{T}}$ be indexed by l and g and the triangles in $\mathscr{M}_i^{\mathcal{T}}$ be indexed by q and r. The resulting shape similarity function has the form

$$\begin{aligned} D^2(\varphi_{v_i} \circ \mathscr{M}_0^{\mathcal{T}}, \mathscr{M}_i^{\mathcal{T}}) = &\sum_l \sum_g K_{\mathcal{Z}}(c(l), c(g))\eta(l)\cdot\eta(g) \\ &- 2\sum_l \sum_q K_{\mathcal{Z}}(c(l), c_i(q))\eta(l)\cdot\eta_i(q) \\ &+ \sum_q \sum_r K_{\mathcal{Z}}(c_i(q), c_i(r))\eta_i(q)\cdot\eta_i(r), \end{aligned} \tag{4.7}$$

with $\cdot$ denoting the scalar product in $\mathbb{R}^3$. Intuitively, the first and last terms measure deformations to the local geometry within the two surfaces, and the middle term measures the mismatch in local geometry between the two surfaces.

Thanks to the procedure outlined in this section, given a set of FoSs, we are able to register them to a fixed template $\mathscr{M}_0$. As a result, the information regarding the geometry of the data is stored in terms of the estimates $\{\hat{v}_i\} \subset \mathcal{V}$ of $\{v_i\}$ in Model 4.1. These are estimated so that $\varphi_{\hat{v}_i} \circ \mathscr{M}_0$ resembles the geometry of the ith FoS. Moreover, we obtain a set of functions $\{\hat{X}_i\}$ on the fixed template, that are a first estimate of the functions $\{X_i\}$ in Model 4.1.

In practice the space of smooth functions $\mathcal{V}$ is implemented as a Reproducing Kernel Hilbert Space (RKHS), as described in Appendix B.1.

4.3.2 Functional Registration

The aim of this section is the introduction of a novel functional registration algorithm for functional data whose domain is a fixed two-dimensional manifold. The functional registration algorithm can then be applied to align the set of functions $\{\hat{X}_i : \hat{X}_i \in L^2(\mathscr{M}_0)\}$, estimated in Section 4.3.1, by registering them to a template function $X_0 \in L^2(\mathscr{M}_0)$, which can be in first instance approximated by the cross-sectional sample mean of $\{\hat{X}_i\}$. The rationale for such a procedure is that, as well known in FDA, the functions $\{\hat{X}_i\}$ on $\mathscr{M}_0$

should in principle be able to drive a better registration, on the assumption that the underlying functions $\{X_i\}$ in Model 4.1 have a preponderant mean effect, with respect to its second order variation.

In fact, each estimated function $\hat{X}_i$, strongly depends on the associated deformation map $\varphi_{\hat{v}_i}$, whose estimation is usually driven only by geometric features. Hence, a systematic mis-registration, due to a naive approximation of the deformation maps, could introduce fictitious functional variability on the functions $\{\hat{X}_i\}$, which in fact should be accounted for by geometric variability, in particular in a setting where obvious landmarks are not available and the deformations $\{\varphi_{\hat{v}_i}\}$ are estimated while ignoring the functional information. The functional registration algorithm can be regarded as a correction step to $\{\hat{X}_i\}$, and thus $\{\hat{v}_i\}$, estimated from Section 4.3.1.

A review on the registration of functional data can be found in Marron et al. (2015). However, most of the FDA literature treats only the case of functions whose domain is an interval of the real line. Registration of 2D images has also been well studied (see e.g., Zitová and Flusser, 2003, for a review). Methods that preserve invertibility of the deformation have also been proposed for 2D/3D Euclidean images (Vercauteren et al., 2009) and extended to functions with spherical domains in Yeo et al. (2010). However, to the best of our knowledge, these methods are not able to deal with the registration of a collection of functions whose domain is a fixed generic two-dimensional manifold embedded in $\mathbb{R}^3$.

Alternatively, in the case of landmark based registration, functional information can be introduced into the registration process, by modifying the algorithm that provides the landmarks, to account for function similarity. In the case where landmarks are not available functional information can be introduced by equipping the shape similarity functional (4.7) with a functional similarity term, as done in Charon and Trouvé (2014) and Charlier et al. (2017).

Definitions

To set the notation of this chapter, we recall some of the basic differential geometry definitions presented more in details in Chapter 1. Let $T_p\mathcal{M}_0$ be the tangent space on the point $p \in \mathcal{M}_0$ and let g_p be the metric on $\mathcal{M}_0$, i.e. a scalar product on the tangent space $T_p\mathcal{M}_0$. In our case it is natural to consider the scalar product induced by the Euclidean embedding space $\mathbb{R}^3$, i.e. the first fundamental form. Define the tangent bundle to be the disjoint union of tangent spaces $T\mathcal{M}_0 = \dot{\bigcup}_{p\in\mathcal{M}_0} T_p\mathcal{M}_0 = \bigcup_{p\in\mathcal{M}_0}\{p\} \times T_p\mathcal{M}_0$. A section of the tangent bundle $T\mathcal{M}_0$ is the formalization of the concept of a vector field on $\mathcal{M}_0$, an

example of which is shown in Figure 4.5. We denote with $L^2(T\mathcal{M}_0)$ the Hilbert space of square integrable sections of $T\mathcal{M}_0$. Moreover, let Δ_{BL} be the Bochner-Laplacian operator. The Bochner-Laplacian of a smooth vector field v, i.e. $\Delta_{BL}v$, is a vector field on $\mathcal{M}_0$, whose L^2 norm gives a measure of the smoothness of the vector field v. i.e. low values for smooth vector fields v and high values for rough vector fields. A more formal definition of the Bochner-Laplacian operator, from the Levi-Civita operator, is given in Chapter 1.

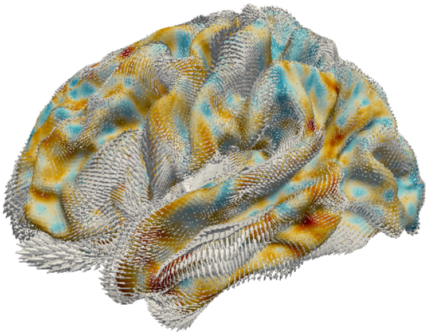


Figure 4.5 A section of the tangent bundle $T\mathcal{M}_0$, which has been computed by minimizing the linearized version of the equation (4.9).

Estimation

The registration of $\{\hat{X}_i\}$ is performed in an iterative fashion, which means that each function $\hat{X}_i$ is aligned to the function X_0 by composition of small diffeomorphic deformations. Let $\{s_i : \mathcal{M}_0 \to \mathcal{M}_0\}$ be the set deformation maps estimated from the previous iterations of the algorithm, such that $\hat{X}_i \circ s_i$ is a registered version of $\hat{X}_i$ to the function X_0. The functions $\{s_i\}$ can be the set of identity maps in the first iteration. Moreover, let $\{p_j : j = 1,\dots,S\} \subset \mathcal{M}_0$ be a collection of S control points where the functions $\{\hat{X}_i\}$ are sub-sampled. In practice, these will be the nodes of the triangulation $\mathcal{M}_0^{\mathcal{T}}$, i.e. the points where the functions are actually observed.

With a slight abuse of notation, let the diffeomorphic function $\phi_u : \mathcal{M}_0 \to \mathcal{M}_0$ be the solution generated at time $t = 1$ by the ODE

$$\begin{cases} \frac{\partial \phi_u}{\partial t}(t,x) = u \circ \phi_u(t,x) & t \in [0,1], x \in \mathcal{M}_0, \\ \phi_u(0,x) = x & x \in \mathcal{M}_0 \end{cases} \tag{4.8}$$

where u is a sufficiently smooth vector field on $\mathcal{M}_0$. If $\mathcal{M}_0$ has a boundary, than we assume u vanishes, with their derivatives, on the boundary. Such an ODE is used here as a tool to

generate a diffeomorphic function ϕ_u from a vector field u that needs only to be smooth. Then, we propose to estimate a set of functional registration maps, each aligning $\hat{X}_i \circ s_i$ to X_0 by minimizing

$$E_{\mathscr{M}_0}(u_i) = \sum_{j=1}^{S} \left(X_0(p_j) - \hat{X}_i \circ s_i \circ \phi_{u_i}(p_j)\right)^2 + \lambda \|\Delta_{BL} u_i\|^2_{L^2(T\mathscr{M}_0)}, \tag{4.9}$$

where $\|\Delta_{BL} u_i\|^2_{L^2(T\mathscr{M}_0)}$ is the L^2 norm of the vector field $\Delta_{BL} u_i$, which imposes smoothness on u_i. The constant λ is a weighting coefficient between the data fidelity term, i.e. how well aligned we want $\hat{X}_i \circ s_i \circ \phi_{u_i}$ to be to X_0, and the smoothing term, i.e. how smooth we want the vector field u_i to be.

The term $\hat{X}_i \circ s_i \circ \phi_{u_i}$ in Equation (4.9), is then linearized with respect to u_i. This results in the approximation

$$\hat{X}_i \circ s_i \circ \phi_{u_i} \approx \hat{X}_i \circ s_i + L_{u_i},$$

where L_{u_i} is a first order approximation of $\hat{X}_i \circ s_i \circ \phi_{u_i} - \hat{X}_i \circ s_i$. By means of Vector Finite Elements, an approximate solution $\hat{u}_i$, at the nodes of $\mathscr{M}_0^{\mathscr{T}}$, can be characterized in terms of the solution of a linear system. An approximate vector field $\hat{u}_i$ on the triangulation $\mathscr{M}_0^{\mathscr{T}}$ is then computed by linear interpolation of the solution found at the nodes of the triangulation. Details of this procedure can be found in Appendix B.2. The main steps of the functional registration algorithm are summarized in Algorithm 4.

Each iteration of the functional registration algorithm result in a newly estimated set of functions $\{\hat{X}_i \circ s_i^k\}$, representing a re-aligned correction of the maps $\{\hat{X}_i\}$. The composition $\hat{X}_i \circ s_i^k$ means that for all $x \in \mathscr{M}_0$ the functional value $\hat{X}_i(x)$ is, after k iterations, relocated on the point $(s_i^k)^{-1}(x) \in \mathscr{M}_0$. Thus, the functional registration also has the effect of correcting the overall geometric deformations $\{\varphi_{\hat{v}_i} : \mathscr{M}_0 \to \mathscr{M}_i\}$, estimated in Section 4.3.1, to be

$$\varphi_{\hat{v}_i} \circ (s_i^k)^{-1}, \qquad i = 1, \ldots, n. \tag{4.10}$$

The geometric registration model in Section 4.3.1 and the functional registration model, introduced in this section, are similar in spirit, as they both rely on the idea that given a smooth vector field we can generate a diffeomorphic vector field by means of an ODE. However, they also differ in many aspects. For instance, they differ in the way smoothness is imposed. In the geometric registration model, smoothness is imposed by penalizing through the norm of a RKHS. In the functional registration model, within each iteration, smoothness is imposed by means of a differential operator, while the overall smoothness is controlled by the number of iterations. Moreover, in the geometric

Algorithm 4 Functional Registration Algorithm

1: Initialization:
   (a) Initialize $\{s_i^0 \leftarrow Id : i = 1,\dots,n\}$ to be the identity functions on $\mathscr{M}_0$
   (b) Initialize $\{\hat{X}_i : i = 1,\dots,n\}$ to be the functions estimated from Section 4.3.1
   (c) Initialize the functional template to be $X_0 \leftarrow \frac{1}{n}\sum_i \hat{X}_i$
2: Compute $\{\hat{u}_i^k : i = 1,\dots,n\}$, the solution at the kth iteration, from the second order functional

$$E_{\mathscr{M}_0}(u_i) = \sum_{j=1}^{S} \left(X_0(p_j) - \hat{X}_i \circ s_i^{k-1}(p_j) - L_{u_i}(p_j)\right)^2 + \lambda \|\Delta_{BL} u_i\|^2_{L^2(T\mathscr{M}_0)},$$

3: Compute the registration maps $\{\phi_{\hat{u}_i^k} : i = 1,\dots,n\}$ by solving the ODE

$$\begin{cases} \frac{\partial \phi_{u_i^k}}{\partial t}(t,x) = u_i^k \circ \phi_{u_i^k}(t,x) & t \in [0,1], x \in \mathscr{M}_0, \\ \phi_{u_i^k}(0,x) = x & x \in \mathscr{M}_0 \end{cases}$$

4: Update current registration maps and functional template:
   $\{s_i^k \leftarrow s_i^{k-1} \circ \phi_{\hat{u}_i^k} : i = 1,\dots,n\}$
   $X_0 \leftarrow \frac{1}{n}\sum_i \hat{X}_i \circ s_i^k$
5: Output and analysis (e.g. fPCA) of the result of the current iteration:
   $\{\hat{X}_i \circ s_i^k : i = 1,\dots,n\}$
6: Repeat Steps 2–5 until until a stopping criterion is satisfied

registration, the ODE is defined for a time-variant vector field, instead in the functional model the ODE is defined for a stationary vector field. Some of these aspects, including the link between penalizing through the norm of a RKHS and penalizing by means of a differential operator, are discussed in Appendix B.2.

The functional registration model introduced in this section, as opposed to the geometric registration model in Section 4.3.1, is based on the composition of small deformations, where at the kth iteration, $\{\hat{X}_i \circ s_i^k : i = 1,\dots,n\}$ represent the re-aligned versions of $\{\hat{X}_i \circ s_i^{k-1} : i = 1,\dots,n\}$. The constant λ in the model, controls the change between the functions $\{\hat{X}_i \circ s_i^{k-1}\}$ and those estimated at the next iteration, as in fact large values of λ privilege small deformations. This has the advantage that the fPCA analysis can be re-performed on the functions $\{\hat{X}_i \circ s_i^k\}$ at each iteration k. The output of this analysis can provide useful information for the next step of the functional registration algorithm, as for instance a stopping criterion in a similar fashion to Kneip and Ramsay (2008).

In summary, we have introduced a method that exploits the functional information to achieve a better registration by updating the functional estimates $\{\hat{X}_i\}$ to be $\{\hat{X}_i \circ s_i^{k_{\text{stop}}}\}$, and the diffeomorphic geometric deformations $\varphi_{\hat{v}_i}$ to be $\{\varphi_{\hat{v}_i} \circ (s_i^{k_{\text{stop}}})^{-1}\}$, where k_{stop} denotes the iteration where the functional algorithm is stopped. As pointed out at many stages in this work, it is however important to have a representation of the final update of the deformations maps $\{\varphi_{\hat{v}_i} \circ (s_i^{k_{\text{stop}}})^{-1}\}$ in terms of elements of the linear Hilbert space $\mathcal{V}$, so that we can perform linear statistics. To this purpose, we can estimate such elements by applying the geometric deformation model in Section 4.3.1, i.e. by solving

$$\hat{v}_i^{k_{\text{stop}}} = \underset{v_i \in \mathcal{V}}{\arg\min}\, D^2(\varphi_{v_i} \circ \mathcal{M}_0^{\mathcal{T}}, \varphi_{\hat{v}_i} \circ (s_i^{k_{\text{stop}}})^{-1} \circ \mathcal{M}_0^{\mathcal{T}}) + \lambda \| v_i \|_{\mathcal{V}}^2,$$

where D^2 denotes the landmark distance defined in equation (4.6).

The overall procedure in this section results in a set of corrected estimates $\{\hat{X}_i \circ s_i^{k_{\text{stop}}}\}$ and $\{\hat{v}_i^{k_{\text{stop}}}\}$, that exploit functional information, estimating respectively $\{X_i\}$ and $\{v_i\}$ in Model 4.1. To ease the notation, in the next section, we drop the index on the number of iterations of the functional registration algorithm, denoting with $\{\hat{X}_i\}$ and $\{\hat{v}_i\}$ the corrected estimates of functions and geometric deformations respectively.

Remarks on computational times

It is also important to highlight that the idea of alternating between each iteration of the functional registration algorithm and the fPCA analysis on the functions is ultimately enabled by the computational efficiency of the proposed functional registration algorithm and fPCA algorithm on the functions. In the case of the application, in Section 4.5, each FoS is represented by a 32K nodes triangulated surface and the associated 32K functional values on the nodes. In this setting, the computational time of one iteration of the functional registration algorithm, applied between two functions, is in the order of 2 minutes on a Intel Core i5-3470 3.20GHz workstation, with 4 GB of RAM. The computational time for a singular PC of the functions is 15 seconds, on the same workstation, with the fPCA implementation proposed in Section 4.3.3. Instead, the landmark driven geometric registration of the 32K nodes template to a 32K nodes surface, representing a cerebral cortex, takes approximately 3 hours on a cluster's node equipped with a Dell T620 server and a NVIDIA K20 GPU.

4.3.3 Functional Principal Component Analysis

In Sections 4.3.1-4.3.2 we introduced the estimation procedure for the objects $\{\hat{v}_i : i = 1,\dots,n\}$ representing the geometries and $\{\hat{X}_i : i = 1,\dots,n\}$ representing the functions, from a set of n FoSs. In this section, the aim is to outline the estimation procedure to the empirical PC component functions from the observed objects $\{\hat{v}_i\}$ and $\{\hat{X}_i\}$, in analogy to what was proposed in Section 4.2.3, in terms of PCs of the underlying random functions V and X.

Geometric variability

The empirical PC functions are in practice computed from the eigen-decomposition of the empirical covariance operator $\hat{C}_{\mathcal{V}}$, defined as

$$\hat{C}_{\mathcal{V}}(v) = \frac{1}{n}\sum_{i=1}^{n}\langle v, \hat{v}_i - \bar{v}\rangle_{\mathcal{V}}(\hat{v}_i - \bar{v}), \qquad v \in \mathcal{V}, \tag{4.11}$$

where $\bar{v} = \frac{1}{n}\sum_{i=1}^{n}\hat{v}_i$ and $\langle\cdot,\cdot\rangle_{\mathcal{V}}$ denotes the scalar product in $\mathcal{V}$. An explicit solution of this eigenvalue problem can be derived by expanding v and $\hat{v}_i$ in (4.11) over a basis of $\mathcal{V}$ or discretizing the problem over a fine grid of $\mathbb{R}^3$. Since the number of observations in this setting is small with respect to the size of the space, an appropriate choice of the basis is given by the collection of the actually observed vector fields $\hat{v}_i$ (Ramsay and Silverman, 2005). Thus, the eigenvalue problem $\hat{C}_{\mathcal{V}}(\psi_j^G) = \kappa_j^G\psi_j^G$ can be re-formulated to a discrete eigenvector problem in terms of the basis expansion coefficients, leading to the empirical PC functions estimates $\{\hat{\psi}_j^G\}$ and empirical variance estimates $\{\hat{\kappa}_j^G\}$. The empirical PC scores vectors can be estimated by projecting $\{\hat{v}_i\}$ on the estimated PC functions, i.e. the ith element of the jth scores vector is given by

$$\hat{A}_{i,j}^G = \langle\hat{v}_i - \bar{v}, \hat{\psi}_j^G\rangle_{\mathcal{V}} \qquad i = 1,\dots,n,\ j = 1,\dots,K^G.$$

The empirical mean $\bar{v}$ can be neglected, as the underlying random function is assumed to have zero mean. The empirical jth mode of geometric variation is thus represented by the PC function $\hat{\psi}_j^G$, which is associated to the deformations $\varphi_{\pm\sqrt{\hat{\kappa}_j^G}\hat{\psi}_j^G}$ of $\mathbb{R}^3$ that applied to the geometric template correspond to the change of geometry described by $\varphi_{\pm\sqrt{\hat{\kappa}_j^G}\hat{\psi}_j^G}\circ\mathcal{M}_0$. The observed vector fields can be finally expressed in terms of the basis expansion, also

known as the Karhunen-Loève expansion:

$$\hat{v}_i \approx \sum_{j=1}^{K^G} \hat{A}^G_{i,j} \hat{\psi}^G_j. \tag{4.12}$$

Equation (4.12) emphasizes the fact that the matrix $(\hat{A}^G_{i,j})_{ij}$ is such that the ith row is a compact description of the vector field $\hat{v}_i$.

Functional variability

From similar arguments, we can build an estimator for the PC functions and PC scores vectors for the functions $\{\hat{X}_i\}$. The estimated functions $\{\hat{X}_i\}$ are noisy estimates of the realization of the underlying unobserved random function X. A pre-smoothing of the noisy functions could be considered, however here we rely on the Regularized-Estimate fPCA algorithm proposed in Chapter 2, where the regularization term is applied directly to the PC functions to be estimated.

In fact, the PC functions $\{\psi^F_j\}$ of the centered random function $X-\mu$, satisfy the following property

$$\{\psi^F_m\}_{m=1}^M = \underset{(\{\psi_m\}_{m=1}^M : \langle \psi_m, \psi_l \rangle_{L^2(\mathscr{M}_0)} = \delta_{ml})}{\arg\min} \mathbb{E} \int_{\mathscr{M}_0} \left\{ X - \mu - \sum_{m=1}^{M} \langle X-\mu, \psi_m \rangle_{L^2(\mathscr{M}_0)} \psi_m \right\}^2, \tag{4.13}$$

where $\int_{\mathscr{M}_0}$ denotes the surface integral over $\mathscr{M}_0$ and $\langle \cdot, \cdot \rangle_{L^2(\mathscr{M}_0)}$ denotes the scalar product in $L^2(\mathscr{M}_0)$. In (4.13) we can see that the PC functions minimize the loss of information caused by the truncation of the series expansion to the first M components. Let $\{p_j : j = 1, \ldots, S\} \subset \mathscr{M}_0$ be a collection of S points where the estimated functions $\{\hat{X}_i\}$ are sub-sampled. In practice, these will be the nodes of the triangulation $\mathscr{M}_0^{\mathscr{T}}$, i.e. the points where the functions are actually observed. Let Δ be the Laplace-Beltrami operator (see e.g. Chavel, 2006). The Laplace-Beltrami operator of a smooth function $f \in L^2(\mathscr{M}_0)$ is a function in $L^2(\mathscr{M}_0)$ that gives a measure of the local curvature of the function f.

The first PC function $\hat{\psi}^F_1 \in L^2(\mathscr{M}_0)$ and associated first scores vector $(\hat{A}^F_{1,1}, \ldots, \hat{A}^F_{n,1})$ are estimated by minimizing the following regularized empirical version of (4.13):

$$(\hat{\psi}^F_1, \{\hat{A}^F_{i,1}\}_{i=1}^n) = \underset{\psi_1, \{A_{i,1}\}_{i=1}^n}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{S} \left(\hat{X}_i(p_j) - \bar{X}(p_j) - A_{i,1} \psi_1(p_j) \right)^2 + \lambda \| \Delta \psi_1 \|_{L^2(\mathscr{M}_0)}, \tag{4.14}$$

where $\bar{X}$ denotes the sample mean function of $\{\hat{X}_i\}$ and λ is a weighting coefficient between the empirical and regularizing term. The regularization term imposes smoothness on the estimated PC function $\hat{\psi}_1^F$, coherently with the structure of the manifold $\mathscr{M}_0$. Subsequent PCs can be estimated by reapplying (4.14) to the residuals. Details of the implementation and an application to functional Magnetic Resonance Imaging can be found in Chapter 2.

The observed functions $\{\hat{X}_i\}$ can be finally expressed in terms of the basis expansion

$$\hat{X}_i \approx \bar{X} + \sum_{j=1}^{K^F} \hat{A}_{i,j}^F \hat{\psi}_j^F. \tag{4.15}$$

The matrix $(\hat{A}_{i,j}^F)_{ij}$ is such that the ith row is a compact description of the function $\hat{X}_i$.

4.3.4 Geometric and Functional variability relation

The matrices $(\hat{A}_{i,j}^G)_{ij}$ and $(\hat{A}_{i,j}^F)_{ij}$, computed in Section 4.3.3, are such that their ith row represents a compact description of the geometry and functions of the ith FoS $(\mathscr{M}_i, Y_i)$. Each row of these matrices could also be regarded as the estimated empirical ith realization of the random vector $(a_1^G, \dots, a_{K^G}^G)$ and $(a_1^F, \dots, a_{K^F}^F)$ defined in Section 4.2.3. As outlined in that section, the matrices $(\hat{A}_{i,j}^G)_{ij}$ and $(\hat{A}_{i,j}^F)_{ij}$ can then be used to study the relation between geometric variability and functional variability of the given collection of FoSs. To this end we can perform, for instance, a linear regression analysis where we try to explain the jth mode of functional variability as a linear combination of the K^G modes of geometric variation.

Alternatively, we could perform CCA, and look for the lth mode of co-variation $(\hat{\mathbf{w}}^{G,l}, \hat{\mathbf{w}}^{F,l})$, representing the lth maximally correlated linear combination $\hat{\mathbf{w}}^{G,l} \in \mathbb{R}^{K^G}$, of the K^G modes of geometric variation with the linear combination $\hat{\mathbf{w}}^{F,l} \in \mathbb{R}^{K^F}$ of the K^F modes of functional variation. The lth mode of co-variation $(\hat{\mathbf{w}}^{G,l}, \hat{\mathbf{w}}^{F,l})$ can be visualized as the sequence of FoSs

$$\begin{cases} \mathscr{M}_{\mathrm{CCA},l} = \varphi_{c\,\hat{\psi}^G_{\mathrm{CCA},l}} \circ \mathscr{M}_0, \\ Y_{\mathrm{CCA},l} = c\,\hat{\psi}^F_{\mathrm{CCA},l} \circ \varphi^{-1}_{c\,\hat{\psi}^G_{\mathrm{CCA},l}}, \end{cases} \tag{4.16}$$

obtained by varying $c \in \mathbb{R}$ in an interval containing 0, with $\hat{\psi}^G_{\mathrm{CCA},l} = \sum_{j=1}^{K^G} \hat{w}_j^{G,l} \hat{\psi}_j^G$ and $\hat{\psi}^F_{\mathrm{CCA},l} = \sum_{j=1}^{K^F} \hat{w}_j^{F,l} \hat{\psi}_j^F$, where $\{\hat{\psi}_j^G\}$ and $\{\hat{\psi}_j^F\}$ are the estimated geometric and functional PC component functions, while $\hat{w}_j^{G,l}$ and $\hat{w}_j^{F,l}$ denote the jth element of $\hat{\mathbf{w}}^{G,l}$ and $\hat{\mathbf{w}}^{F,l}$

respectively. An example of such a visualization is shown in Figures 4.17-4.18, for the real application.

4.3.5 Choice of the hyper-parameters

In the proposed models, various hyper-parameters have to be chosen. In particular, in the geometric registration step in Section 4.3.1, we have to choose the regularization weighting parameter λ. The regularization weighting parameter, in our analysis, does not play a large role. In fact, if the surfaces were noisy reconstructions, its choice would have been more delicate. However, in practice, the surfaces are extracted from a regularized segmentation process of 3D images, and thus are smooth. For this reason, the regularization weighting parameter λ, in the geometric registration, is chosen to be small.

As previously mentioned, $\mathcal{V}$ is in practice a RKHS. Important to the registration problem is the choice of $\sigma_{\mathcal{V}}$, the size of the kernel of the RKHS $\mathcal{V}$ (see Appendix B.1). In fact a RKHS with a large kernel size $\sigma_{\mathcal{V}}$ is able to better capture large deformations (e.g. size differences), while under-fitting local differences. A RKHS with a small kernel size has an opposite behaviour. Following the approach of Bruveris et al. (2012), we take a sum of two Gaussian kernels, which allows the space $\mathcal{V}$ to account for both large and small deformations.

The functional registration has also a regularization weighting parameter λ, which determines how slowly the algorithm approaches an optimal solution. As in Kneip and Ramsay (2008), after some experimentation, we choose the value λ that achieves a smooth variation on functional PC functions, obtained from the functional variability analysis, between each iteration. To determine the number of iterations needed, we examine the eigenvalue plots (scree plots) to determine when stability of these plots has been reached in a analogous manner to Kneip and Ramsay (2008). Finally, the regularization weighting parameter of the fPCA algorithm applied to the functions, has been chosen by K-fold cross-validation, with $K = 5$, details of which can be found in Chapter 2.

On a more general note, choosing the hyper-parameters of the registration algorithms, in a data-driven fashion, is admittedly a very difficult problem and it has been very little explored in the current literature, even in simpler situations such as for functions on the real line. The above only represents one possible method of choosing them, which appears to work well in our application, although further work would be needed for very different settings.

4.4 Analysis of a synthetic dataset

In this section, we validate the estimation framework introduced in Section 4.3, by performing a study on a dataset generated from Model 4.1. We thus proceed with defining the unknown quantities of such model. We will not use different notation for the theoretical objects and their respective computer representations, unless necessary.

Thus, we denote with $\mathcal{M}_0$, the template temporal lobe shown in Figure 4.6. We set

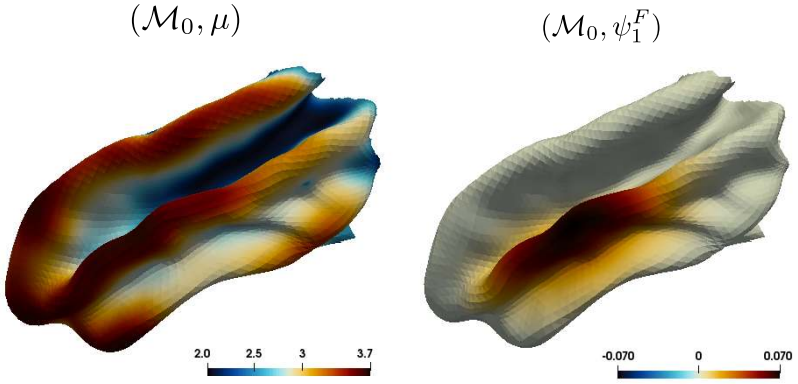


Figure 4.6 On the left, a template of the temporal lobe $\mathcal{M}_0$ with an associated cortical thickness map μ. On the right, the function ψ_1^F used to generate subject-specific functional variability.

the deformation operator φ to be the diffeomorphic deformation operator introduced in Section 4.2.4. We then choose two orthonormal vector fields $\psi_1^G, \psi_2^G \in \mathcal{V}$, visualized in Figure 4.7 as the deformations $\varphi_{\pm c\psi_1^G}, \varphi_{\pm c\psi_2^G}$ applied to the template $\mathcal{M}_0$, where $c \in \mathbb{R}$ is a constant regulating the norm of the two orthonormal vector fields, for visualization purposes. The vector field ψ_1^G encodes a change in the length of the temporal lobe, while the vector field ψ_2^G encodes a change in the size of temporal lobe.

We set the mean function $\mu \in L^2(\mathcal{M}_0)$, to be the thickness maps in Figure 4.7, which is a sharpened version of the cross-sectional average thickness of 100 real subjects. Note that despite it being computed from real data, this plays the role of an unknown quantity of the model. Moreover, we introduce localized functional variability through the single mode of variation $\psi_1^F \in L^2(\mathcal{M}_0)$, this also visualized in Figure 4.6.

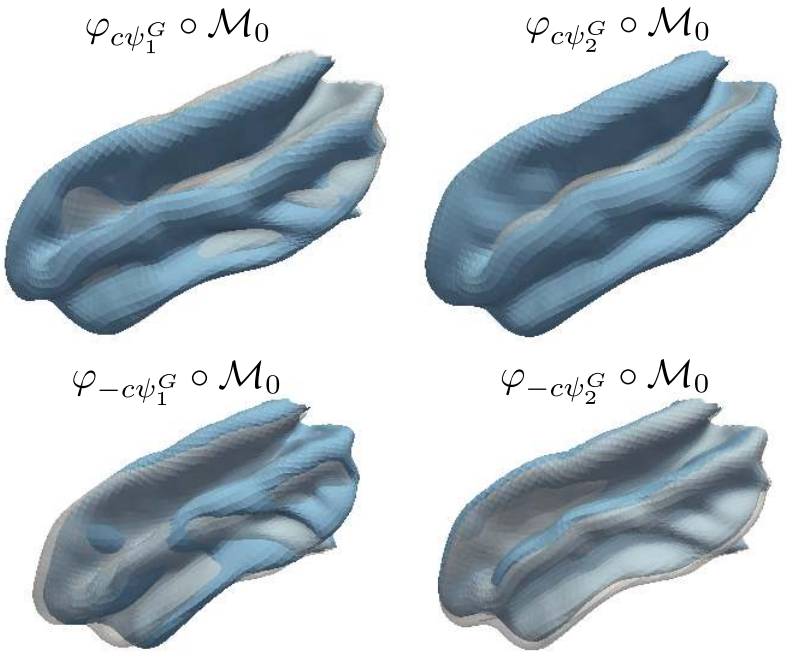


Figure 4.7 From left to right, first and second geometric modes of variation of the generated FoSs, here visualized as $\varphi_{\pm c\psi_1^G} \circ \mathcal{M}_0, \varphi_{\pm c\psi_2^G} \circ \mathcal{M}_0$, where $c \in \mathbb{R}$ is a constant regulating the magnitude for visualization purposes.

We then generate $n = 50$ FoSs $(\mathcal{M}_1, Y_1), \ldots, (\mathcal{M}_n, Y_n)$ by

$$\begin{cases} \mathcal{M}_i &= \varphi_{a_{i1}\psi_1^G + a_{i2}\psi_2^G} \circ \mathcal{M}_0, \\ X_i &= \mu + \delta a_{i2}\psi_1^F, \\ Y_i &= X_i \circ \varphi^{-1}_{a_{i1}\psi_1^G + a_{i2}\psi_2^G}, \end{cases} \tag{4.17}$$

where a_{i1}, a_{i2} are independent random variables distributed as $a_{il} \sim N(0, \sigma_l^2)$, with $\sigma_1 = 15$ and $\sigma_2 = 10$. The constant $\delta = 0.1$ determines the scale that relates variations in the functional term $\delta a_{i2}\psi_1^F$ and variations in the geometric term $a_{i2}\psi_2^G$. Finally, normally distributed noise with variance $\sigma = 0.3$, is added to each node of the mesh where the function is observed. The generative model proposed here is a simplistic implementation of the one proposed in Model 4.1, with $v_i = a_{i1}\psi_1^G + a_{i2}\psi_2^G$ and $Z_i = a_{i2}\psi_1^F$.

The generative model (4.17) seeks to reproduce a situation where the FoSs have two modes of geometric variation. The first one is a mode of variation which is not correlated with a variation in the functions. The second one, which encodes the size of the temporal temporal lobe, has an effect on the function, formalized with a linear relation between

the scores of the second geometric mode of variation ψ_2^G and the scores of the functional mode of variation ψ_1^F. The generated FoSs are such that larger temporal lobe have larger cortical thickness in proximity of the central gyrus of the cerebral cortex, independently of the first geometric mode of variation. We hope to recover this relation through the approximation pipeline introduced in Section 4.3.

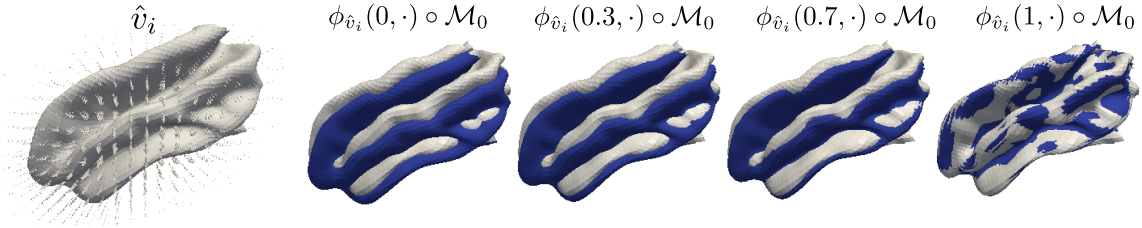


Figure 4.8 On the left, the template $\mathcal{M}_0$ with an estimated vector field $\hat{v}_i \in \mathcal{V}$ generating the diffeomorphic deformation $\varphi_{\hat{v}_i}$ that registers the template to the ith subject surface. Next, the evolution of the flow generating the diffeomorphic deformations $\phi_{\hat{v}_i}(t,\cdot)$ through the ODE (4.3), which registers the template to the target surface at time $t = 1$.

Figure 4.9 Two vector fields estimated from the functional registration algorithm, generating, for two different subjects, the flow which aligns two different functions to the cross-sectional mean function.

In particular, we perform non-landmarked diffeomorphic registration of the template to the single surfaces, resulting in the estimated vector fields $\{\hat{v}_i : i = 1,\dots,n\}$. The ith vector field $\hat{v}_i$ is such that $\varphi_{\hat{v}_i}\circ\mathcal{M}_0$ resembles the geometry $\mathcal{M}_i$ of the ith FoS, with φ the diffeomorphic deformation operator. In Figure 4.8, we show an estimated vector field $\hat{v}_i \in \mathcal{V}$ and the ODE's (4.3) flow $\phi_{\hat{v}_i}(t,\cdot)$, generated from the estimated vector field, which deforms the template to match the target.

The estimated diffeomorphic deformations $\{\varphi_{\hat{v}_i} = \phi_{\hat{v}_i}(1,\cdot)\}$ are then used to transport the functions $\{Y_i\}$ on the template surface, thus leading to the estimates $\{\hat{X}_i\}$. Subsequently, the cross-sectional mean map of $\{\hat{X}_i\}$ is computed and each function $\hat{X}_i$ is iteratively registered to it through the functional registration algorithm presented in Section 4.3.2. In

Figure 4.9, we show the template surface, with the tangential vector fields that generate the deformations that align two different functions to the cross-sectional mean function.

At each iteration of the functional registration algorithm, the cross-sectional mean and the first 2 functional PCs, from the functionally aligned versions of $\{\hat{X}_i\}$, are computed. The results are shown in Figure 4.10. We can notice that while the cross-sectional mean does not change from iteration to iteration of the functional registration algorithm, the estimates of the PC functions do. In particular, the first PC function is supposed to capture ψ_1^F. However, where no functional registration is applied, the first estimated PC component is a mix of the ψ_1^F and fictitious variability due to misalignment, while the second PC function is a flat and corrupted version of ψ_1^F. After only one iteration of the functional registration algorithm, the estimated first PC function starts resembling the shape of ψ_1^F, shifting the misalignment component to the second PC function. With the subsequent iteration the first estimated PC function becomes a sharper estimation of ψ_1^F, while the misalignment component disappears also from the second component, in favour of a flat PC function, which is a regularized PC function of the noise added to the functions.

Subsequently, we perform fPCA on the estimated vector fields $\{\hat{v}_i\}$ representing the overall deformation, due to both geometric and functional registration. In Figure 4.11 we show the estimated main modes of variation before the functional registration has been applied. By comparison with Figure 4.7, we can see that the first two PCs capture the main geometric modes of variations introduced in the generative process of the FoSs. The estimated geometric PC function do not change, in a visible manner, from iteration to iteration of the functional registration algorithm, because the functional registration brings only small deformations.

We finally plot, in Figure 4.12, the scores associated to the PCs describing the geometric variability and those describing the functional variability, for the estimated quantity without functional registration and after seven iterations of the functional registration. Note that without performing functional registration, not only is the first functional mode of variation a spurious version of the true underlying component, but this is also correlated to the geometric mode of variations, which might lead to misleading conclusions. Functional registration removes from the first PC the misalignment effect, bringing to light the true underlying linear dependence between the functional mode of variation and the second geometric mode of variation.

In practice, the above procedure is particularly useful if the discovered PCs have biological interpretations. However, in practice, the discovered PCs tend to vary, depending

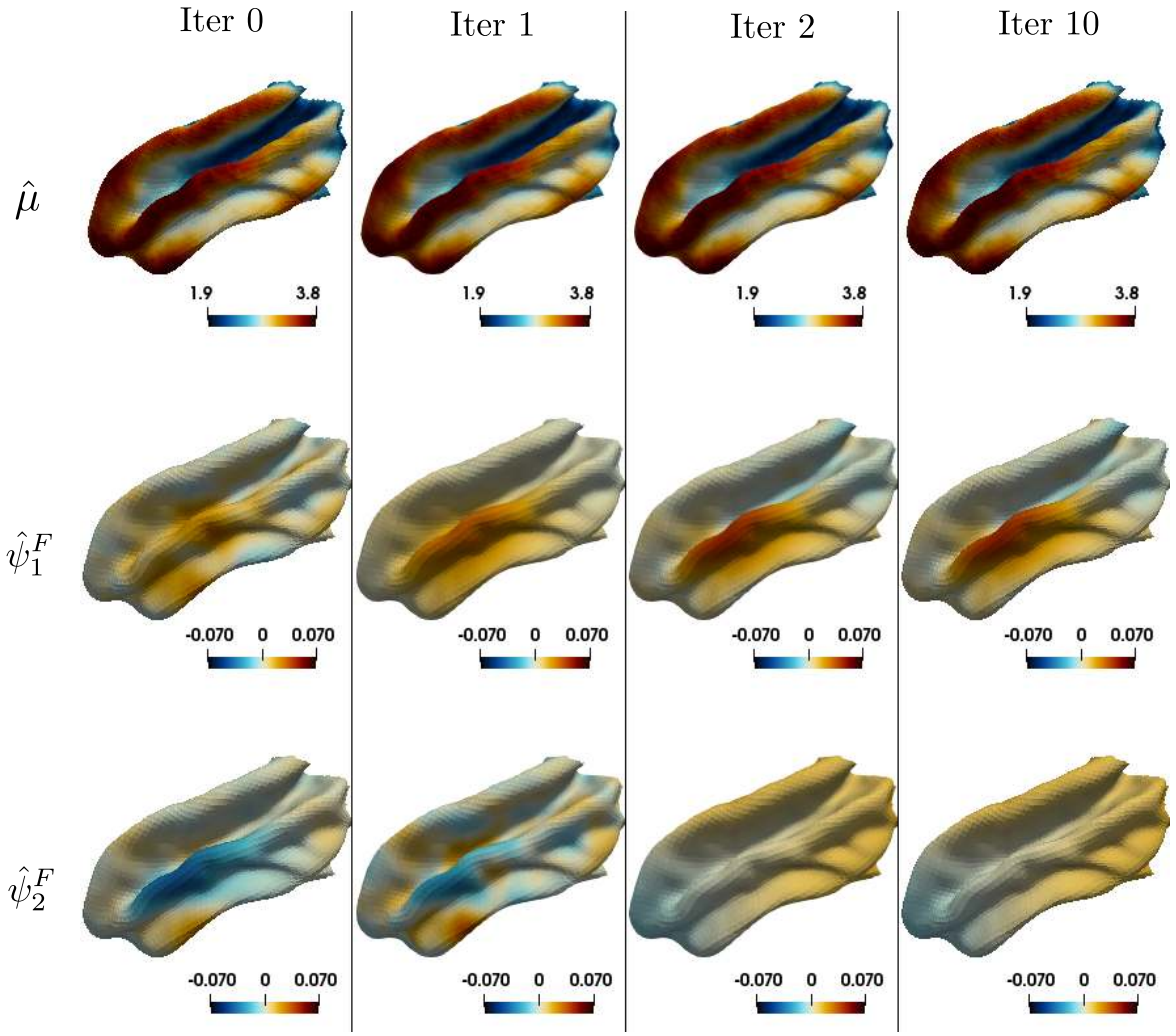


Figure 4.10 From left to right, the mean and first two functional PC functions estimates of the functions, computed after 0, 1, 2 and 10 iterations of the functional registration algorithm.

for instance on the pre-registration method applied or on the scalar product adopted to impose orthogonality between the PC functions. For these reasons, if the aim is to study the relation between geometry and function, we advocate CCA (see Section 4.3.4). We perform a CCA on the estimated scores of the geometric and functional variability, after seven iterations of the functional registration algorithm. In detail, we construct a $n \times 3$ matrix $\mathbb{X}^F$ with the scores of the first three components of the fPCA applied to the functions. Moreover, we construct a $n \times 5$ matrix $\mathbb{X}^G$ with the scores of the first five components of the fPCA applied to the deformations. The lth canonical correlation component is the pair of vectors $\hat{\mathbf{w}}^{F,l} \in \mathbb{R}^3$ and $\hat{\mathbf{w}}^{G,l} \in \mathbb{R}^5$. The resulting main mode of co-variation

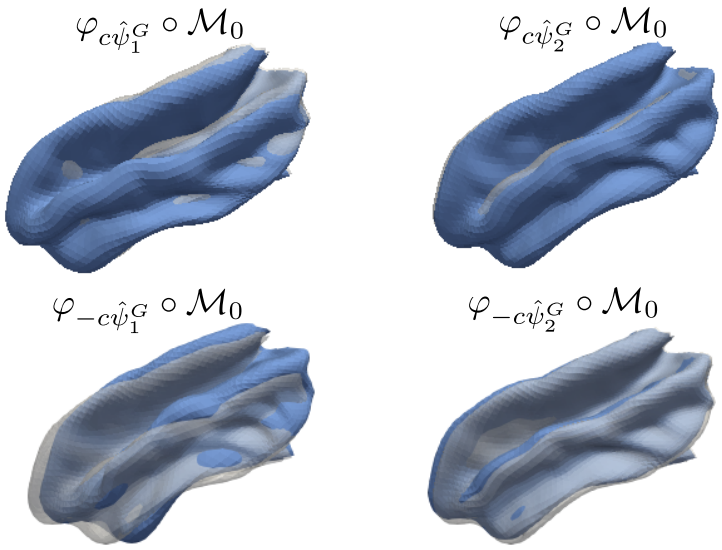


Figure 4.11 From left to right, the first two geometric PC functions computed on the space of initial vector fields. These are visualized as $\varphi_{\pm c\hat{\psi}_j^G} \circ \mathcal{M}_0$, where $\hat{\psi}_j^G$ is the estimated jth geometric PC function.

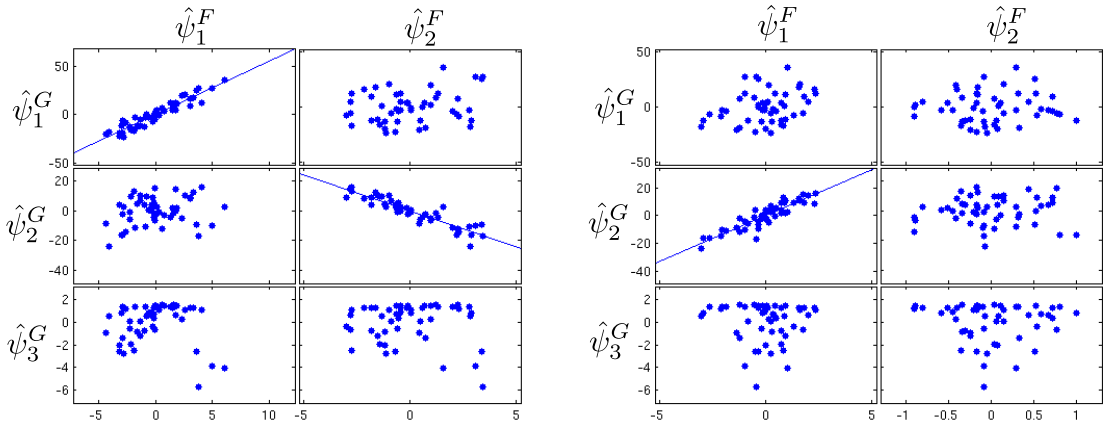


Figure 4.12 From left to right, scatter plots of the scores obtained from the fPCA on the function and the geometric fPCA, respectively without and with functional registration. After functional registration, these show only the linear dependence imposed between the first PC function on the functions ψ_1^F and the second geometric PC function ψ_2^G. Without functional registration, also the spurious PC function, due to misalignment, is correlated with the first geometric PC function.

$(\hat{\mathbf{w}}^G, \hat{\mathbf{w}}^F) = (\hat{\mathbf{w}}^{G,1}, \hat{\mathbf{w}}^{F,1})$ is visualized in Figure 4.13 as

$$\begin{cases} \mathcal{M}_{\text{CCA}} = \varphi_{c\,\hat{\psi}^G_{\text{CCA}}} \circ \mathcal{M}_0, \\ Y_{\text{CCA}} = c\,\hat{\psi}^F_{\text{CCA}} \circ \varphi^{-1}\, c\,\hat{\psi}^G_{\text{CCA}}, \end{cases}$$

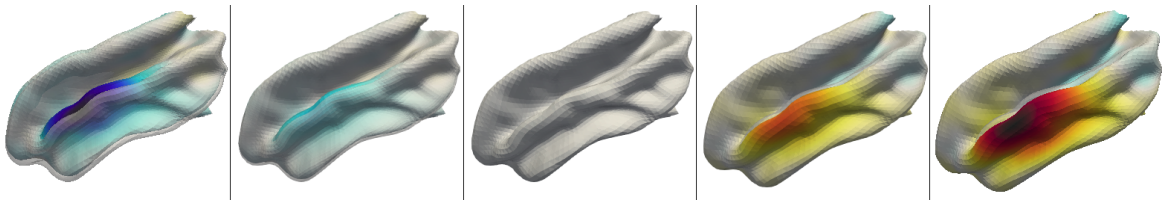


Figure 4.13 First main mode of co-variation of geometric and functional components of the CCA analysis, representing the most correlated linear combinations of the first five geometric modes of variation and first three functional modes of variation. From left to right, this is visualized by plotting the FoS in (4.16) for a sequence of constants c.

with $\hat{\psi}^G_{\text{CCA}} = \sum_{j=1}^3 \hat{w}^G_j \hat{\psi}^G_j$ and $\hat{\psi}^F_{\text{CCA}} = \sum_{j=1}^5 \hat{w}^F_j \hat{\psi}^F_j$, where $(\hat{\psi}^G_j)$ and $(\hat{\psi}^F_j)$ are the estimated functional and geometric PC functions. $c \in \mathbb{R}$ is a constant varied for visualization purposes in an interval containing 0. As we can see in Figure 4.13, the dependence between the magnitude and the thickening of the function is captured.

Moreover, we test for the statistical significance of the obtained modes of co-variation. Specifically, we test the hypotheses

$$H_0^l : \hat{\rho}_1 \neq 0, \hat{\rho}_2 \neq 0, \ldots, \ldots, \hat{\rho}_l \neq 0, \hat{\rho}_{l+1} = \ldots = 0, \tag{4.18}$$

with $\hat{\rho}_l = \text{corr}(\mathbb{X}^G \hat{\mathbf{w}}^{G,l}, \mathbb{X}^F \hat{\mathbf{w}}^{F,l})$. According to a likelihood ratio test, with the Bartlett χ^2 approximation of the test statistic distribution (see Johnson and Wichern, 2002, Chapter 10.6), only the sample correlation between the first canonical correlation variables, i.e. $\mathbb{X}^G \hat{\mathbf{w}}^{G,1}$ and $\mathbb{X}^F \hat{\mathbf{w}}^{F,1}$, is significantly different from zero (p-value $5e-19$), while for $l = 2,3$ we get p-values 0.7759 and 0.9587 respectively.

4.5 Application

The publicly available data set considered in this work has been collected by the Human Connectome Project Consortium (HCP, Van Essen et al., 2012), with the ultimate goal of elucidating the understanding of the brain functions, by collecting multi-modal neuroimaging 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. A minimal preprocessing pipeline have been applied to the dataset (Glasser et al., 2013).

4.5.1 Preprocessing

A 3D structural MRI scan has been performed for each individual, returning a 3D image describing the internal structure of the brain. A slice of the 3D image is shown on the left panel of Figure 4.1. The cerebral cortex is the outermost layer of the brain, mostly consisting of neuronal cell bodies. With automatic segmentation techniques, it is possible to separate the cerebral cortex from the other parts of the brain. Subsequently the two surfaces enclosing the cerebral cortex can be computed. The inner surface represents the boundary between the cerebral cortex and the white matter (second panel in Figure 4.1), while the outer surface corresponds to the boundary between the cerebral cortex and the cerebrospinal fluid (fourth panel in Figure 4.1). The resolution of the MRI images (0.7 mm isotropic, in this study) and the effectiveness of the segmentation algorithm determine the level of details at which such surfaces can be reconstructed.

The geometry of the cerebral cortex is generally represented by the mid-thickness surface, which is the surface fitting the middle-points of the inner and outer surfaces, an example of which is shown on the third panel of Figure 4.1. Thus, it is natural to expect the resulting surfaces to have wider sulci and thinner gyri than what we could observe from a picture of the brain surface. Moreover, the mid-thickness surface can be equipped with a function representing the thickness of the cerebral cortex, computed from the inner and outer surface, as described in Fischl and Dale (2000). A comparison of the various methods for the cerebral cortex thickness estimation can be found in Lerch and Evans (2005). In Figure 1, we show the reconstructed (mid-thickness) surfaces of the left hemisphere of 3 different subjects with the associated cerebral cortex thickness maps. Each surface is represented by a 32K nodes mesh, and at each node of the mesh an evaluation of the function is available.

The mid-thickness surfaces of the collected cohort are pre-registered to the Conte69 template, on the left in Figure 4.14, through a surface-based registration algorithm driven by geometric features that describe measures of cortical shape folding, such as sulcal depth or local curvature (Fischl et al., 1999; Glasser et al., 2013). Registrations are ensured to be one-to-one by introducing, in the objective function, a term related to the metric distortion of the registration maps and a term that enforces the positivity of the signed areas of the triangles on the surfaces (see Fischl et al., 1999, for details). Such a procedure defines a one-to-one correspondence between the 32K nodes of the template and the 32K nodes of each of the mid-thickness surfaces, which can be regarded as a set of 32K landmarks.

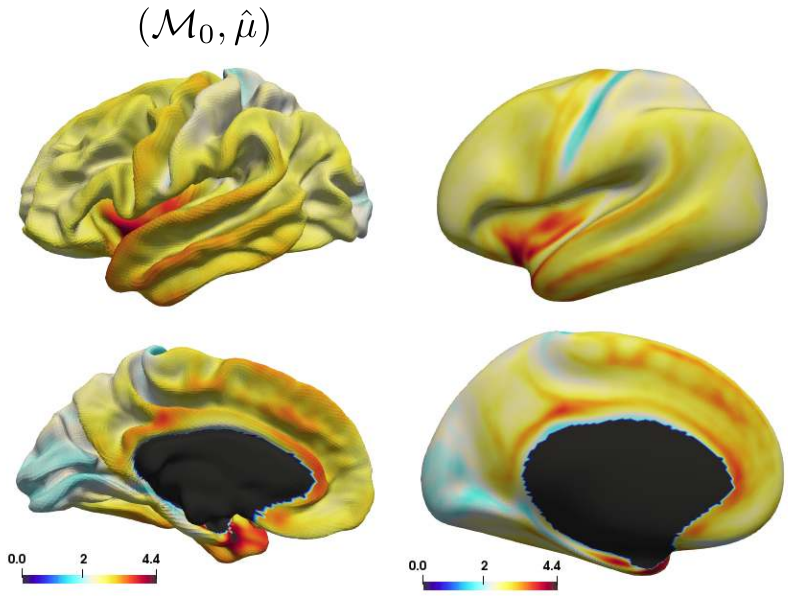


Figure 4.14 On the left, the Conte69 template, used as a template surface for the registration of the individual surfaces. This is equipped with the cross-sectional mean function computed post-geometric registration. On the right, the cross-sectional mean function visualized on an inflated version of the template.

4.5.2 Analysis

The relation between geometric features of the brain has raised great interest in the recent years, since it can potentially help us understand the principles underlying brain development. Classically, these studies have been confined to correlation studies on variables summarizing particular geometric features. For instance, in Im et al. (2008), for each subject, the average cortical volume and absolute mean curvature, among other, are computed. This set of real variables are then compared to the average cerebral cortex thickness computed on each subject. Moreover, a more localized analysis is performed by parcellating each cortical surface in the 4 lobes. Subsequently, the analysis is performed independently on each of lobe. However, there are two limitations of such approach. Firstly, the description of the geometric properties through summary statistics is in general incomplete. Secondly, the parcellation of the cortical surfaces determines a priori which areas of the cortical surface can have a different behaviour.

The fact that a geometric registration has already been performed on the HCP data, without relying on the diffeomorphic registration framework in Section 4.3.1, is not in contrast with the proposed analysis. In fact, diffeomorphic-like constraints can be imposed in many different ways when it comes to the estimation of registration maps. However,

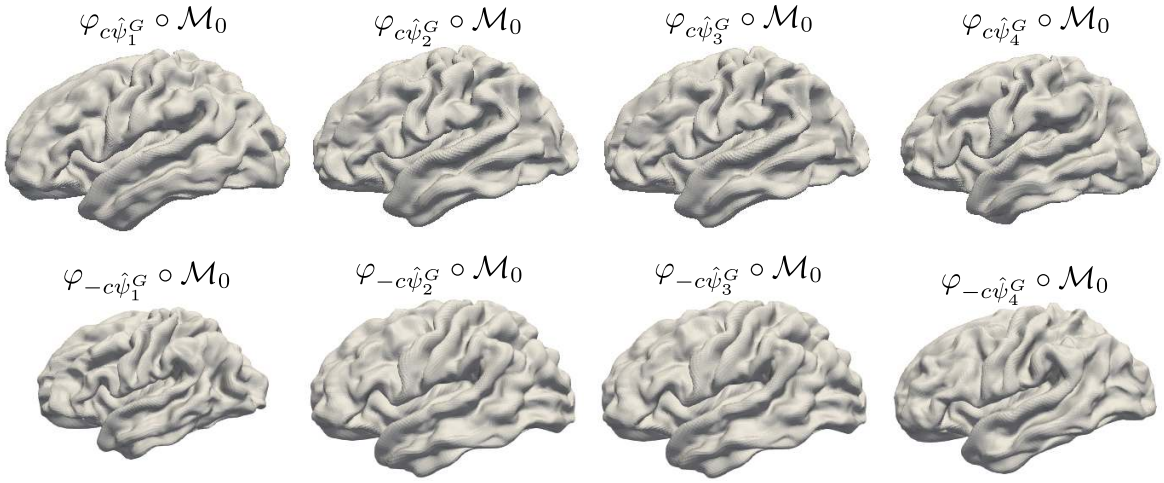


Figure 4.15 From left to right, the first four geometric PC functions, computed on the space of initial vector fields. These are visualized as $\varphi_{\pm c\hat{\psi}_j^G} \circ \mathcal{M}_0$, where $\hat{\psi}_j^G$ is the jth geometric PC function.

if the aim is the estimation of a low-dimensional subspace of the diffeomorphic space, these alternative approaches cannot be extended to this more general problem. For this reason, we use the landmarks defined by the pre-processing geometric registration to estimate the vector fields that represent such registrations and then perform fPCA on the estimated vector fields, as described in Section 4.3.3. The estimated first four geometric PCs are shown in Figure 4.15. Not surprisingly, they are mostly related to the size of the brain or the size of sub-parts of the brain.

We then perform fPCA on the functions registered on the Conte69 template. The results are shown in the top two rows of Figure 4.16. Subsequently, we perform functional registration of the functions on the Conte69 template and recompute the functional modes of variation at each iteration. In the bottom two rows of Figure 4.16 we show the PC functions after 2 iterations of the functional registration algorithm.

We finally perform a CCA on the first eight geometric and functional PC functions scores. The resulting first two main modes of co-variation, the only significant ones from the likelihood ratio test (4.18), are shown in Figures 4.17-4.18. From the left to the right panel of Figure 4.17, we can see the presence of a correlation between a decrease in thickness in the frontal lobe and an increase in size of the entire brain, while in the temporal lobe, an increase in thickness seems associated to an increase in size of the entire brain. Moreover, in the second main mode of co-variation a more localized phenomenon is captured in proximity of the high average cortical thickness area on the lateral sulcus (see Figure 4.14), where an association between an increase in the cortical thickness and

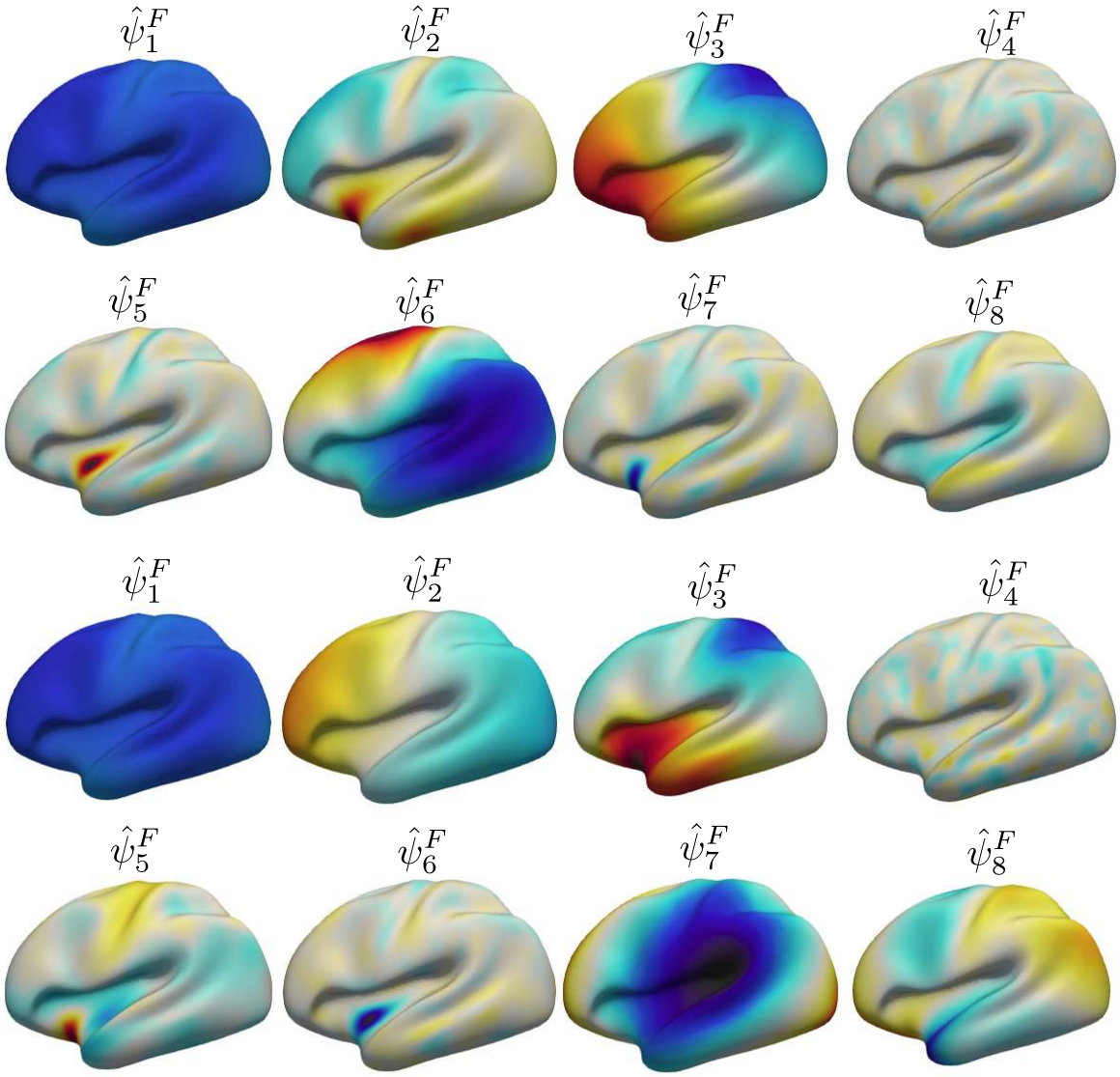


Figure 4.16 Results of the fPCA on the functions. On the top two rows, the first eight functional modes of variation computed without performing functional registration. On the bottom two rows, the first eight functional modes of variation computed after performing functional registration. These are shown on an inflated version of the template to easy their visualization.

an increase in the size of the brain is suggested. Note that such local effect would have not been captured by a study confined to study individually each lobe of the brain, and such relation would have probably been ascribed to the entire lobe containing that area.

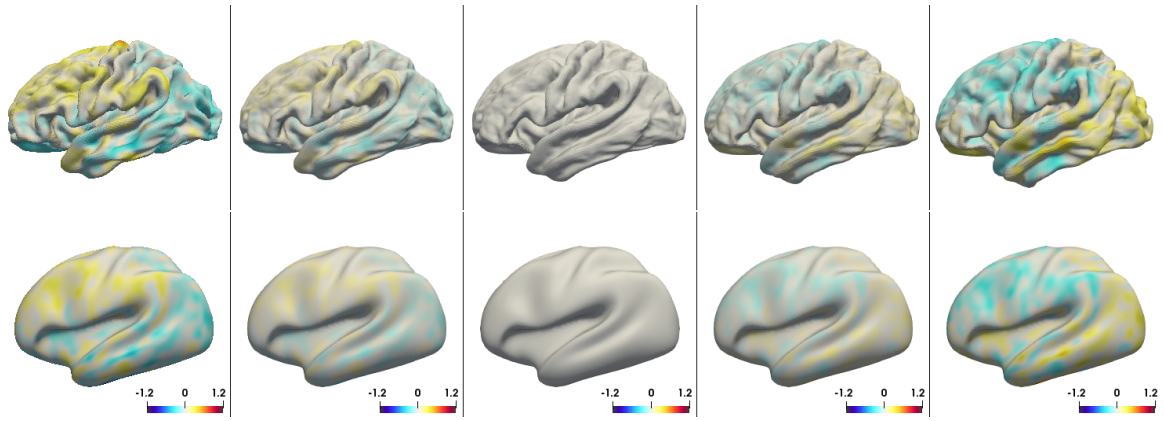


Figure 4.17 A representation of the first main mode of co-variation of the geometric and functional components of the CCA analysis, representing the most correlated linear combinations of the first eight geometric and functional PC functions. From left to right, this is visualized by plotting the FoS in (4.16) for a sequence of constants c.

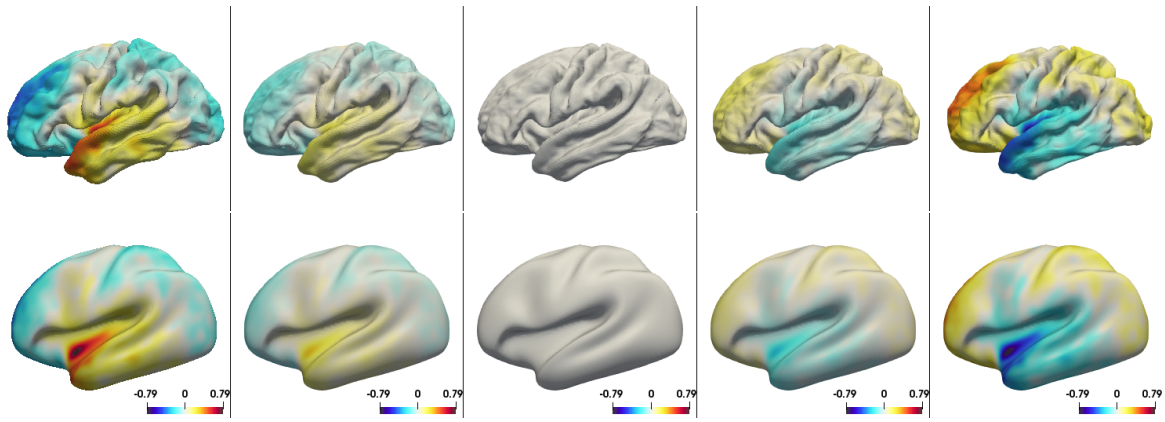


Figure 4.18 A representation of the second main mode of co-variation of the geometric and functional components of the CCA analysis, representing the second most correlated linear combinations of the first eight geometric and functional PC functions. From left to right, this is visualized by plotting the FoS in (4.16) for a sequence of constants c.

4.6 Discussion

In this chapter, motivated by the analysis of neuroimaging data, we introduce a framework for the analysis of FoSs. In particular, a statistical model describing the phenomenon is formulated, and the estimators of the unknown quantities of the model are introduced. The construction of such estimators is complicated by the necessity of the resulting estimates to lie in the non-linear subspace of 'sensible' solutions, here taken to be deformations. Moreover, in such high dimensional setting, it is fundamental for the estimator to incorporate prior information on the geometry and the smoothness of the data, achieved

by regularizing the estimates through differential operators. Motivated by simulation studies, we address the necessity of using the functional information to achieve a better registration, a well known fact in FDA, by introducing a novel diffeomorphic registration algorithm for functional data on a two-dimensional manifold.

While the main motivation of this work was taken from a neuroimaging application into assessing the inherent variabilities of cortical thickness, the methodology has wider applications in medical imaging as a whole, where FoS appear in cardiovascular (e.g. Huang et al. (2016)), muskuloskeletal (e.g. Treece and Gee (2015)) and many other imaging areas. More generally, this methodology is an example of the use of differential operators as regularisers in statistics, a field where not only statistical but also numerical techniques are needed to facilitate solutions.

A future interesting aspect is the exploration of the applicability of the Optimal Transport framework to the registration problem, as suggested in Panaretos and Zemel (2016) in a discrete context, and its links with the diffeomorphic deformation framework. This is of potential interest in the surface registration framework, where we usually lack physical models that can describe the phenomena, and thus a 'least action' approach could well be effective.

Chapter 5

Conclusions and Future work

The main contributions of this thesis can be summarized as follows. We introduce a Functional Principal Component Analysis (FPCA) model for functional data whose domain is a two-dimensional manifold. We adopt a Regularized-Estimation approach and incorporate additional geometric information on the supporting manifold by means of a differential operator penalty, namely the Laplace-Beltrami operator. We apply the proposed model to connectivity maps computed from resting-state fMRI. We then formulate a FPCA model for functional data in an inverse problems setting. This can be applied to either first-order or second-order functional data, allowing us to study statistical objects that are covariance functions. We apply the proposed model to study changes in functional connectivity across time and subjects. Finally, we present a novel joint geometric-functional model for functional data that take values on subject-specific supporting manifold domains. We model geometric variations by means of diffeomorphic flows. We apply the model to study associations between variations of brain shape and thickness of the cerebral cortex.

The framework developed naturally leads to many interesting questions for future research. Here, we list a few of them.

- From a neuroimaging prospective, it is necessary to move towards the integration of multiple sources of information. In this thesis, we have mainly focused on shape and functional modeling, due to the methodological challenges they present. However, a more comprehensive investigation could include, for instance, task-fMRI data, family structure across subjects, lifestyle, demographic and psychometric measures. A promising approach for combing these heterogeneous set of variables is Joint and Individual Variation Explained (Lock et al., 2013).

- In our work, we have mainly focused on healthy individuals. It is certainly of interest to extend the proposed approaches to understand associations between a disease outcome and geometric-functional variations. This requires the generalization of the proposed methodology to a regression framework. The proposed FPCA models can often be interpreted as regularized tensor decompositions, it is thus natural to explore the idea of embedding these decompositions in a tensor regression framework (Zhou et al., 2013).

- From a statistical prospective, the developed models will require theoretical analysis of the convergence properties of the proposed estimators. Due to the differential operator regularizers, this requires tools from both partial differential equations theory and statistical theory.

Appendix A

FPCA over manifolds

A.1 Simulation on the sphere

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

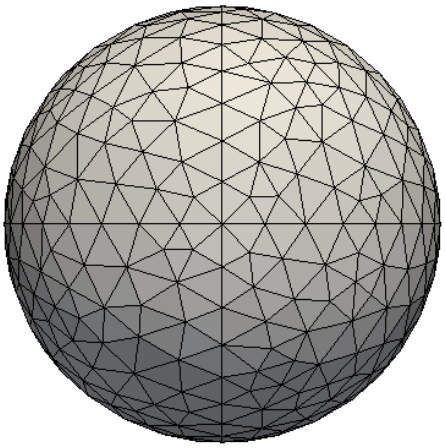


Figure A.1 The triangulated surface approximating the sphere with 488 points.

A.1.1 Noisy obervations

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

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

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 $i \in \{1,2\}$ and $\int_{\mathcal{M}} v_i v_k = 0$ for $i \neq k$ with $i,k \in \{1,2\}$. The PC functions are plotted in Figure A.2. The functions x_i are sampled at locations coinciding with the nodes of the mesh in Figure A.1. 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.

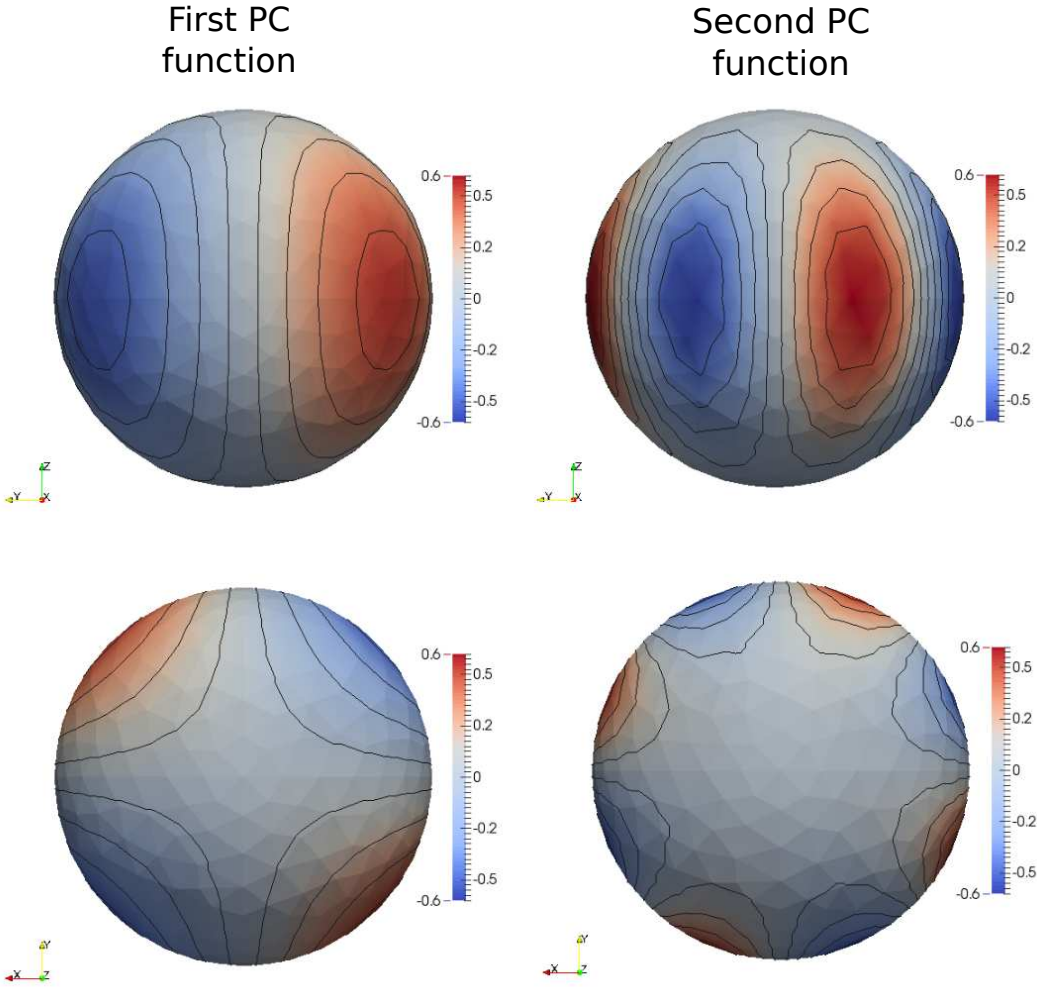


Figure A.2 From the left to the right, two views of the true first and second PC functions.

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 presmoothing 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 A.3.

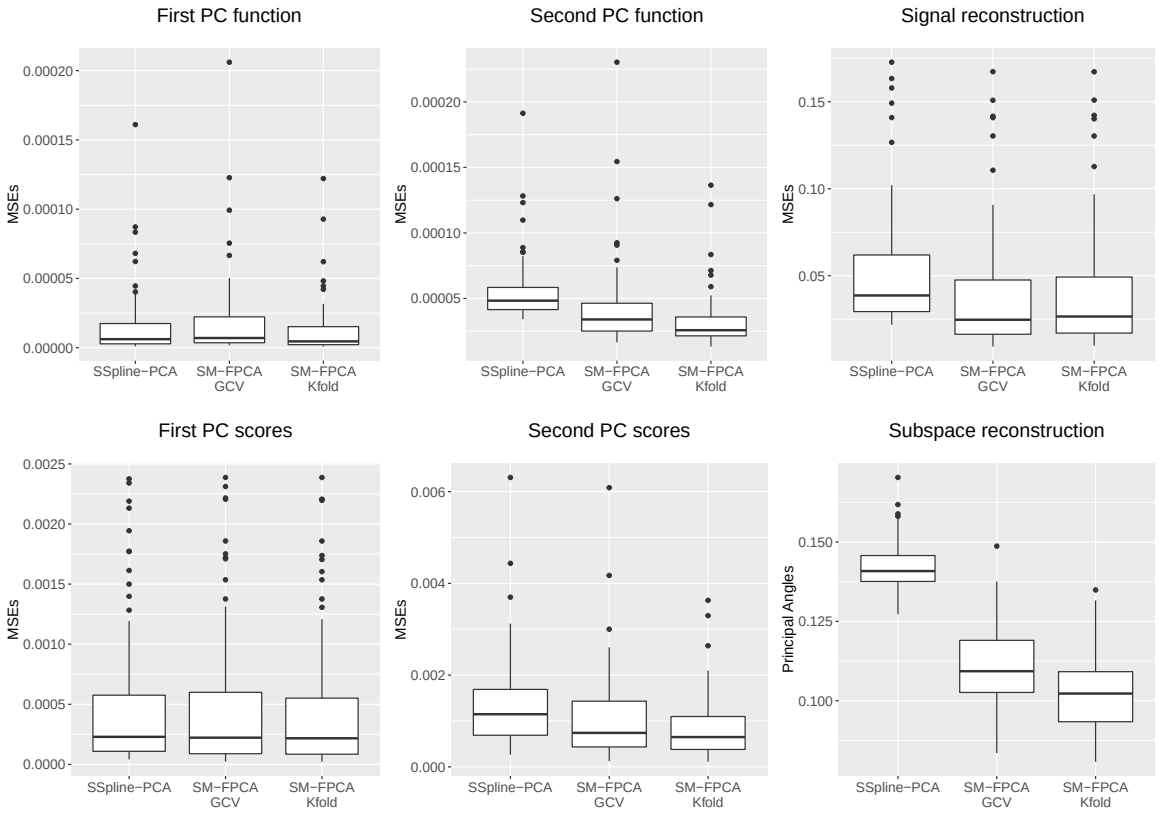


Figure A.3 Boxplots summarizing the performance of SSpline-PCA and SM-FPCA. For the SM-FPCA both GCV and K-fold has been applied for the selection of the smoothing parameter.

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 based 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. Also the MSE on the signal reconstructions, as well as the measure based on the principal angle between the space spanned respectively by $\{v_i\}_{i=1,2}$ and the estimated PC functions $\{\hat{v}_i\}_{i=1,2}$, emphasize the good performance of the introduced algorithm.

A.1.2 Spatial mismatching

In this section we complement the set of simulations in the noisy setting by designing a simple simulation that shows how SM-FPCA behaves when a spatial mismatching effect is introduced. In the motivating application to neuroimaging data, spatial mismatching is introduced by the shape registration algorithm. In this simulation, we consider a spherical domain $\mathcal{M}_{\mathcal{T}}$ and reproduce this spatial mismatching effect, that results in misalignment of the signals on this domain, by including a subject specific shift (in spherical coordinates) of the first PC function. In detail, we generate $n = 50$ smooth functions $x_1, \dots, x_{50}$ on $\mathcal{M}_{\mathcal{T}}$ by

$$x_i = u_{i1} v_{i1}, \qquad i = 1, \dots, n \tag{A.1}$$

where u_{i1} represent the PC scores, generated independently and distributed as $u_{i1} \sim N(0, \sigma^2)$ with $\sigma = 4$, and the functions v_{i1} represent misaligned realization of the PC function v_1. Specifically, we parametrize v_1 in spherical coordinates (θ, ϕ) and set $v_{i1}(\theta, \phi) = v_1(\theta + \theta_i, \phi + \phi_i)$, with θ_i and ϕ_i generated independently with a discrete uniform distribution on the set $\{0, 0.4\}$. In Figure A.4 we show v_{i1} for the four possible realizations of shifting coefficients (θ_i, ϕ_i).

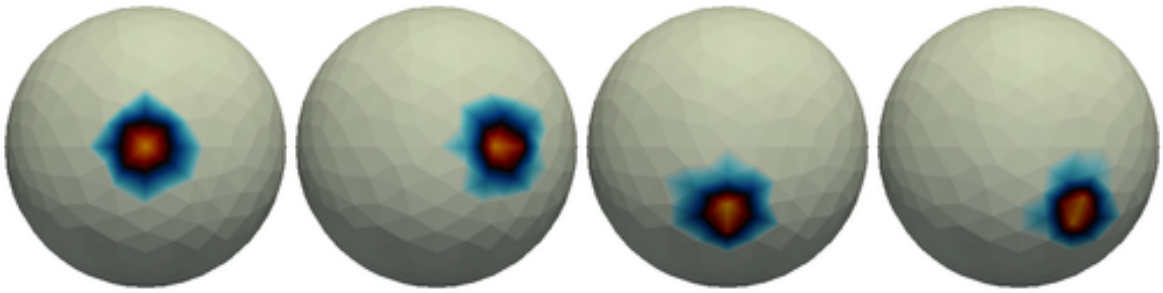


Figure A.4 A plot of the four different realizations of the misaligned PC function v_{i1}.

The interest is to recover the structure of the only PC function v_1, from the misaligned realizations $\{x_i\}_{i=1,\dots,n}$, ignoring the effects introduced by the shifts. To consider purely the misalignment's effect, we do not add noise to the sampled functions x_i. In fact, while the

benefits of SM-FPCA in the noisy setting have already been extensively demonstrated, we aim now at considering separately the effect of a spatial mismatching on the sampled functions from the effect of the presence of noise. Pre-smoothing of the signal, as performed in SSPline-PCA, is thus unnecessary, and we compare directly MV-PCA to SM-FPCA. In fact, as already mentioned, the proposed SM-FPCA model incorporates the smoothing penalty in a more parsimonious way than the pre-smoothing approach, allowing a direct control of the smoothness of the estimated PC function. We would like to show that SM-FPCA, combined with a cross-validation approach for the choice of the smoothing parameter λ, might help removing artefacts introduced by the spatial mismatching.

In Figure A.5 we show the estimates computed with MV-PCA, SM-FPCA GCV and SM-FPCA K-fold ($K = 5$) for four different datasets generated as in (A.1). In the top row we show a situation where the PC function estimated with MV-PCA shows a satisfactory result. In this case also SM-FPCA GCV and SM-FPCA K-fold show a similar behavior. However, in the bottom three rows the estimates of the PC function computed with MV-PCA and SM-FPCA GCV show some artefacts introduced by the misalignment, while the estimate computed with SM-FPCA K-fold better preserves the shape of the PC function, renouncing however to spatial localization. The results obtained with SM-FPCA K-fold suggest to interpret the phenomena at a more macroscopical scale, due to the high local variability introduced by the spatial mismatching.

The different behavior of SM-FPCA, when the smoothing parameter is chosen by GCV with respect to K-fold cross-validation, can be explained by the fact that this first approach concerns with the choice of λ only in the regression step (2.7), where the choice of λ is only driven by the presence of noise on the vector $\mathbf{X}^T\mathbf{u}$. On the contrary, SM-FPCA K-fold is based on a direct comparison of the PC function estimated on the training and validation sets, obtained partitioning the dataset.

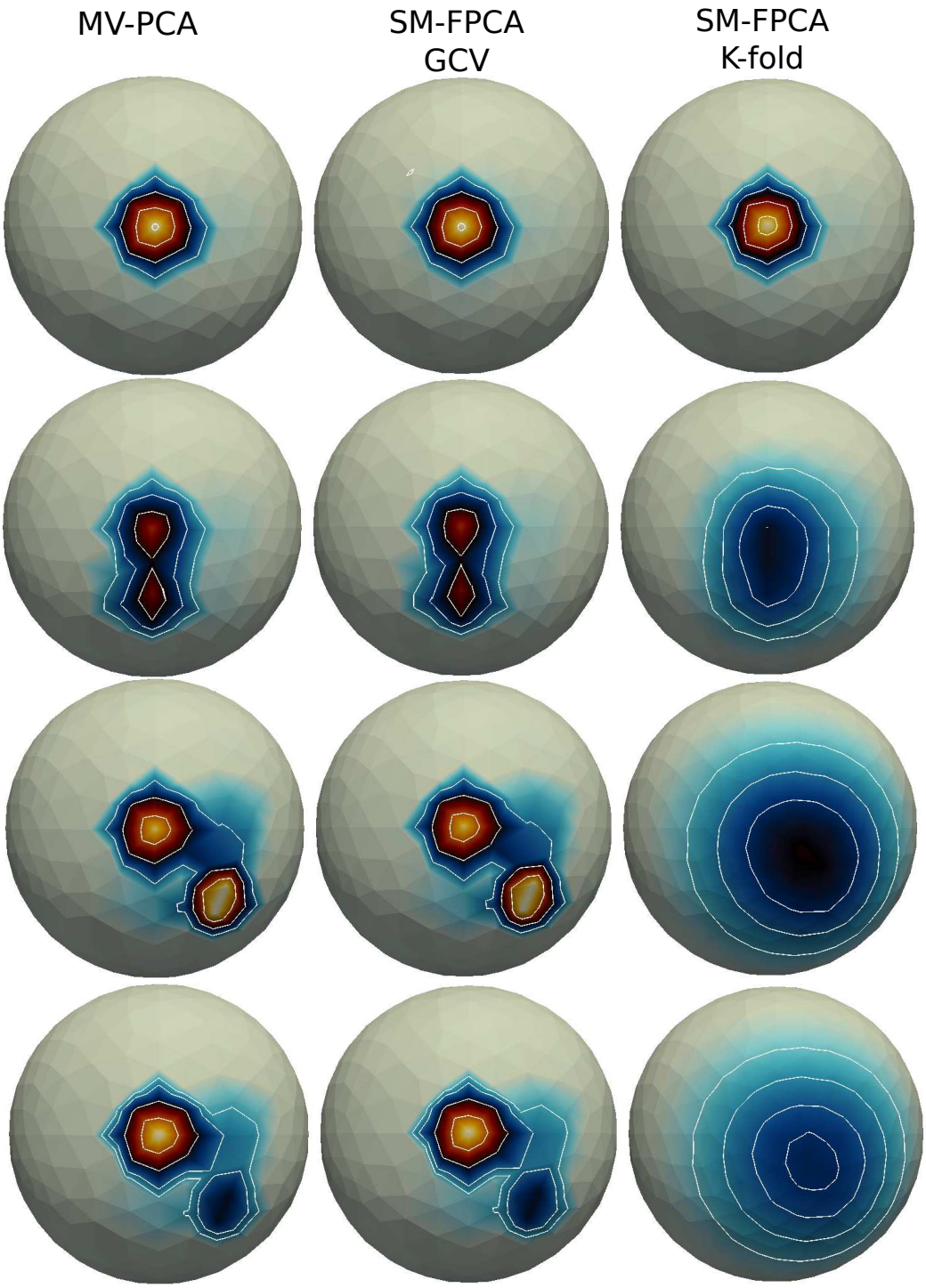


Figure A.5 From top to bottom, plot of the estimates computed on 4 different generated datasets. From left to right, plot of the estimate of the first PC function computed respectively with MV-PCA, SM-FPCA GCV and SM-FPCA K-fold.

Appendix B

FoSs

B.1 Geometric Diffeomorphic Registration

Here we present the algorithmic details of the estimation framework introduced in Section 4.3.

The space of smooth vector fields $\mathcal{V}$, in the geometric registration model (4.5), is usually constructed as a RKHS (Miller et al., 2015). In detail, let $K_{\mathcal{V}} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ be a bounded symmetric positive definite function. $K_{\mathcal{V}}$ is usually referred to as the kernel of $\mathcal{V}$ and a typical choice for it is the Gaussian isotropic kernel, i.e. $K_{\mathcal{V}}(x,y) = \exp(-\|x-y\|_2^2/(2\sigma_{\mathcal{V}}^2))\mathrm{Id}_{3\times 3}$, with $\mathrm{Id}_{3\times 3}$ denoting a 3×3 identity matrix and $\sigma_{\mathcal{V}}$ reflecting the rigidity of the space. Define the pre-Hilbert space $\mathcal{V}_0 = \mathrm{span}\{K_{\mathcal{V}}(\cdot,x)\omega | x \in \mathbb{R}^3, \omega \in \mathbb{R}^3\}$. Given $f, g \in \mathcal{V}_0$ we can write them as $f = \sum_{i=1}^N K_{\mathcal{V}}(\cdot,x_i)\omega_i$ and $g = \sum_{i=1}^N K_{\mathcal{V}}(\cdot,y_i)z_i$. We thus define the inner product between f and g to be $\langle f,g\rangle_{\mathcal{V}} = \sum_{i,j=1}^N \omega_i^T K_{\mathcal{V}}(x_i,y_j)z_j$. The space $(\mathcal{V},\langle\cdot,\cdot\rangle_{\mathcal{V}})$, defined as the closure of $\mathcal{V}_0$, is a (Reproducing Kernel) Hilbert space of smooth vector fields.

For modeling purposes, the time-variant vector-field v_t, introduced in Section 4.2.4, is assumed to be of the form (see e.g. Vaillant et al., 2004)

$$v_t(\cdot) = \sum_{k=1}^{k_g} K_{\mathcal{V}}(\phi_v(t,c_k),\cdot)\alpha_k(t), \tag{B.1}$$

for a set of control points $\{c_k : k = 1,\dots,k_g\} \subset \mathbb{R}^3$ and the auxiliary variables $\{\alpha_k(t) : \mathbb{R} \to \mathbb{R}^3\}$ called momenta of the deformation. The control points $\{c_k\}$ are commonly chosen to be the nodes of the triangulated representation of the surface to be deformed. ϕ_v denotes the solution of the ODE (4.3) given the time-variant vector field $\{v_t : t \in [0,1]\}$. The associated

deformation energy is defined to be

$$\int_0^1 \|v_t\|_{\mathcal{V}}^2 = \int_0^1 \sum_{k,l=1}^{k_g} \alpha_k(t)^T K_{\mathcal{V}}(\phi_v(t,c_k),\phi_v(t,c_l))\alpha_l(t). \tag{B.2}$$

Denoting with ∇_1 the gradient with respect to the first variable, the vector field v_t generating geodesics, with respect to the energy term $\int_0^1 \|v_t\|_{\mathcal{V}}^2$, can be characterized as the solution of the coupled ODE system, known as the EPDiff equation (Miller et al., 2015)

$$\begin{cases} \frac{\partial c_k(t)}{dt} & = \sum_{l=1}^{k_g} K_{\mathcal{V}}(c_k(t),c_l(t))\alpha_l(t) \\ \frac{\partial \alpha_k(t)}{dt} & = -\frac{1}{2}\left(\sum_{l=1}^{k_g} \nabla_1 K_{\mathcal{V}}(c_k(t),c_l(t))\alpha_l(t)\right)^T \alpha_k(t), \end{cases} \tag{B.3}$$

for a set of initial conditions $\{\alpha_k = \alpha_k(0)\} \subset \mathbb{R}^3$, parameterizing the initial vector field v_0. This means that the energy minimizing vector fields, generating diffeomorphisms, can be determined by (B.3) and fully controlled by the initial vector field

$$v_0(\cdot) = \sum_{k=1}^{k_g} K_{\mathcal{V}}(\cdot,c_k)\alpha_k,$$

parametrized in terms of the initial momentum vector $\{\alpha_k : k = 1,\dots,k_g\}$. Moreover, along a geodesic path the instantaneous deformation energies $\|v_t\|_{\mathcal{V}}$ are constant, meaning that the total deformation energy $\int_0^1 \|v_t\|_{\mathcal{V}}^2 dt$ can be equivalently represented by the initial deformation energy $\|v_0\|_{\mathcal{V}}^2 = \sum_{k,l} \alpha_k K_{\mathcal{V}}(c_k,c_l)\alpha_l$.

Thanks to the finite dimensional representation underlying the element of the RKHS $\mathcal{V}$, the minimization of (4.5) can be cast in a finite dimensional setting and can be approached, for instance, with a gradient descent algorithm on the initial momentum vector parametrizing the initial velocity field (see, among others, Vaillant et al., 2004).

The MATLAB toolkit fshapesTk (https://github.com/fshapes/fshapesTk) offers an implementation of the described geometric registration algorithm, and its extension to the fshapes framework (Charlier et al., 2017).

B.2 Registration of Functional Data on a two-dimensional manifold

Functional Registration Model

Let now $M, F : \mathscr{M}_0 \to \mathbb{R}$ be respectively a 'moving' and 'fixed' image. We recall here the objective function of the functional registration model (4.9), in terms of M and F:

$$E_{\mathscr{M}_0}(u) = \sum_{j=1}^{S} \left(F(p_j) - M \circ s \circ \phi_u(p_j)\right)^2 + \lambda \|\Delta_{BL} u\|^2_{L^2(T\mathscr{M}_0)}, \tag{B.4}$$

with $\{p_j, j = 1, \dots, S\} \subset \mathscr{M}_0$ the set of control points on the template and ϕ_u denoting the solution of the ODE (4.8) for the vector field u, at time $t = 1$.

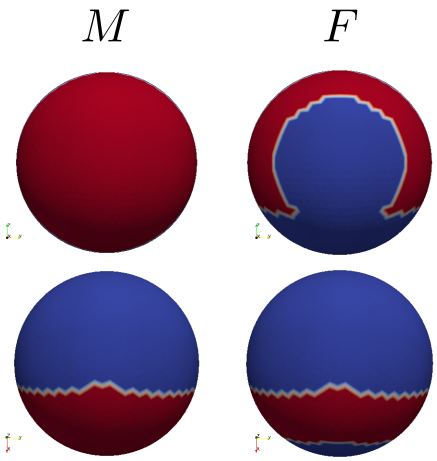


Figure B.1 On the left two views of a semi-circle image on the unit sphere, representing the moving image M, while on the right two views of a C-shaped image on the unitary sphere, representing the fixed image F.

The term $M \circ s \circ \phi_u$ is then linearized with respect to u. This results in the approximation

$$M \circ s \circ \phi_u \approx M \circ s + L_u,$$

where L_u is a first order approximation of $M \circ s \circ \phi_u - M \circ s$. In practice L_u is chosen to be of the form

$$L_u(p) = g_p(J(p), u(p)), \qquad p \in \mathscr{M}_0,$$

with $J(p) \in T_p\mathscr{M}_0$ for all $p \in \mathscr{M}_0$. Two classical choices for J, in the planar case, are $J = \nabla_D(M \circ s)$ and $J = \frac{1}{2}(\nabla_D(M \circ s) + \nabla_D(F))$ (Vercauteren et al., 2009), where ∇_D denotes

a discrete estimate of the gradient. Plugging the linearized term in (B.4) we obtain the objective function

$$E_{\mathscr{M}_0}(u) = \sum_{j=1}^{S} \left(F(p_j) - (M \circ s)(p_j) - g_{p_j}(J(p_j), u(p_j))\right)^2 + \lambda \|\Delta_{BL} u\|^2_{L^2(T\mathscr{M}_0)}. \tag{B.5}$$

The minimization of (B.4) can be achieved by iteratively minimizing the associated problem (B.5) and updating the current deformation s with $s \rightarrow s \circ \phi_u$, with ϕ_u denoting the solution of the ODE (4.8) at time $t = 1$.

B.2.1 Problem reformulation

To minimize the objective function in (B.5) we opt for a finite elements discretization approach. Finite element discretization has been previously applied to the discretization of FDA problems on manifolds, for instance, in Ettinger et al. (2016) and Lila et al. (2016). Here, we extend the methodology to the estimation of smooth vector fields on a generic two-dimensional manifold. To this end, we first reformulate the minimization of (B.5) in terms of the Euler-Lagrange equation associated to this minima problem.

Define now the space of smooth vector fields on the template to be $\mathscr{W} = \{u \in L^2(T\mathscr{M}_0) | \Delta_{BL} u \in L^2(T\mathscr{M}_0)\}$. Let the vector field $u \in \mathscr{W}$, in the functional (B.5), be perturbed by an ε amount along the arbitrary direction $\varphi \in \mathscr{W}$. The minimization problem is reformulated by imposing the Gateaux derivative $\partial_\varphi E_{\mathscr{M}_0}(u)$ of the energy functional to be 0 for all $\varphi \in \mathscr{W}$.

This leads to the problem reformulation: find $\hat{u} \in \mathscr{W}$ such that

$$\begin{aligned} \sum_{j=1}^{S} g_{p_j}(\varphi(p_j), J(p_j)) g_{p_j}(\hat{u}(p_j), J(p_j)) + \lambda \langle \Delta_{BL}\varphi, \Delta_{BL}\hat{u} \rangle_{L^2} = \\ \sum_{j=1}^{S} g_{p_j}(\varphi(p_j), J(p_j))(F(p_j) - M \circ s(p_j)) \end{aligned} \tag{B.6}$$

for every $\varphi \in \mathscr{W}$. Moreover, equation (B.6) can be reformulated as the problem of finding $(\hat{f}, \hat{h}) \in \mathscr{W} \times L^2(T\mathscr{M}_0)$ that satisfies

$$\begin{cases} \langle \Delta_{BL}\hat{u}, v \rangle_{L^2} - \langle \hat{h}, v \rangle_{L^2} = 0 \\ \lambda \langle \hat{h}, \Delta_{BL}\varphi \rangle_{L^2} + \sum_{j=1}^{S} g_{p_j}(\varphi(p_j), J(p_j)) g_{p_j}(\hat{u}(p_j), J(p_j)) = \sum_{j=1}^{S} g_{p_j}(\varphi(p_j), J(p_j))(F(p_j) - M \circ s(p_j)) \end{cases} \tag{B.7}$$

for all $(\varphi, v) \in \mathscr{W} \times L^2(T\mathscr{M}_0)$. In this last reformulation, we have introduced the auxiliary function $\hat{h}$, which has been imposed to be equal, in a weak sense, to $\Delta_{BL}\hat{u}$. Now, asking the auxiliary function v and the test functions φ to be such that $v, \varphi \in \mathscr{W}^1 = \{u \in L^2(T\mathscr{M}_0) | \nabla u \in L^2(T^*\mathscr{M}_0 \otimes T\mathscr{M}_0)\}$, and by exploiting the definition of the Bochner-Laplacian, we can rewrite the problem only in terms of the connection operator ∇, and consequently be able to formulate it in a finite dimensional space involving only first order polynomials, as done in equation (B.9).

B.2.2 Vector Finite Element discretization

Here we introduce a linear finite element space for vector fields on a triangulated surface, where we seek for the discrete solution of the problem (B.7). To this end, consider the triangulated surface $\mathscr{M}_0^c alT$, approximated representation of the manifold $\mathscr{M}_0$. $\mathscr{M}_0^c alT$ is not a smooth surface, so it is not even clear what the tangent space on a vertex of the triangulation is. For this reason, we use elements of computer graphics to define an interpolation basis on the triangulated surface, as done for instance in Knöppel et al. (2013); Zhang et al. (2006).

Let now $\xi_1, \ldots, \xi_K$ be the vertices of $\mathscr{M}_0^c alT$. For each vertex ξ_k consider the subset of $\mathscr{M}_0^c alT$ composed by the triangles adjacent to ξ_k, that we call here one-ring. Following the approach in Knöppel et al. (2013), the one-ring surface is idealized by normalizing the sum of the angles incident to the vertex ξ_k to add up to 2π, i.e. by 'flattening' the vertex and uniformly distributing that curvature to the flat triangles of the one-ring. To the vertex ξ_k they associate a unit vector basis (e_k^1, e_k^2) representing a reference orientation, so that an element of the tangent vector $u_k \in T_{\xi_k}\mathscr{M}_0^c alT$ will be represented by its coefficients $\mathbf{u}_k \in \mathbb{R}^2$ respect to the local basis. Then, an interpolation basis can be defined on the idealized one-ring of the vertex ξ_k by parallel transporting through geodesics (e_k^1, e_k^2) to the interior points of the one-ring and by scaling them with a piecewise linear function which takes value 1 on ξ_k and 0 one the other vertices of the one-ring (see Knöppel et al., 2013, for details).

What is important to this work is that the outlined procedure leads to a basis of K functions, whose kth function has support localized on the triangles adjacent to ξ_k, and that we denote here with the function $\boldsymbol{\psi}_k = (\psi_k^1, \psi_k^2)$, with ψ_k^1 and ψ_k^2 vector fields on $\mathscr{M}_0^c alT$. For this basis functions the FE matrices $\langle \boldsymbol{\psi}_k, \boldsymbol{\psi}_{k'} \rangle_{L^2}$ and $\langle \nabla\boldsymbol{\psi}_k, \nabla\boldsymbol{\psi}_{k'} \rangle_{L^2}$ are provided.

We can finally define the FE function space $\mathcal{W}_h$ to be

$$\mathcal{W}_h = \{u_h = \sum_{k=1}^{K} \boldsymbol{\psi}'_k \mathbf{u}_k | \mathbf{u}_k \in \mathbb{R}^2\}. \tag{B.8}$$

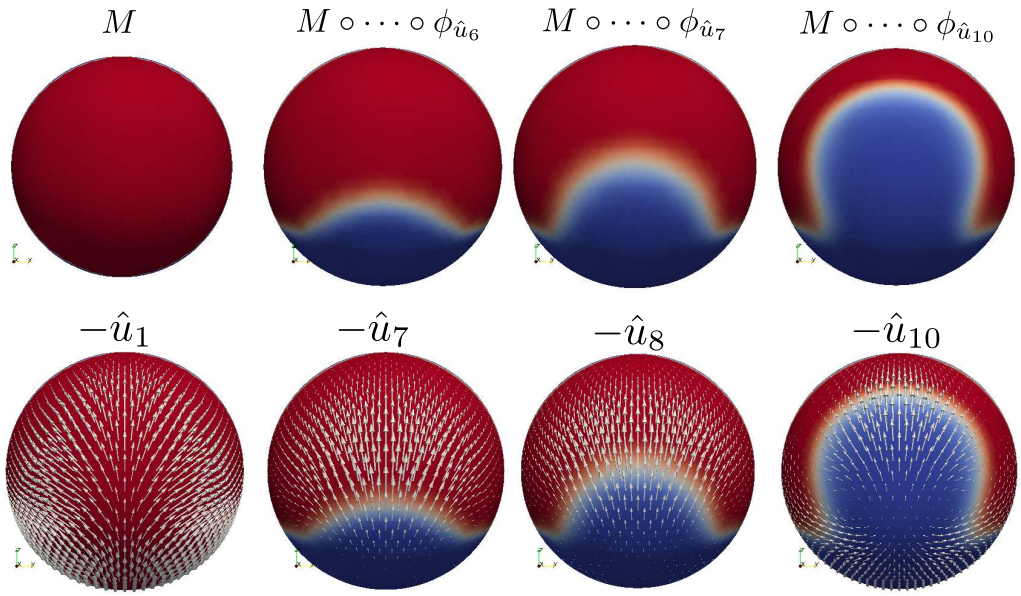


Figure B.2 From left to right, the estimated vector fields, and associated deformations of M, at 4 different iterations of the functional registration algorithm. The target is the C-shaped image F.

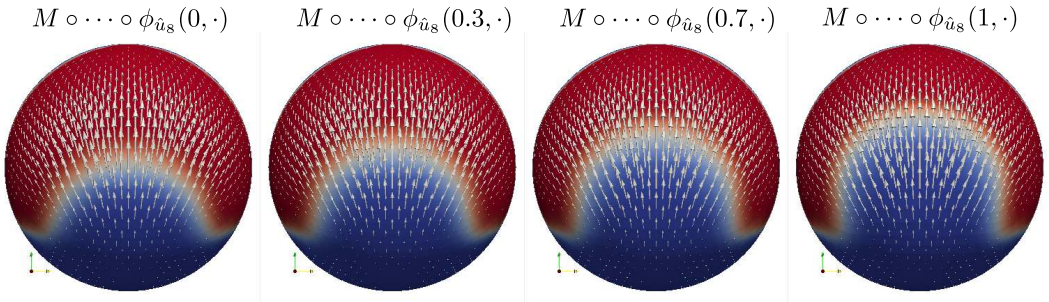


Figure B.3 From left to right, evolution of the flow through the ODE (4.8) for a fixed vector field. The vector field is obtained by the minimization of the linearized objective function (B.5) at the 8th iteration.

The solution in the restricted space $\mathcal{W}_h$ is finally given by the discrete approximations $\hat{u}_h, \hat{h}_h \in \mathcal{W}_h$, obtained by solving

$$\begin{cases} \langle \nabla \hat{u}_h, \nabla \varphi_h \rangle_{L^2} - \langle \hat{h}_h, \varphi_h \rangle_{L^2} = 0 \\ \lambda \langle \nabla \hat{h}_h, \nabla v_h \rangle_{L^2} + \sum_{j=1}^{S} g_{p_j}(v_h(p_j), J(p_j)) g_{p_j}(\hat{u}(p_j), J(p_j)) = \sum_{j=1}^{S} g_{p_j}(v_h(p_j), J(p_j))(F(p_j) - M \circ s(p_j)) \end{cases} \tag{B.9}$$

for all $\varphi_h, v_h \in \mathcal{W}_h$.

Exploiting the representation (B.8) of functions in $\mathcal{W}_h$ we can rewrite (B.9) as a linear system as follows. Let $\hat{\mathbf{u}}$ be a $2K$ vector obtained from the vectorization of the set coefficients $\{\mathbf{u}_i\}$. In the same way let $\hat{\mathbf{h}}$ be the vectorization of the coefficients of $\hat{h}_h$ in (B.9). Now, introduce the $2K \times S$ matrix Θ_1 and the $2K \times 2K$ matrix Θ_2, such that

$$\begin{aligned} \mathbf{v}' \Theta_1 \mathbf{z} &= \sum_{j=1}^{S} g_{p_j}(v_h(p_j), J(p_j))(F(p_j) - M \circ s(p_j)) \\ \mathbf{v}' \Theta_2 \hat{\mathbf{u}} &= \sum_{j=1}^{S} g_{p_j}(v_h(p_j), J(p_j)) g_{p_j}(\hat{u}(p_j), J(p_j)), \end{aligned}$$

with $\mathbf{z}$ the vector of length S such that its jth element is $(F(p_j) - M \circ s(p_j))$ and $\mathbf{v}$ the $2K$ vector obtained from the vectorization of the set coefficients of v_h. These sparse matrices are defined in Section B.2.3, together with the $2K \times 2K$ mass and stiffness matrices R_0 and R_1, such that

$$\begin{aligned} \hat{\mathbf{h}}' R_0 \boldsymbol{\varphi} &= \langle \hat{g}_h, \varphi_h \rangle_{L^2} \\ \hat{\mathbf{h}}' R_1 \mathbf{v} &= \langle \nabla \hat{g}_h, \nabla v_h \rangle_{L^2}, \end{aligned}$$

where $\boldsymbol{\varphi}$ is a $2K$ vector obtained from the vectorization of the set coefficients φ_h.

Finally, the coefficients $\hat{\mathbf{u}}, \hat{\mathbf{h}}$, of $\hat{u}_h, \hat{h}_h$ are given by the solution of the linear system

$$\begin{bmatrix} \Theta_2 & \lambda R_1 \\ \lambda R_1 & -\lambda R_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{u}} \\ \hat{\mathbf{h}} \end{bmatrix} = \begin{bmatrix} \Theta_1 \mathbf{z} \\ \mathbf{0} \end{bmatrix}, \tag{B.10}$$

where $\mathbf{0}$ is a $2K$ length zero-vector.

The coefficients $\{\hat{\mathbf{u}}_k\}$ extracted from their vectorization $\hat{\mathbf{u}}$ in (B.10) represent the approximated tangent vectors on the vertices $\{\xi_k\}$. They are then linearly interpolated to define a solution on $\mathcal{M}_0^c alT$. This linear piecewise solution on $\mathcal{M}_0^c alT$ is then used to generate a diffeomorphic transformation through the ODE (4.8), which is itself approxi-

mated with the Euler method. At each step of the Euler method the image of the solution is re-projected on $\mathscr{M}_0^c alT$. Finally, the current registration is updated by composition with the newly estimated deformation as $s \leftarrow s \circ \phi_{\hat{u}}$, where $\phi_{\hat{u}}$ denotes the solution of the time $t = 1$ given by the Euler method.

In Figure B.1, we show an example of a moving image M, which is a semicircle indicator function, and a fixed image F, which is a C-shaped indicator function. They both live on the same spherical domain. This example tries to replicate the C-shaped planar registration problem, where image registration algorithms are usually tested, as for instance done in Vercauteren et al. (2009). In Figure B.2 we show the vector fields estimated at four different iterations of the Algorithm 4. While in Figure B.3, for one particular iteration, we show the evolution of the flow generated by the ODE (4.8). In this specific example, the domain is chosen to be spherical for visualization purposes, however it can be any smooth two-dimensional manifold, as for instance, in Section 4.5. The performances of the algorithm, with these synthetic data, are excellent. In fact, only 12 iterations are necessary to register the semicircled indicator function to the C-shaped indicator function.

Finally, it could be argued that being the proposed approximation of the vector field $\hat{u}$ only piecewise linear, and not of higher regularity, this could lead to deformations that are not diffeomorphic. However, the use of reasonably fine triangulated meshes $\mathscr{M}_0^c alT$ should solve the problem. After all, in practice, even for higher regularity vector fields, the computer resolution of the ODE relies on a finite number of sampled values from the vector field, and thus on a non smooth vector field.

B.2.3 Finite element matrices

Assume, for simplicity, that the points $\{p_j\}$ coincide with the nodes $\{\xi_k : 1, \ldots, K\}$ of the mesh $\mathscr{M}_0^c alT$. The non-zero entries of the matrices Θ_1 and Θ_2 are

$$\{\Theta_1\}_{2k,k} = g_{\xi_k}(J(\xi_k), e_1^k),$$
$$\{\Theta_1\}_{2k+1,k} = g_{\xi_k}(J(\xi_k), e_2^k)$$

and

$$\{\Theta_2\}_{2k,2k} = g^2_{\xi_k}(J(\xi_k), e_1^k), \qquad \{\Theta_2\}_{2k,2k+1} = g_{\xi_k}(J(\xi_k), e_1^k) g_{\xi_k}(J(\xi_k), e_2^k),$$
$$\{\Theta_2\}_{2k+1,2k} = g_{\xi_k}(J(\xi_k), e_1^k) g_{\xi_k}(J(\xi_k), e_2^k), \qquad \{\Theta_2\}_{2k+1,2k+1} = g^2_{\xi_k}(J(\xi_k), e_2^k)$$

with the matrices indexed from zero and $k = 0, \dots, K-1$. The computation of the entries $g_{\xi_k}(J(\xi_k), e_1^k)$ can be performed by representing the tangent vectors $J(\xi_k)$ and e_1^k as vectors in $\mathbb{R}^3$ and computing the $\mathbb{R}^3$ Euclidean scalar product between them, as in fact the manifold $\mathscr{M}_0$, and its associated triangulated mesh $\mathscr{M}_0^c alT$, are embedded in $\mathbb{R}^3$. The entries of the $2K \times 2K$ matrices R_0 and R_1 in (B.9) are computed in (Knöppel et al., 2013, Section 6.1.1), for the purpose of computing eigen-vectors of the Bochner-Laplacian operator.

B.2.4 Boundary Conditions

The deformations generated by the functional registration algorithm are by definition constrained to be maps with their image on the template surface, since the ODE (4.8) is defined on the manifold itself. However, if the template is a manifold with a boundary, as in the simulations performed in Section 3.5, the vector might generate deformations that transport the functions outside the boundary. This can be avoided by imposing homogeneous Dirichlet boundary conditions on the estimated vector field. Dirichlet boundary conditions can be implemented in different ways. Here, we opt for applying them after the linear system (B.10) has been built. In particular given a boundary node k, we add a large constant M to the entries $2k, 2k$ and $2k+1, 2k+1$ of the left hand side matrix and set to 0 the entries $2k$ and $2k+1$ of the right hand side vector. As a consequence, the vector fields estimated from the modified linear system will smoothly vanish as approaching the boundary.

B.3 Further Simulations

As previously mentioned, the functional registration algorithm introduced in Section 4.3.2 is not the only option to account for functional information in the registration process. Here we compare our methodology to the joint functional and geometric registration algorithm proposed in Charon and Trouvé (2014), where the shape similarity functional (4.7) is extended to include a functional similarity term.

Suppose now that the template mesh $\mathscr{M}_0^c alT$ is equipped with a functional object $\mu^{\mathcal{T}} : \mathscr{M}_0^c alT \to \mathbb{R}$, which in first instance can be the cross-sectional mean of the functions $\hat{X}_i$ estimated after the geometric registration described in Section 4.3.1. We briefly recall the notation in Section 4.3.1, introduced to define (4.7). We define $K_{\mathcal{Z}} : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^{3\times 3}$ to be a Gaussian isotropic kernel of variance $\sigma_{\mathcal{Z}}^2$, i.e. $K_{\mathcal{Z}}(x, y) = \exp(-\|x-y\|_2^2/(2\sigma_{\mathcal{Z}}^2))\mathrm{Id}_{3\times 3}$,

with $\mathrm{Id}_{3\times3}$ denoting a 3×3 identity matrix. Additionally, we introduce a scalar Gaussian kernel $K_{\mathscr{F}}:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ of the form $K_{\mathscr{F}}(x,y)=\exp(-(x-y)^2/(2\sigma_{\mathscr{F}}^2))$.

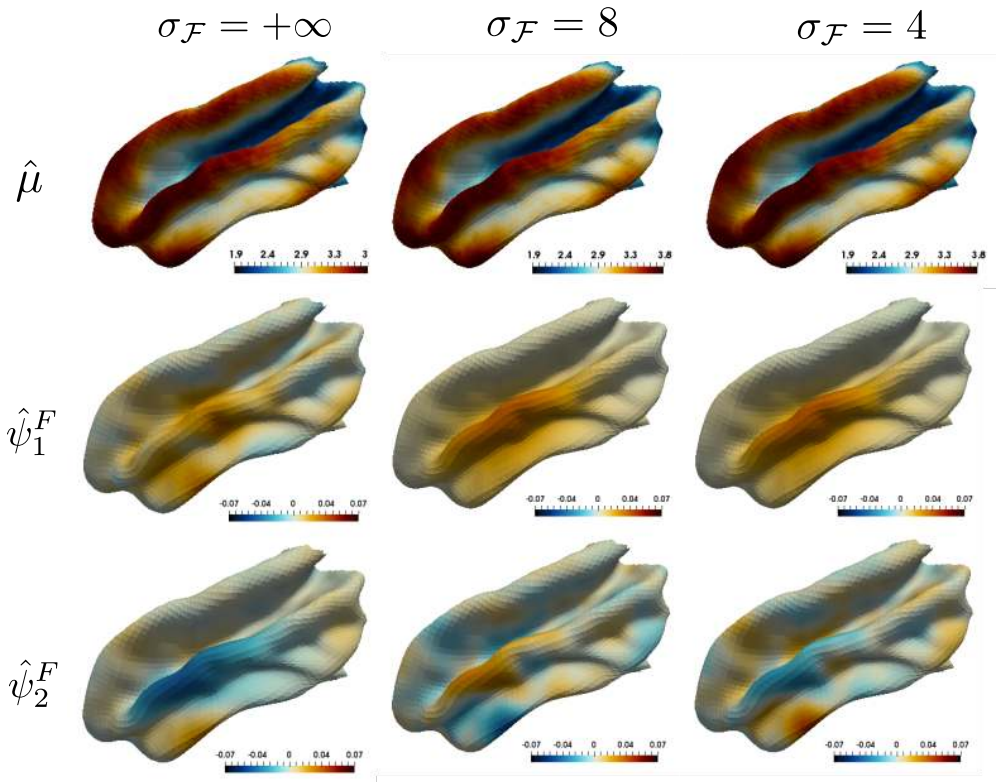


Figure B.4 From left to right, the mean and first two functional PC functions estimates of $\{\hat{X}_i\}$, estimated by using the registration maps computed by solving (4.5) with the extended matching function in (B.11), for different choices of $\sigma_{\mathscr{F}}$.

Moreover, we denote with $c(l)$ and $\eta(l)$, respectively, the center point and the normal vector of the lth triangle of the mesh $\varphi_{v_i}\circ\mathscr{M}_0^{\mathscr{T}}$. We denote with $c_i(q)$ and $\eta_i(q)$, respectively, the center point and the normal vector of the qth triangle of the mesh $\mathscr{M}_i^{\mathscr{T}}$. Additionally, we introduce $y(l)$, denoting the functional value $\mu^{\mathscr{T}}$, associated to the mesh $\varphi_{v_i}\circ\mathscr{M}_0^{\mathscr{T}}$, at the center point of the lth triangle. We denote with $y_i(q)$ the functional value associated to the ith FoSs at the center point of the qth triangle of the mesh $\mathscr{M}_i$.

Let the triangles of the mesh $\varphi_{v_i}\circ\mathscr{M}_0^{\mathscr{T}}$ be indexed by l and g and the triangles in $\mathscr{M}_i^{\mathscr{T}}$ be indexed by q and r. The shape similarity functional (4.7) can be extended to include

functional informations as follows (Charon and Trouvé, 2014).

$$
\begin{aligned}
D^2\big((\varphi_{v_i}\circ\mathscr{M}_0^{\mathscr{T}},\mu^{\mathscr{T}}\circ\varphi_{v_i}^{-1}),(\mathscr{M}_i^{\mathscr{T}},Y_i^{\mathscr{T}})\big) = \\
&\sum_l\sum_g K_{\mathscr{F}}(y(l),y(g))K_{\mathscr{Z}}(c(l),c(g))\eta(l)\cdot\eta(g) \\
&-2\sum_l\sum_q K_{\mathscr{F}}(y(l),y_i(q))K_{\mathscr{Z}}(c(l),c_i(q))\eta(l)\cdot\eta_i(q) \\
&+\sum_q\sum_r K_{\mathscr{F}}(y_i(q),y_i(r))K_{\mathscr{Z}}(c_i(q),c_i(r))\eta_i(q)\cdot\eta_i(r),
\end{aligned}
\tag{B.11}
$$

with $\cdot$ denoting the scalar product in $\mathbb{R}^3$. Each term now, measures not only differences in geometry but also differences in the functional values between the template and the target FoS.

Subsequently, given the FoSs $\{(\mathscr{M}_i^{\mathscr{T}},Y_i^{\mathscr{T}})\}$ generated as described in Section 3.5, we perform the landmark-free geometric registration by minimizing the objective function in (4.5), with the shape similarity functional (4.7), which is equivalent to the similarity functional (B.11) with $\sigma_{\mathscr{F}}=+\infty$. Thanks to the estimated registration maps we can estimate the functions $\{\hat{X}_i\}$ and compute the cross-sectional mean function $\mu^{\mathscr{T}}$. Subsequently a second registration step can be performed, by minimizing the objective function in (4.5) but this time with the similarity functional (B.11). We have performed this for different choice of $\sigma_{\mathscr{F}}$. The smaller $\sigma_{\mathscr{F}}$, the more we are weighting the functional matching term as opposed to the geometric matching term.

In Figure B.4 we show the results of the fPCA applied to the functions $\{\hat{X}_i\}$ for different choices of $\sigma_{\mathscr{F}}$. These need to be compared with the results in Figure 4.10, obtained by applying the iterative functional registration algorithm in Section 4.3.2. On the left panel of Figure B.4 we can see the mean and first two PC functions estimated when functional information is ignored, which coincide with the one showed on the left panel of Figure 4.10, as they are computed in the same way. On the other two panels of Figure B.4 we can see the mean and first two PC functions estimated when functional information is introduced. As we can see the estimated first PC function resembles the true underlying first PC function, but some fictitious variability is left on the second estimated PC function.

Trying to further decrease $\sigma_{\mathscr{F}}$, to remove the residual fictitious variability, resulted in estimated registration maps failing to bring the template in geometric correspondence to the target surface. Such problem has been the limiting factor in successfully applying the same method to the data in the real application, where the differences in geometries between the template and the target FoSs are much bigger. In fact, this is one of the motivations underlying the introduction of the functional registration algorithm in Sec-

tion 4.3.2, where the ‘moving’ functions are instead ‘constrained’ to lie in the predefined geometry.

Bibliography

Alfeld, P., Neamtu, M., and Schumaker, L. L. (1996). Bernstein-Bézier polynomials on spheres and sphere-like surfaces. *Computer Aided Geometric Design*, 13(4):333–349.

Amini, A. A. and Wainwright, M. J. (2012). Sampled forms of functional PCA in reproducing kernel Hilbert spaces. *The Annals of Statistics*, 40(5):2483–2510.

Arridge, S. R. (1999). Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41–R93.

Audette, M. A., Ferrie, F. P., and Peters, T. M. (2000). An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis*, 4(3):201–217.

Belkin, M. and Niyogi, P. (2002). Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. In *Advances in Neural Information Processing Systems 14*. The MIT Press.

Benko, M., Härdle, W., and Kneip, A. (2009). Common functional principal components. *The Annals of Statistics*, 37(1):1–34.

Bookstein, F. L. (1992). *Morphometric tools for landmark data: Geometry and Biology*. Cambridge University Press, Cambridge.

Bookstein, F. L. (1997). Shape and the Information in Medical Images: A Decade of the Morphometric Synthesis. *Computer Vision and Image Understanding*, 66(2):97–118.

Bosq, D. (2000). *Linear Processes in Function Spaces*, volume 149 of *Lecture Notes in Statistics*. Springer New York, New York, NY.

Braess, D. (2007). *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge University Press, Cambridge.

Bruveris, M., Risser, L., and Vialard, F.-X. (2012). Mixture of Kernels and Iterated Semidirect Product of Diffeomorphisms Groups. *Multiscale Modeling & Simulation*, 10(4):1344–1368.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The Brain's Default Network. *Annals of the New York Academy of Sciences*, 1124(1):1–38.

Cai, D., He, X., Han, J., and Huang, T. S. (2011). Graph Regularized Nonnegative Matrix Factorization for Data Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1548–1560.

Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004.

Charlier, B., Charon, N., and Trouvé, A. (2017). The Fshape Framework for the Variability Analysis of Functional Shapes. *Foundations of Computational Mathematics*, 17(2):287–357.

Charon, N. and Trouvé, A. (2014). Functional Currents: A New Mathematical Tool to Model and Analyse Functional Shapes. *Journal of Mathematical Imaging and Vision*, 48(3):413–431.

Chavel, I. (2006). *Riemannian Geometry: A Modern Introduction*. Cambridge University Press, Cambridge.

Chevallier, E., Kalunga, E., and Angulo, J. (2017). Kernel Density Estimation on Spaces of Gaussian Distributions and Symmetric Positive Definite Matrices. *SIAM Journal on Imaging Sciences*, 10(1):191–215.

Chung, M. K., Robbins, S. M., Dalton, K. M., Davidson, R. J., Alexander, A. L., and Evans, A. C. (2005). Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25(4):1256–1265.

Cornea, E., Zhu, H., Kim, P., and Ibrahim, J. G. (2017). Regression models on Riemannian symmetric spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):463–482.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Dai, X., Lin, Z., and Müller, H.-G. (2018). Modeling Longitudinal Data on Riemannian Manifolds.

Dai, X. and Müller, H.-G. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *The Annals of Statistics*, 46(6B):3334–3361.

Dassi, F., Ettinger, B., Perotto, S., and Sangalli, L. M. (2015). A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Applied Numerical Mathematics*, 90:111–131.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154.

De Vito, E., Rosasco, L., Caponnetto, A., De Giovannini, U., Odone, F., Vito, D., and De Vito, F. O. (2005). Learning from Examples as an Inverse Problem. *Journal of Machine Learning Research*, 6:883–904.

Descary, M.-H. (2017). *Functional data analysis by matrix completion*. PhD thesis, École polytechnique fédérale de Lausanne.

Descary, M.-H. and Panaretos, V. M. (2019). Functional data analysis by matrix completion. *The Annals of Statistics*, 47(1):1–38.

Dobriban, E., Leeb, W., and Singer, A. (2017). Optimal prediction in the linearly transformed spiked model.

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

Dryden, I. L. and Mardia, K. V. (2016). *Statistical Shape Analysis, with Applications in R*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, Chichester, UK.

Duchon, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In Schempp, W. and Zeller, K., editors, *Constructive Theory of Functions of Several Variables. Lecture Notes in Mathematics*. Springer, Berlin, Heidelberg.

Dupuis, P., Grenander, U., and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*, 56(3):587–600.

Dziuk, G. (1988). Finite Elements for the Beltrami operator on arbitrary surfaces. In Hildebrandt, S. and Leis, R., editors, *Partial Differential Equations and Calculus of Variations. Lecture Notes in Mathematics*. Springer, Berlin, Heidelberg.

Dziuk, G. and Elliott, C. M. (2013). Finite element methods for surface PDEs. *Acta Numerica*, 22(April):289–396.

Eltzner, B., Huckemann, S., and Mardia, K. V. (2018). Torus principal component analysis with applications to RNA structure. *The Annals of Applied Statistics*, 12(2):1332–1359.

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

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London.

Fischl, B. and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055.

Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, 8(4):272–284.

Flury, B. N. (1984). Common Principal Components in k Groups. *Journal of the American Statistical Association*, 79(388):892–898.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., and Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, 80:105–124.

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, 26(1):288–303.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric Regression and Generalized Linear Models*. Springer US, Boston, MA.

Hadjipantelis, P. Z., Aston, J. A. D., Müller, H. G., and Evans, J. P. (2015). Unifying Amplitude and Phase Analysis: A Compositional Data Approach to Functional Multivariate Mixed-Effects Modeling of Mandarin Chinese. *Journal of the American Statistical Association*, 110(510):545–559.

Hagler, D. J., Saygin, A. P., and Sereno, M. I. (2006). Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. *NeuroImage*, 33(4):1093–1103.

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

Harrison, S. J., Woolrich, M. W., Robinson, E. C., Glasser, M. F., Beckmann, C. F., Jenkinson, M., and Smith, S. M. (2015). Large-scale probabilistic functional modes from resting state fMRI. *NeuroImage*, 109:217–231.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, volume 200 of *Springer Series in Statistics*. Springer New York, New York, NY.

Hotz, T., Huckemann, S., Le, H., Marron, J. S., Mattingly, J. C., Miller, E., Nolen, J., Owen, M., Patrangenaru, V., and Skwerer, S. (2013). Sticky central limit theorems on open books. *The Annals of Applied Probability*, 23(6):2238–2258.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd, Chichester, UK.

Hu, Y. and Jacob, M. (2012). Higher Degree Total Variation (HDTV) Regularization for Image Recovery. *IEEE Transactions on Image Processing*, 21(5):2559–2571.

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

Huang, Y., Teng, Z., Elkhawad, M., Tarkin, J. M., Joshi, N., Boyle, J. R., Buscombe, J. R., Fryer, T. D., Zhang, Y., Park, A. Y., Wilkinson, I. B., Newby, D. E., Gillard, J. H., and Rudd, J. H. F. (2016). High Structural Stress and Presence of Intraluminal Thrombus Predict Abdominal Aortic Aneurysm 18 F-FDG Uptake. *Circulation: Cardiovascular Imaging*, 9(11):1–9.

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

Im, K., Lee, J. M., Won Seo, S., Hyung Kim, S., Kim, S. I., and Na, D. L. (2008). Sulcal morphology changes and their relationship with cortical thickness and gyral white matter volume in mild cognitive impairment and Alzheimer's disease. *NeuroImage*, 43(1):103–113.

Jin, K. H., McCann, M. T., Froustey, E., and Unser, M. (2017). Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522.

Johnson, R. A. and Wichern, D. W. (2002). *Applied multivariate statistical analysis*. Prentice Hall, Upper Saddle River, NJ, 5. ed edition.

Johnstone, I. M. and Silverman, B. W. (1990). Speed of Estimation in Positron Emission Tomography and Related Inverse Problems. *The Annals of Statistics*, 18(1):251–280.

Jolliffe, I. T., Trendafilov, N. T., and Uddin, M. (2003). A Modified Principal Component Technique Based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3):531–547.

Joshi, S. and Miller, M. I. (2000). Landmark matching via large deformation diffeomorphism. *IEEE Image Proc.*, 9(8)(8):1357–1370.

Katsevich, E., Katsevich, A., and Singer, A. (2015). Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem. *SIAM Journal on Imaging Sciences*, 8(1):126–185.

Kim, Y. T. and Park, H. S. (2013). Geometric structures arising from kernel density estimation on Riemannian manifolds. *Journal of Multivariate Analysis*, 114(2012):112–126.

Kneip, A. and Ramsay, J. O. (2008). Combining Registration and Fitting for Functional Models. *Journal of the American Statistical Association*, 103(483):1155–1165.

Knöppel, F., Crane, K., Pinkall, U., and Schröder, P. (2013). Globally optimal direction fields. *ACM Transactions on Graphics*, 32(4):1.

Kurtek, S., Srivastava, A., Klassen, E., and Ding, Z. (2012). Statistical Modeling of Curves Using Shapes and Related Features. *Journal of the American Statistical Association*, 107(499):1152–1165.

Lee, J. M. (1997). *Riemannian Manifolds: An Introduction to Curvature*, volume 176 of *Graduate Texts in Mathematics*. Springer New York, New York, NY.

Lee, J. M. (2012). *Introduction to Smooth Manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer New York, New York, NY.

Lefkimmiatis, S., Bourquard, A., and Unser, M. (2012). Hessian-Based Norm Regularization for Image Restoration With Biomedical Applications. *IEEE Transactions on Image Processing*, 21(3):983–995.

Lerch, J. P. and Evans, A. C. (2005). Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, 24(1):163–173.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351.

Lila, E. (2014). Smooth principal component analysis over two-dimensional manifolds with application to neuroimaging. Master's Thesis, Politecnico di Milano.

Lila, E., Aston, J. A. D., and Sangalli, L. M. (2016). Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging. *The Annals of Applied Statistics*, 10(4):1854–1879.

Lin, L., St. Thomas, B., Zhu, H., and Dunson, D. B. (2017). Extrinsic Local Regression on Manifold-Valued Data. *Journal of the American Statistical Association*, 112(519):1261–1273.

Lin, Z. and Yao, F. (2018). Intrinsic Riemannian Functional Data Analysis.

Lock, E. F., Hoadley, K. A., Marron, J. S., and Nobel, A. B. (2013). Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *The Annals of Applied Statistics*, 7(1):523–542.

Lustig, M., Donoho, D., Santos, J., and Pauly, J. (2008). Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82.

Maguire, E. A., Woollett, K., and Spiers, H. J. (2006). London taxi drivers and bus drivers: A structural MRI and neuropsychological analysis. *Hippocampus*, 16(12):1091–1101.

Mammen, E. (1991). Estimating a Smooth Monotone Regression Function. *The Annals of Statistics*, 19(2):724–740.

Mammen, E. and Thomas-Agnan, C. (1999). Smoothing Splines and Shape Restrictions. *Scandinavian Journal of Statistics*, 26(2):239–252.

Mardia, K. V. and Patrangenaru, V. (2005). Directions and projective shapes. *The Annals of Statistics*, 33(4):1666–1699.

Marron, J. S. and Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753.

Marron, J. S., Ramsay, J. O., Sangalli, L. M., and Srivastava, A. (2015). Functional Data Analysis of Amplitude and Phase Variation. *Statistical Science*, 30(4):468–484.

Mathé, P. and Pereverzev, S. V. (2006). Regularization of some linear ill-posed problems with discretized random noisy data. *Mathematics of Computation*, 75(256):1913–1929.

McCann, M. T., Jin, K. H., and Unser, M. (2017). Convolutional Neural Networks for Inverse Problems in Imaging: A Review. *IEEE Signal Processing Magazine*, 34(6):85–95.

Miller, M. I., Trouvé, A., and Younes, L. (2006). Geodesic Shooting for Computational Anatomy. *Journal of Mathematical Imaging and Vision*, 24(2):209–228.

Miller, M. I., Trouvé, A., and Younes, L. (2015). Hamiltonian Systems and Optimal Control in Computational Anatomy: 100 Years Since D'Arcy Thompson. *Annual Review of Biomedical Engineering*, 17(1):447–509.

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011:1–9.

Panaretos, V. M. and Zemel, Y. (2016). Amplitude and phase variation of point processes. *The Annals of Statistics*, 44(2):771–812.

Patrangenaru, V. and Bhattacharya, R. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29.

Patrangenaru, V. and Ellingson, L. (2015). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press.

Pelletier, B. (2006). Non-parametric regression estimation on closed Riemannian manifolds. *Journal of Nonparametric Statistics*, 18(1):57–67.

Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics*, 47(2):691–719.

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

Pizer, S. M., Jung, S., Goswami, D., Vicory, J., Zhao, X., Chaudhuri, R., Damon, J. N., Huckemann, S., and Marron, J. S. (2013). Nested Sphere Statistics of Skeletal Models. In Breuß, M., Bruckstein, A., and Maragos, P., editors, *Innovations for Shape Analysis. Mathematics and Visualization.*, pages 93–115. Springer, Berlin, Heidelberg.

Ramsay, J. and Silverman, W. B. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer-Verlag, New York.

Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):307–319.

Repetti, A., Pereyra, M., and Wiaux, Y. (2019). Scalable Bayesian Uncertainty Quantification in Imaging Inverse Problems via Convex Optimization. *SIAM Journal on Imaging Sciences*, 12(1):87–118.

Rice, J. A. and Silverman, B. W. (1991). Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(1):233–243.

Riesz, F. and Szokefalvi-Nagy, B. (1955). *Functional analysis*. F. Ungar Pub. Co., New York.

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

Seo, S., Chung, M. K., and Vorperian, H. K. (2010). Heat Kernel Smoothing Using Laplace-Beltrami Eigenfunctions. In *Med. Imag. Comput Assist Inverv*, pages 505–512.

Shen, H. and Huang, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis*, 99(6):1015–1034.

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

Srivastava, A., Klassen, E., Joshi, S. H., and Jermyn, I. H. (2011). Shape Analysis of Elastic Curves in Euclidean Spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7):1415–1428.

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

Tang, R. and Müller, H. G. (2008). Pairwise curve synchronization for functional data. *Biometrika*, 95(4):875–889.

Tenorio, L. (2001). Statistical Regularization of Inverse Problems. *SIAM Review*, 43(2):347–366.

Thirion, J.-P. (1995). Fast Non-Rigid Matching of 3D Medical Images. Research Report RR-2547, INRIA.

Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260.

Treece, G. and Gee, A. (2015). Independent measurement of femoral cortical thickness and cortical bone density using clinical CT. *Medical Image Analysis*, 20(1):249–264.

Trouvé, A. and Younes, L. (2005). Metamorphoses through lie group action. *Foundations of Computational Mathematics*, 5(2):173–198.

Tward, D., Miller, M., Trouvé, A., and Younes, L. (2017). Parametric Surface Diffeomorphometry for Low Dimensional Embeddings of Dense Segmentations and Imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1195–1208.

Vaillant, M. and Glaunès, J. (2005). Surface Matching via Currents. In Christensen, G. and Sonka, M., editors, *Information Processing in Medical Imaging. IPMI 2005*, pages 381–392. Springer, Berlin, Heidelberg.

Vaillant, M., Miller, M., Younes, L., and Trouvé, A. (2004). Statistics on diffeomorphisms via tangent space representations. *NeuroImage*, 23(SUPPL. 1):S161–S169.

Vaillant, M., Qiu, A., Glaunès, J., and Miller, M. I. (2007). Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus. *NeuroImage*, 34(3):1149–1159.

Van Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Della Penna, S., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S., Prior, F., Schlaggar, B., Smith, S., Snyder, A., Xu, J., and Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, 62(4):2222–2231.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72.

Vogel, C. R. (2002). *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics.

Wahba, G. (1981). Spline Interpolation and Smoothing on the Sphere. *SIAM Journal on Scientific and Statistical Computing*, 2(1):5–16.

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

Yao, K. D., Patrangenaru, V., and Lester, D. (2017). 3D mean Projective Shape Difference for Face Differentiation from Multiple Digital Camera Images. pages 1–13.

Yeo, B., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., and Golland, P. (2010). Spherical Demons: Fast Diffeomorphic Landmark-Free Surface Registration. *IEEE Transactions on Medical Imaging*, 29(3):650–668.

Younes, L. (2010). *Shapes and Diffeomorphisms*. Applied Mathematical Sciences. Springer, Berlin, Heidelberg.

Yuan, Y., Zhu, H., Lin, W., and Marron, J. S. (2012). Local polynomial regression for symmetric positive definite matrices. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(4):697–719.

Zaetz, J. and Kurtek, S. (2015). A Novel Riemannian Framework for Shape Analysis of Annotated Surfaces. In *Procedings of the Proceedings of the 1st International Workshop on DIFFerential Geometry in Computer Vision for Analysis of Shapes, Images and Trajectories 2015*, number Lddmm, pages 3.1–3.11. British Machine Vision Association.

Zhang, E., Mischaikow, K., and Turk, G. (2006). Vector field design on surfaces. *ACM Transactions on Graphics*, 25(4):1294–1326.

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321.

Zhdanov, M. (2002). *Inverse Theory and Applications in Geophysics*. Elsevier.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor Regression with Applications in Neuroimaging Data Analysis. *Journal of the American Statistical Association*, 108(502):540–552.

Zhou, L. and Pan, H. (2014). Principal Component Analysis of Two-Dimensional Functional Data. *Journal of Computational and Graphical Statistics*, 23(3):779–801.

Zhu, H., Chen, Y., Ibrahim, J. G., Li, Y., Hall, C., and Lin, W. (2009). Intrinsic Regression Models for Positive-Definite Matrices With Applications to Diffusion Tensor Imaging. *Journal of the American Statistical Association*, 104(487):1203–1212.

Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Zou, H., Hastie, T., and Tibshirani, R. (2006). Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15(2):265–286.