

Representation and reconstruction of covariance operators in linear inverse problems

Eardi Lila^{1,4} , Simon Arridge²  and John A D Aston³

¹ Department of Biostatistics, University of Washington, United States of America

² Centre for Medical Image Computing, Department of Computer Science, University College London, United Kingdom

³ Statistical Laboratory, DPMMS, University of Cambridge, United Kingdom

E-mail: elila@uw.edu, s.arridge@cs.ucl.ac.uk and j.aston@statslab.cam.ac.uk

Received 10 October 2019, revised 28 March 2020

Accepted for publication 6 April 2020

Published 20 August 2020



CrossMark

Abstract

We introduce a framework for the reconstruction and representation of functions in a setting where these objects cannot be directly 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 images 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 approach to the investigation of functional connectivity.

Keywords: inverse problems, covariance operator, principal component analysis, dynamic functional connectivity, magnetoencephalography

(Some figures may appear in colour only in the online journal)

1. Introduction

An inverse problem is the process of recovering missing information from indirect and noisy observations. Not surprisingly, inverse problems play a central role in numerous fields such as, to name a few, geophysics (Zhdanov 2002), computer vision (Hartley and Zisserman 2004), medical imaging (Arridge 1999, Lustig *et al* 2008) and machine learning (De Vito *et al* 2005).

⁴ Author to whom any correspondence should be addressed.



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

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

$$y = \mathcal{K}x + \varepsilon, \quad (1)$$

where y and ε belong to an either finite or infinite dimensional Banach space. The map \mathcal{K} is called a forward operator and is generally assumed to be known, although its uncertainty has also been taken into account in the literature (Arridge *et al* 2006, Golub and van Loan 1980, Gutta *et al* 2019, Kluth and Maass 2017, Lehtikoinen *et al* 2007, Nissinen *et al* 2009, Zhu *et al* 2011). The term ε represents observational error.

Problem 1 is a well-studied problem within applied mathematics (for early works in the field, see Adorf 1995, Calderón 1980, Geman 1990). 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 the function x in (1) is generally tackled by minimizing a functional which is the sum of a data (fidelity) term and a regularizing term encoding prior information on the function to be recovered (see, among others, Cavalier 2008, Hu and Jacob 2012, Lefkimmatis *et al* 2012, Mathé and Pereverzev 2006, Tenorio 2001). For convex optimization functionals, modern efficient optimization methods can be applied (Beck and Teboulle 2009, Boyd *et al* 2010, Burger *et al* 2016, Chambolle and Pock 2011, Chambolle and Pock 2016). Alternatively, when it is important to assess the uncertainty associated with the estimates, a Bayesian approach could be adopted (Calvetti and Somersalo 2007, Kaipio and Somersalo 2005, Repetti *et al* 2019, Stuart 2010). The deep convolutional neural network approach has also been applied to this setting (Jin *et al* 2017, McCann *et al* 2017).

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

$$\mathcal{S}_i = \mathcal{K}_i \circ \mathcal{C}_i \circ \mathcal{K}_i^* + \mathcal{E}_i, \quad i = 1, \dots, n, \quad (2)$$

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

As opposed to more classical linear inverse problems formulations, problem 2 introduces the following additional difficulties:

- We are in a setting where each sample is a high-dimensional object that is a covariance operator; it is important to take advantage of the information from all the samples to reconstruct and represent each of them.
- The elements $\{\mathcal{C}_i\}$ and $\{\mathcal{S}_i\}$ live on non-Euclidean spaces, as they belong to the positive semidefinite cone, and it is important to account for this manifold structure 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.

More general non-Euclidean settings could also be accommodated. Specifically, the error term could be defined on a tangent space and mapped to the original space through the exponential mapping. Another setting of interest is the case of error terms that push the observables out of the original space. In our applications this is not an issue, as the contaminated observations are themselves empirical covariance matrices, which belong to the non-Euclidean space of positive semidefinite matrices.

We tackle problem 2 by generalizing the concept of principal component analysis (PCA) to optimally represent and understand the variation associated with samples that are indirectly observed covariance operators. The proposed model is also able to deal with the simpler case of samples that are indirectly observed functional images belonging to a linear functional space.

1.1. Motivating application — functional connectivity

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

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 and, as shown in figure 1 in 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 paper for convenience, however it is important to highlight that the formulation of the problem is much more general than the setting of this specific application.

The signals on the brain space are related to the signals on the sensors space by a forward operator, derived from the physical modeling of the electrical/magnetic propagation, from the cerebral cortex to the sensors. This is generally referred to as the forward problem. For soft-field methods like EEG, MEG and functional near-infrared spectroscopy (Eggebrecht *et al* 2014, Ferrari and Quaresima 2012, Mosher *et al* 1999, Singh *et al* 2014, Ye *et al* 2009), 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 sensors space, is strongly ill-posed. In fact, signals with fairly different intensities on the brain space, due to the diffusion effect, result in signals with similar intensities in the sensors space. In figure 1, we show an example of a signal on the brain space and the associated signal on the sensors space.

From a practical perspective, it is crucial to understand how the different parts of the brain interact, which is sometimes known as functional connectivity. A possible way to understand these interactions is by analyzing the covariance function associated with the signals describing the cerebral activity of an individual on the brain space (Fransson *et al* 2011, Lee *et al* 2013,

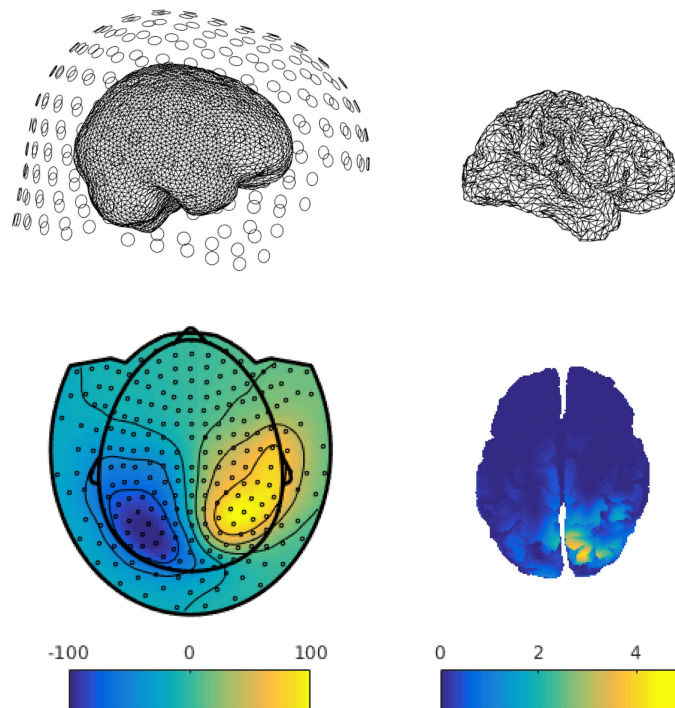


Figure 1. On the top left, head model of a subject and superimposition of the 248 MEG sensors positioned around the head, called ‘sensors space’. On the top right, brain model of the same subject represented by a triangular mesh of 8k nodes, which represents the ‘brain space’. On the bottom left, an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

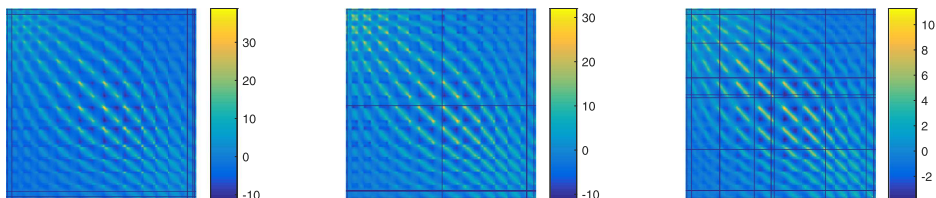


Figure 2. Covariance matrices of the signal detected by the MEG sensors from three different subjects of the human connectome project. The size of the matrices is 248×248 . The dark blue bands represent missing data, which are due to the exclusion of some channels after a quality check of the signal.

Li *et al* 2009). 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 2, we show the covariance matrices, on the sensors space, computed from the MEG signals of three different subjects.

The remainder of this paper is organized as follows. In section 2 we give a formal description of the problem. We then introduce a model for indirectly observed smooth functional images in section 3 and present the more general model associated with problem 2 in section 4. In section 5, we perform simulations to assess the validity of the estimation framework. In section 6 we apply the proposed models to MEG data and we finally give some concluding remarks in section 7.

2. Mathematical description of the problem

We now introduce the problem using our driving application as an example. To this purpose, let \mathcal{M} be a closed smooth two-dimensional manifold embedded in \mathbb{R}^3 , which in our application represents the geometry of the cerebral cortex. An example of such a surface is shown on the top right of figure 1. We denote with $L^2(\mathcal{M})$ the space of square integrable functions on \mathcal{M} . Define X to be a random function with values in a Hilbert functional space $\mathcal{F} \subset L^2(\mathcal{M})$ with mean $\mu = \mathbb{E}[X]$, finite second moment, and assume the continuity and square integrability of its covariance function $C_X(v, v') = \mathbb{E}[(X(v) - \mu(v))(X(v') - \mu(v'))]$. The associated covariance operator \mathcal{C}_X is defined as $\mathcal{C}_X g = \int_{\mathcal{M}} C_X(v, v') g(v) dv$, for all $g \in L^2(\mathcal{M})$. Mercer's lemma (Riesz and Szokefalvi-Nagy 1955) guarantees the existence of a non-increasing sequence $\{\gamma_r\}$ of eigenvalues of \mathcal{C}_X and an orthonormal sequence of corresponding eigenfunctions $\{\psi_r\}$, such that

$$C_X(v, v') = \sum_{r=1}^{\infty} \gamma_r \psi_r(v) \psi_r(v'), \quad \forall v, v' \in \mathcal{M}. \quad (3)$$

As a direct consequence, X can be expanded⁵ as $X = \mu + \sum_{r=1}^{\infty} \zeta_r \psi_r$, where the random variables $\{\zeta_r\}$ are uncorrelated and are given by $\zeta_r = \int_{\mathcal{M}} \{X(v) - \mu(v)\} \psi_r(v) dv$. The collection $\{\psi_r\}$ defines the modes of variation of the random function X , in descending order of strength, and these are called principal component (PC) functions. The associated random variables $\{\zeta_r\}$ are called PC scores. Moreover, the defined PC functions are the best finite basis approximation in the L^2 -sense, therefore for any fixed $R \in \mathbb{N}$, the first R PC functions of X minimize the reconstruction error, i.e.

$$\{\psi_r\}_{r=1}^R = \underset{\{\phi_r\}_{r=1}^R: \langle \phi_r, \phi_{r'} \rangle = \delta_{rr'}}{\operatorname{argmin}} \mathbb{E} \int_{\mathcal{M}} \left\{ X(v) - \mu(v) - \sum_{r=1}^R \langle X - \mu, \phi_r \rangle \phi_r(v) \right\}^2 dv, \quad (4)$$

where $\langle \cdot, \cdot \rangle$ denotes the $L^2(\mathcal{M})$ inner product and $\delta_{rr'}$ is the Kronecker delta; i.e. $\delta_{rr'} = 1$ for $r = r'$ and 0 otherwise.

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 $\{K_l : l = 1, \dots, m\}$ be a collection of $s \times p$ real matrices, representing the potentially sample-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. The matrices $\{K_l\}$ are

⁵ More precisely, we have that $\lim_{R \rightarrow \infty} \sup_{v \in \mathcal{M}} \mathbb{E} \left\{ X(v) - \mu(v) - \sum_{r=1}^R \zeta_r \psi_r(v) \right\}^2 = 0$, i.e. the series converges uniformly in mean-square.

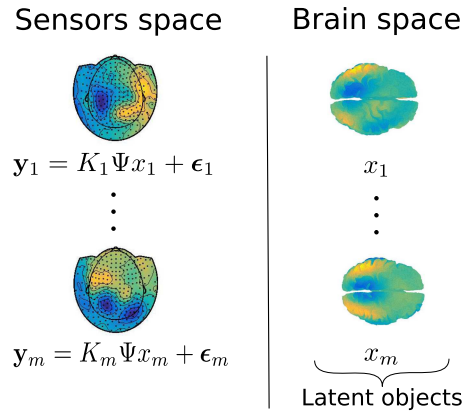


Figure 3. Illustration of the setting introduced with model (5).

discrete versions of the forward operator \mathcal{K} introduced in section 1. Moreover, define the evaluation operator $\Psi : \mathcal{F} \rightarrow \mathbb{R}^p$ to be a vector-valued functional that evaluates a function $f \in \mathcal{F}$ at the p pre-specified points $\{v_j\} \subset \mathcal{M}$, returning the p dimensional vector $(f(v_1), \dots, f(v_p))^T$. The operators Ψ and $\{K_l\}$ are known. However, in the described problem the random function X can be observed only through indirect measurements $\{\mathbf{y}_l \in \mathbb{R}^s : l = 1, \dots, m\}$ generated from the model

$$\begin{cases} x_l = \mu + \sum_{r=1}^{\infty} \zeta_{l,r} \psi_r \\ \mathbf{y}_l = K_l \Psi x_l + \varepsilon_l, \quad l = 1, \dots, m \end{cases} \quad (5)$$

where $\{x_l\}$ are m independent realizations of X , and thus expandible in terms of the PC functions $\{\psi_r\}$ and the coefficients $\{\zeta_{l,r}\}$ given by $\zeta_{l,r} = \int_{\mathcal{M}} \{x_l(v) - \mu(v)\} \psi_r(v) dv$. The terms $\{\varepsilon_l\}$ represent observational errors and are independent realizations of 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.

We consider the problem of estimating the PC functions $\{\psi_r\}$ in (5), and associated scores $\{\zeta_{l,r}\}$, from the observations $\{\mathbf{y}_l\}$. In figure 3 we give an illustration of the introduced setting. Note that it would not be necessary to define the evaluation operator if the forward operators were defined to be functionals $\{\mathcal{K}_l : \mathcal{F} \rightarrow \mathbb{R}^p\}$, relating directly the functional objects on the brain space to the real vectors on the sensors space. It is however the case that the operators $\{K_l\}$ are computed in a matrix form by third party software (see section 6 for details) for a pre-specified set of points $\{v_j\} \subset \mathcal{M}$ and it is thus convenient to take this into account in the model through the introduction of an evaluation operator Ψ .

In the case of single subject studies, the surface \mathcal{M} is the subject's reconstructed cortical surface, an example of which is shown on the right panel of figure 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, \mathcal{M} is assumed to be a template cortical surface. We are thus assuming that the individual cortical surfaces have been registered to the template \mathcal{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 \mathcal{M} , where the PC functions are defined.

However, notice that when it comes to the computation of the forward operators, we are not assuming the brain geometries of the single subjects to be all equal to a geometric template, as in fact the model in (5) allows for sample-specific forward operators $\{K_i\}$. The individual cortical surfaces could also have different number of mesh points, in that case the subject-specific ‘resampling’ operator could be absorbed into the definition of sample-specific evaluation operators $\{\Psi_i\}$.

The estimation of the PC functions in (5) has been classically dealt with by reconstructing each observation x_i independently and subsequently performing PCA. However, such an approach can be sub-optimal in particular in a low signal-to-noise setting, as when estimating one signal, the information from all the other sampled signals is systematically ignored. The statistical analysis of data samples that are random functions or surfaces, i.e. functional data, has also been explored in the functional data analysis (FDA) literature (Ramsay and Silverman 2005), however, most of those works focus on the setting of fully observed functions. An exception to this is the sparse FDA literature (see e.g. Yao *et al* 2005), where instead the functional samples are assumed to be observable only through irregular and noisy evaluations.

In the case of direct but noisy observations of a signal, previous works on statistical estimation of the covariance function, and associated eigenfunctions, have been made, for instance, in Bunea and Xiao (2015) for regularly sampled functions and in Huang *et al* (2008), Yao *et al* (2005) for sparsely sampled functions. A generalization to functions whose domain is a manifold is proposed in Lila *et al* (2016) and appropriate spatial coherence is introduced by penalizing directly the eigenfunctions of the covariance operator to be estimated, i.e. the PC functions. In the indirect observations setting, Tian *et al* (2012) propose a separable model in time and space for source localization. The estimation of PC functions of functional data in a linear space and on linear domains, from indirect and noisy samples, has been previously covered in Amini and Wainwright (2012). They propose a regularized M-estimator in a reproducing kernel Hilbert space (RKHS) framework. Due to the fact that in practice the introduction of an RKHS relies on the definition of a kernel, i.e. a covariance function on the domain, this approach cannot be easily extended to non-linear domains. 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. However, a regularized approach is crucial in our setting, due to the strong ill-posedness of the inverse problem considered. In the discrete setting, also other forms of regularization have been adopted, e.g. sparsity on the inverse covariance matrix (Friedman *et al* 2008, Liu and Zhang 2019).

2.2. The case of indirectly observed covariance operators

A natural generalization of the setting introduced 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 underlying covariance operators $\{C_i : i = 1, \dots, n\}$ on the brain space. In our driving application, each covariance function $C_i : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ describes the functional connectivity of the i th individual or the functional connectivity of the same individual at the i th time-point.

We consider the problem of defining and estimating a set of covariance functions, that we call PC covariance functions, which enable the description of $\{C_i\}$ through the ‘linear combinations’ of few components. Such a reduced order description is of interest, for example, in understanding how functional connectivity varies among individuals or over time.

We define a model for the PC covariance functions of $\{C_i\}$ from the set of indirectly observed covariance matrices, computed from the signals on the sensors space, and thus given

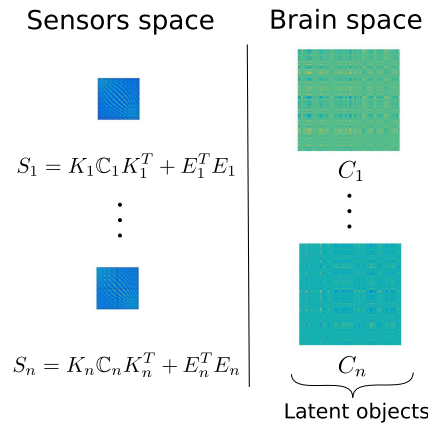


Figure 4. Illustration of the setting introduced with model (6).

by $\{S_i \in \mathbb{R}^{s \times s}, i = 1, \dots, n\}$ with

$$S_i = K_i C_i K_i^T + E_i^T E_i, \quad i = 1, \dots, n, \quad (6)$$

where $C_i = (C_i(v_j, v_f))_{j,f}$, and $\{v_j : j = 1, \dots, p\}$ are the sampling points associated with 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 an $s \times s$ matrix such that each entry is an independent sample of a Gaussian distribution with mean zero and standard deviation σ . Model (6) could be regarded as an implementation of the idealized problem 2, where the covariance operators are represented by the associated covariance functions. An illustration of the setting introduced can be found in figure 4.

The problem introduced in this section has not been extensively covered in the literature. In the discrete case, Dryden *et al* (2009) introduce a tangent PCA model for directly observed covariance matrices. An extension to directly observed covariance operators has been proposed in Pigoli *et al* (2014). Also related to our work is the setting considered in Petersen and Müller (2019), where the authors propose a regression framework for responses that are random objects (e.g. covariance matrices) with Euclidean predictors. The proposed regression model is applied to study associations between age and low-dimensional correlation matrices, representing functional connectivity, which have been computed from a parcellation of the brain. In section 4, we propose a novel PCA approach for indirectly observed high-dimensional covariance matrices.

3. Principal components of indirectly observed functions

The aim of this section is to define a model for the estimation of the PC functions $\{\psi_r\}$ from the observations $\{y_i\}$, defined in (5). Although the model proposed in this section is not the main contribution of this work, it allows us to introduce some of the concepts necessary to the definition of the more general model for indirectly observed covariance functions in section 4.

3.1. Model

Now let $\mathbf{z} = (z_1, \dots, z_m)^T$ be an m -dimensional real column vector and $H^2(\mathcal{M})$ be the Sobolev space of functions in $L^2(\mathcal{M})$ with first and second distributional derivatives in $L^2(\mathcal{M})$. From

now on \mathcal{F} is instantiated with $H^2(\mathcal{M})$. We propose to estimate $\hat{f} \in H^2(\mathcal{M})$, the first PC function of X , and the associated PC scores vector \mathbf{z} , by solving the equation

$$(\hat{\mathbf{z}}, \hat{f}) = \arg \min_{\mathbf{z} \in \mathbb{R}^m, f \in H^2(\mathcal{M})} \sum_{l=1}^m \|\mathbf{y}_l - z_l K_l \Psi f\|^2 + \lambda \mathbf{z}^T \mathbf{z} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f, \quad (7)$$

where $\|\cdot\|$ is the Euclidean norm and Δ is the Laplace–Beltrami operator, which enables a smoothing regularizing effect on the PC function \hat{f} . The data fit term encourages $K_l \Psi f$ to capture the strongest mode of variation of $\{\mathbf{y}_l\}$. 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 applying model (7) on the residuals $\{\mathbf{y}_l - \hat{z}_l K_l \Psi \hat{f}\}$, 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 (5), with the terms $\{z_l\}$ playing the role of estimates of the variables $\{\zeta_{l,1}\}$.

In the simplified case of a single forward operator $K := K_1 = \dots = K_m$, the minimization problem (7) can be reformulated in a more classical form. In fact, fixing f in (7) and minimizing over \mathbf{z} gives

$$z_l = \frac{\mathbf{y}_l^T K \Psi f}{\|K \Psi f\|^2 + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f}, \quad l = 1, \dots, m, \quad (8)$$

which can then be used to show that the minimization problem (7) is equivalent to maximizing

$$\frac{(\Psi f)^T K^T \mathbb{Y}^T \mathbb{Y} K (\Psi f)}{\|K \Psi f\|^2 + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f}, \quad (9)$$

with \mathbb{Y} an $m \times s$ real matrix, where the l th row of \mathbb{Y} is the observation \mathbf{y}_l^T . This reformulation gives further insights on the interpretation of \hat{f} in (7). In fact, \hat{f} is such that $K \Psi \hat{f}$ maximizes $(K \Psi \hat{f})^T \left\{ \frac{1}{m} \mathbb{Y}^T \mathbb{Y} \right\} (K \Psi \hat{f})$ subject to a norm constraint. The term $\left\{ \frac{1}{m} \mathbb{Y}^T \mathbb{Y} \right\}$ is the empirical covariance matrix in the sensors space. The term $\mathbf{z}^T \mathbf{z}$ in (7) places the regularization term $\lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f$ in the denominator of the equivalent formulation (9). Thus, \hat{f} is regularized by the choice of norm in the denominator of (9), in a similar fashion to the classic functional principal component formulation of Silverman (1996). Ignoring the spatial regularization, the point-wise evaluation of the PC function Ψf in (9) can be interpreted as the first PC vector computed from the dataset of backprojected data $[K_1^T \mathbf{y}_1, \dots, K_m^T \mathbf{y}_m]^T$, similarly to what is proposed in Dobriban et al (2017) in the context of optimal prediction.

3.2. Algorithm

Here we propose a minimization approach for the objective function in (7), which we approach by alternating the minimization of \mathbf{z} and f in an iterative algorithm. In (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 $\|\mathbf{z}\| = 1$, which leads to a normalized version of the estimator (8):

$$z_l = \frac{\mathbf{y}_l^T K_l \Psi f}{\sqrt{\sum_{l=1}^m \mathbf{y}_l^T K_l \Psi f}}, \quad l = 1, \dots, m. \quad (10)$$

For a given \mathbf{z} , solving (7) with respect to f will turn out to be equivalent to solving an inverse problem, which we discretize adopting a mixed finite elements approach (Azzimonti et al 2014). Specifically, consider now a triangulated surface $\mathcal{M}_{\mathcal{T}}$, union of the finite set of triangles \mathcal{T} , giving an approximated representation of the manifold \mathcal{M} . We then consider the linear finite element space V consisting 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 $\phi_1, \dots, \phi_{\kappa}$ associated with the nodes $\xi_1, \dots, \xi_{\kappa}$, corresponding to the vertices of the triangulation $\mathcal{M}_{\mathcal{T}}$. Such basis functions are Lagrangian, meaning that $\phi_i(\xi_j) = 1$ if $i = j$ and $\phi_i(\xi_j) = 0$ otherwise. Setting $\mathbf{c} = (f(\xi_1), \dots, f(\xi_{\kappa}))^T$ and $\phi = (\phi_1, \dots, \phi_{\kappa})^T$, every function $f \in V$ has the form

$$f(v) = \sum_{k=1}^{\kappa} f(\xi_k) \phi_k(v) = \mathbf{c}^T \phi(v) \quad (11)$$

for all $v \in \mathcal{M}_{\mathcal{T}}$. To ease the notation, we assume that the p points $\{v_j\}$ associated with the evaluation operator Ψ coincide with the nodes of the triangular mesh $\xi_1, \dots, \xi_{\kappa}$, 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 $\{K_l\}$ to be $s \times \kappa$ matrices, relating the κ points on the cortical surface of the i th sample, in one-to-one correspondence to $\xi_1, \dots, \xi_{\kappa}$, to the s -dimensional signal detected on the sensors for the i th sample.

Let now M and A be the mass and stiffness $\kappa \times \kappa$ matrices defined as $(M)_{jj'} = \int_{\mathcal{M}_{\mathcal{T}}} \phi_j \phi_{j'}$ and $(A)_{jj'} = \int_{\mathcal{M}_{\mathcal{T}}} \nabla_{\mathcal{M}_{\mathcal{T}}} \phi_j \cdot \nabla_{\mathcal{M}_{\mathcal{T}}} \phi_{j'}$, where $\nabla_{\mathcal{M}}$ is the gradient operator on the manifold \mathcal{M} . Practically, $\nabla_{\mathcal{M}_{\mathcal{T}}} \phi_j$ is a constant function on each triangle of $\mathcal{M}_{\mathcal{T}}$, and can take an arbitrary value on the edges⁶.

Let $h = \max_{\tau \in \mathcal{T}}(\text{diam}(\tau))$ denote the maximum diameter of the triangles forming $\mathcal{M}_{\mathcal{T}}$, then the solution \hat{f}_h of (7), in the discrete space V , is given by the following proposition.

Proposition 1. *The surface finite element solution $\hat{f}_h \in V$ of model (7), for a given unitary norm vector \mathbf{z} , is $\hat{f}_h = \hat{\mathbf{c}}^T \phi$ where $\hat{\mathbf{c}}$ is the solution of*

$$\hat{\mathbf{c}} = \left(\sum_{l=1}^m z_l^2 K_l^T K_l + \lambda A M^{-1} A \right)^{-1} \sum_{l=1}^m z_l K_l^T \mathbf{y}_l. \quad (12)$$

Equation (12) has the form of a penalized regression, where the discretized version of the penalty term is $A M^{-1} A$.

The sparsity of the linear system (12), namely the number of zeros, depends on the sparsity of its components. The matrices M and A are very sparse, however M^{-1} is not, in general. To overcome this problem, 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 $\tilde{M}_{jj} = \sum_{j'} M_{jj'}$ (Fried and Malkus 1975, Zienkiewicz et al 2013). The penalty operator $\tilde{A} \tilde{M}^{-1} A$ approximates very well the behavior of $A M^{-1} A$.

Moreover, in the case of longitudinal studies that involve only one subject, we have a single forward operator $K := K_1 = \dots = K_m$ common to all the observed signals, and consequently

⁶Formally, these are weak derivatives, hence uniquely defined almost everywhere (i.e. up to a set of measure zero) and are always evaluated in an integral form (see Dziuk and Elliott 2013, for further details).

Algorithm 1. Inverse problems—PCA algorithm

1: Initialization:

(a) Computation of M and A (b) Initialize \mathbf{z} , the scores vector associated with the first PC function2: PC function's estimation: compute \mathbf{c} such that

$$\left(\sum_{l=1}^m z_l^2 K_l^T K_l + \lambda A n M^{-1} A\right) \mathbf{c} = \sum_{l=1}^m z_l K_l^T \mathbf{y}_l$$

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

3: Scores estimation:

$$z_l \leftarrow \frac{\mathbf{y}_l^T K_l \Psi f_h}{\sqrt{\sum_{l=1}^m \mathbf{y}_l^T K_l \Psi f_h}}, \quad l = 1, \dots, m$$

4: Repeat steps 2 and 3 until convergence

equation (12) can be rewritten as the sparse overdetermined system

$$\begin{bmatrix} K \\ \sqrt{\lambda} \tilde{M}^{-1/2} A \end{bmatrix} \mathbf{c} = \begin{bmatrix} Y^T \mathbf{z} \\ \mathbf{0} \end{bmatrix}, \quad (13)$$

to be interpreted in a least-square sense. A sparse QR solver can be finally applied to efficiently solve the linear system (13).

In algorithm 1 we summarize the main algorithmic steps to compute the PC functions and associated PC scores for indirectly observed functions. The initializing scores \mathbf{z} can be chosen either at random or, when there is a correspondence between the detectors of different samples (e.g. $K_1 = \dots = K_m$), with the scores obtained by performing PCA on the observations in the sensors space.

3.3. Eigenfunctions of indirectly observed covariance operators

Suppose now we are in the case of a single forward operator K . Combining steps 2 and 3 of algorithm 1, and moving the normalization step from (z_l) to f_h , we obtain the iterations

$$(K^T K + \lambda A M^{-1} A) \mathbf{c} = K^T \sum_{l=1}^m (\mathbf{y}_l \mathbf{y}_l^T) K \Psi f_h$$

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

The obtained algorithm depends on the data only through $\sum_{l=1}^m \mathbf{y}_l \mathbf{y}_l^T$ 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 $\{\mathbf{y}_l\}$ are not available, but we are given only the associated $s \times s$ covariance matrix on the sensors space, computed from $\{\mathbf{y}_l\}$. This could be of interest in situations where the temporal resolution is very high and the spatial resolution is low, therefore it is convenient to store the covariance matrix rather than the entire set of observations.

4. Reconstruction and representation of indirectly observed covariance operators

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

relating the signal at the p pre-specified points $\{v_j\}$ on the cortical surface \mathcal{M} with the signal detected on the s sensors.

The aim of this section is to introduce a model for the reconstruction and representation of the covariance functions $\{C_i\}$, on the brain space, associated with the actually observed covariance matrices $\{S_i\}$, on the sensors space. The matrices $\{S_i\}$ are related to the covariance functions $\{C_i\}$ through formula (6) that we recall here being

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

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

First, in section 4.1, we see how the PC model introduced in section 3 could be applied to individually reconstruct the covariance functions $\{C_i\}$. In section 4.2, we introduce a population model that achieves both reconstruction and joint representation of the underlying covariance functions $\{C_i\}$.

4.1. 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 $S_i^{1/2} = D_i^{1/2} V_i^T$ 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 . Another possible choice for the square-root decompositions is $S_i^{1/2} = V_i D_i^{1/2} V_i^T$. The output of the proposed algorithms will not depend on the specific choice of the square-root decompositions.

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

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

where, for each i , $\gamma_{i1} \geq \gamma_{i2} \geq \dots \geq 0$ is a sequence of non-increasing variances and $\{\psi_{ir}\}_r$ a set of orthonormal eigenfunctions. Introduce now $\{f_i \in H^2(\mathcal{M})\}$ and $\{\hat{\mathbf{z}}_i \in \mathbb{R}^s\}$, obtained by applying model (7) to each sample independently, i.e.

$$\left\{ (\hat{\mathbf{z}}_i, \hat{f}_i) \right\}_i = \underset{\{\mathbf{z}_i\} \subset \mathbb{R}^s, \{f_i\} \subset H^2(\mathcal{M})}{\operatorname{arg\,min}} \quad \|S_i^{1/2} - \mathbf{z}_i (K_i \Psi f_i)^T\|_{\mathbb{F}}^2 + \lambda \|\mathbf{z}_i\|^2 \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f_i, \quad i = 1, \dots, n, \quad (14)$$

with $\|\cdot\|_{\mathbb{F}}$ denoting the Frobenius matrix norm. Each estimate \hat{f}_i , from model (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 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 (14). This leads to a set of estimates $\{\hat{f}_{ir}\}$ and $\{\hat{\mathbf{z}}_{ir}\}$.

The unregularized version of model (14) is equivalent to a singular value decomposition applied to each matrix $S_i^{1/2}$ independently, which would lead to a set of orthogonal estimates $\{\hat{\mathbf{z}}_{ir}\}_r \subset \mathbb{R}^s$, for each $i = 1, \dots, n$. In the regularized model orthogonality is not enforced, however the estimated PC components can be orthogonalized post-estimation by means of a QR decomposition.

Define now the empirical variances to be $\hat{\gamma}_{ir} = \|\hat{f}_{ir}\|_{L^2(\mathcal{M})}^2$ and consider the $L^2(\mathcal{M})$ -normalized version of $\{\hat{f}_{ir}\}$. An approximate representation of $S_i = (S_i^{1/2})^T S_i^{1/2}$ is thus given by

$$S_i = K_i \sum_r \left\{ \hat{\gamma}_{ir} (\Psi \hat{f}_{ir})(\Psi \hat{f}_{ir})^T \right\} K_i^T, \quad (15)$$

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

$$C_i = \sum_r \hat{\gamma}_{ir} \hat{f}_{ir} \otimes \hat{f}_{ir},$$

where $\hat{\gamma}_{ir}$ is an estimate of the variance γ_{ir} and \hat{f}_{ir} is an estimate of ψ_{ir} . The tensor product $\hat{f}_{ir} \otimes \hat{f}_{ir}$ is such that $(\hat{f}_{ir} \otimes \hat{f}_{ir})(v, v') = \hat{f}_{ir}(v)\hat{f}_{ir}(v')$ for all $v, v' \in \mathcal{M}$. The regularizing terms in (14) introduce spatial coherence on the estimated $\{\hat{f}_{ir}\}$ and thus on the estimated eigenfunctions of $\{C_i\}$, fundamental in an inverse problems setting.

The reconstructed covariance functions $\{C_i\}$ could be discretized on a dense grid, leading to a collection of covariance matrices $(C_i(v_j, v_l))_{jl}$. Following the approach in Dryden *et al* (2009), a Riemannian metric could be defined on the space of covariance matrices, followed by projection of $(C_i(v_j, v_l))_{jl}$ on the tangent space centered at the sample Fréchet mean. PCA could then be carried out on vectorizations of the tangent space representations. A related approach, for covariance functions, has been adopted in Pigoli *et al* (2014).

However, the aforementioned approaches could be prohibitive in our setting. In fact, performing PCA on tangent space projections produces modes of variation that are geodesics passing through the mean, and whose interpretation in a high-dimensional setting is often challenging. Therefore, in the next section, we propose an alternative model that enables joint reconstruction, and representation on a ‘common basis’, of indirectly observed covariance functions.

4.2. 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:

$$(\{\hat{\mathbf{z}}_i\}, \hat{f}) = \arg \min_{\{\mathbf{z}_i\} \subset \mathbb{R}^s, f \in H^2(\mathcal{M})} \sum_{i=1}^n \|S_i^{1/2} - \mathbf{z}_i (K_i \Psi f)^T\|_{\mathbb{F}}^2 + \lambda \sum_{i=1}^n \|\mathbf{z}_i\|^2 \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \quad (16)$$

The newly defined model, as opposed to model (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 $\hat{\mathbf{z}}_{ir}$.

Define now the empirical variances to be $\hat{\gamma}_{ir} = \|\hat{\mathbf{z}}_{ir}\|^2 \|\hat{f}_r\|_{L^2(\mathcal{M})}^2$ and consider the $L^2(\mathcal{M})$ -normalized version of $\{\hat{f}_r\}$. The empirical term in model (16) suggests an approximate representation of S_i , that is

$$C_i = \sum_r \hat{\gamma}_{ir} \hat{f}_r \otimes \hat{f}_r, \quad (17)$$

where each underlying covariance function C_i is approximated by the sum of the product between a subject-specific constant $\hat{\gamma}_{ir}$ and a component $\hat{f}_r \otimes \hat{f}_r$ common to all the observations. The regularizing term in (16) introduces spatial coherence on the estimated functions $\{\hat{f}_r\}$.

The covariance operators $\{C_i\}$ are said to be commuting if $C_i C_{i'} = C_{i'} C_i$ for all $i, i' = 1, \dots, n$. This property can be equivalently characterized as

$$C_i(v, v') = \sum_{r=1}^{\infty} \gamma_{ir} \psi_r(v) \psi_r(v'), \quad \forall v, v' \in \mathcal{M}, \quad (18)$$

with $\{\gamma_{ir}\}_r$ 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 $\{\hat{\gamma}_{ir}\}$ estimates of $\{\gamma_{ir}\}$.

On the one hand, model (16) constrains the estimated covariances to be of the form $C_i = \sum_r \hat{\gamma}_{ir} \hat{f}_r \otimes \hat{f}_r$ and not of the more general form $C_i = \sum_r \hat{\gamma}_{ir} \hat{f}_{ir} \otimes \hat{f}_{ir}$. On the other hand, such a model takes advantage of all the n samples to estimate the components $\{\hat{f}_r \otimes \hat{f}_r\}$. Moreover, the associated variables $\{\hat{\gamma}_{ir}\}$ give a convenient approximate description of the i th covariance, as they are comparable across samples, as opposed to the one computed from model (14). In fact, the i th covariance function can be represented by the variance vector $(\hat{\gamma}_{i1}, \dots, \hat{\gamma}_{iR})^T$, for a suitable truncation level R , where each entry is associated with the rank-one component $\hat{f}_r \otimes \hat{f}_r$. For each r , a scatter plot of the variances $\{\hat{\gamma}_{ir}\}_i$, as the one in figure 14, helps understand what the average contribution of the r th components is and what its variability across samples is. Model (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 $C = \frac{1}{n} \sum_{i=1}^n C_i$.

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

4.3. Algorithm

The minimization in (14), for each fixed i , is a particular case of the one in (7) (see section 3.2), so we focus on the minimization problem in (16) which is also 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

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

with

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

The estimate of f given $\{\mathbf{z}_i\}$, in the discrete space V introduced in section 3.2, is given by the following proposition.

Algorithm 2. Inverse problems—covariance PCA 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

$$S_i^{1/2} = D_i^{1/2} V_i^T,$$
 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 (14): compute \mathbf{c} such that

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

$$f_h \leftarrow \mathbf{c}^T \phi$$
 - 4: Scores estimation from model (14):

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

$$\mathbf{z}_i \leftarrow \frac{\mathbf{z}_i}{\sqrt{\sum_{i=1}^n \|\mathbf{z}_i\|^2}}, \quad i = 1, \dots, n$$
 - 5: Repeat step 3 and 4 until convergence
-

Proposition 2. The surface finite element solution $\hat{f}_h \in V$ of model (16), given the vectors $\{\mathbf{z}_i\}$, is $\hat{f}_h = \hat{\mathbf{c}}^T \phi$ where $\hat{\mathbf{c}}$ is the solution of

$$\left(\sum_{i=1}^n \|\mathbf{z}_i\|^2 K_i^T K_i + \lambda A M^{-1} A \right) \hat{\mathbf{c}} = \sum_{i=1}^n K_i^T (S_i^{1/2})^T \mathbf{z}_i. \quad (19)$$

Algorithm 2 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 $F_r = \hat{f}_r \otimes \hat{f}_r$, 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 $F_r = \hat{f}_r \otimes \hat{f}_r$ spatial coherence can be imposed by regularizing \hat{f}_r , as in fact done for model (14), and this is fundamental in the setting of indirectly observed covariance functions. Finally, due to their size, it might not be possible to store the full reconstructions of the covariance functions $\{C_i\}$ on the brain space, instead, the representation model in (17) allows for an efficient joint representation of such covariance functions in terms of rank-one components.

5. Simulations

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

5.1. Indirectly observed functions

We consider $\mathcal{M}_{\mathcal{T}}$ to be a triangular mesh, with 8k nodes, representing the cortical surface geometry of a subject, as shown on the left panel of figure 1. Each of the 8k nodes will represent

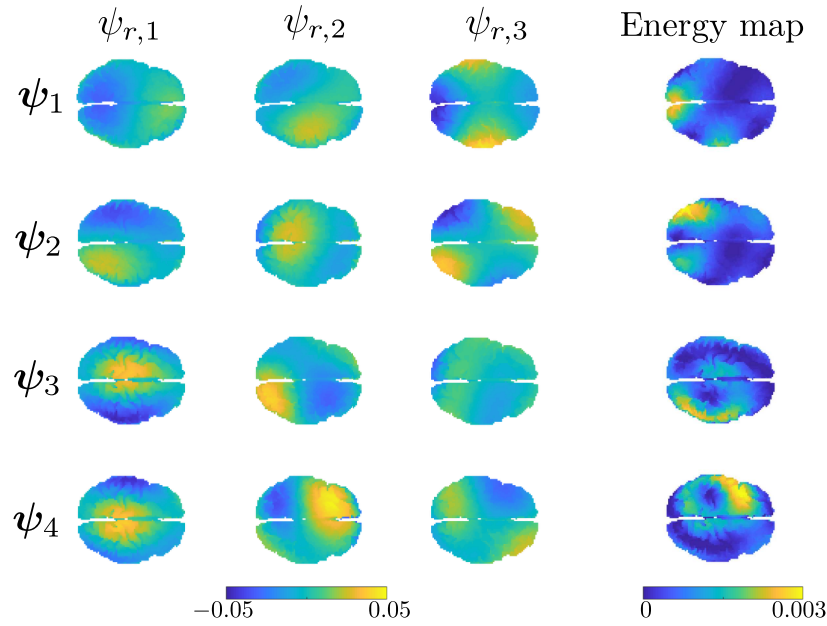


Figure 5. From top to bottom the components and the energy maps of the PC functions ψ_1, \dots, ψ_4 .

a location v_j associated with 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 $241 \times 24k$ matrix.

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 (5). Specifically, on $\mathcal{M}_{\mathcal{T}}$, we construct the four $L^2(\mathcal{M}_{\mathcal{T}})$ orthonormal vector-valued functions $\{\psi_r = (\psi_{r,1}, \psi_{r,2}, \psi_{r,3}) : r = 1, \dots, 4\}$, with $\psi_r : \mathcal{M}_{\mathcal{T}} \rightarrow \mathbb{R}^3$. These represent the PC functions to be estimated. In figure 5 we show the four components of $\{\psi_r\}$ and the associated energy maps $\{\|\psi_r(v)\|^2 : v \in \mathcal{M}_{\mathcal{T}}\}$, with $\|\cdot\|$ denoting the Euclidean norm in \mathbb{R}^3 . We then generate $m = 50$ smooth vector-valued functions $\{\mathbf{x}_l\}$ on $\mathcal{M}_{\mathcal{T}}$ by

$$\mathbf{x}_l = z_{l1}\psi_1 + z_{l2}\psi_2 + z_{l3}\psi_3 + z_{l4}\psi_4 \quad l = 1, \dots, m,$$

where $\{z_{lr}\}$ are i.i.d realizations of the four independent random variables $\{z_r \sim N(0, \gamma_r) : r = 1, \dots, 4\}$, with $\gamma_1 = 3^2$, $\gamma_2 = 2.5^2$, $\gamma_3 = 2^2$ and $\gamma_4 = 1$.

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

In the following, we compare the PC model (7) to an alternative approach that we call the naive approach. In fact, the individual functions $\{\mathbf{x}_l\}$ could be estimated from $\{\mathbf{y}_l\}$ by use of classical inverse problem estimators. Here, we adopt the estimates $\{\hat{\mathbf{x}}_l\}$ defined as

$$\hat{\mathbf{x}}_l = \underset{\substack{\mathbf{f}=(f_1, f_2, f_3): \\ f_1, f_2, f_3 \in H^2(\mathcal{M})}}{\operatorname{argmin}} \sum_{l=1}^m \|\mathbf{y}_l - K\Psi\mathbf{f}\|^2 + \lambda \int_{\mathcal{M}} \|\Delta_{\mathcal{M}}\mathbf{f}\|^2, \quad l = 1, \dots, m, \quad (20)$$

where each $\hat{\mathbf{x}}_l$ is defined in such a way that it balances the fitting term and the regularization term in (20). Due to the fact that \mathbf{f} is vector-valued, $\|\Delta_{\mathcal{M}}\mathbf{f}\|^2$ is defined as

$$\|\Delta_{\mathcal{M}}\mathbf{f}\|^2 = \Delta_{\mathcal{M}}^2 f_1 + \Delta_{\mathcal{M}}^2 f_2 + \Delta_{\mathcal{M}}^2 f_3,$$

with $\{f_q : q = 1, 2, 3\}$ denoting the components of \mathbf{f} . The same penalty operator is also adopted to generalize to vector-valued functions the PC models introduced in sections 3 and 4. In this approach, the constant λ is chosen independently for each of the m functions by partitioning the 241 detectors in roughly equally sized $K = 2$ groups and applying K -fold cross-validation. The criterion for the optimal λ is the average reconstruction error, on the sensors space, computed on the validation groups. Once we obtain the estimates $\{\hat{\mathbf{x}}_l\}$ we can compute the estimated PC functions $\{\psi_r\}$ by applying classical multivariate PC analysis on the reconstructed objects $\hat{\mathbf{x}}_l$.

The estimates are compared to those of the proposed PC function model, as described in algorithm 1, with 15 iterations. Note that, instead, a tolerance could be fixed to test if the algorithm has converged. However, 15 iterations give satisfactory convergence levels in our simulations and application studies. We partition the m observations in equally sized $K = 2$ groups and perform K -fold cross-validation for the choice of the penalty. Specifically, we choose the coefficient λ that minimizes the sensors space reconstruction error, on the validation groups.

To evaluate the performances of the two approaches, we generate 100 datasets as previously detailed. The quality of the estimated r th PC function is then measured with $E(\psi_r, \hat{\psi}_r) = \sum_{q=1}^3 \|\nabla_{\mathcal{M}}(\psi_{r,q} - \hat{\psi}_{r,q})\|^2$. The results are summarized in the boxplots in figure 6, for two different signal-to-noise ratios, where the Gaussian noise has standard deviation $\sigma = 5$ and $\sigma = 10$. In figure 7 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 (i.e. lower estimation errors $E(\psi_r, \hat{\psi}_r)$), when compared to the naive approach. Differences in the estimation error are higher in a low signal-to-noise regime, as it is for the estimation of the fourth PC function, where intuitively, the low variance associated to the PC function makes it more difficult to distinguish this structured signal from the noise component. Also surprising is the stability of the estimates of the proposed algorithm across the generated datasets, as opposed to the naive approach of reconstructing the functional observations independently, which instead returns multiple particularly unsatisfactory reconstructions. An example of such reconstructions is shown in figure 8.

5.2. Indirectly observed covariance functions

In this section, we consider $\mathcal{M}_{\mathcal{T}}$ to be a 8k nodes triangular mesh, this time representing a template geometry of the cortical surface, which is shown in figure 10. 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

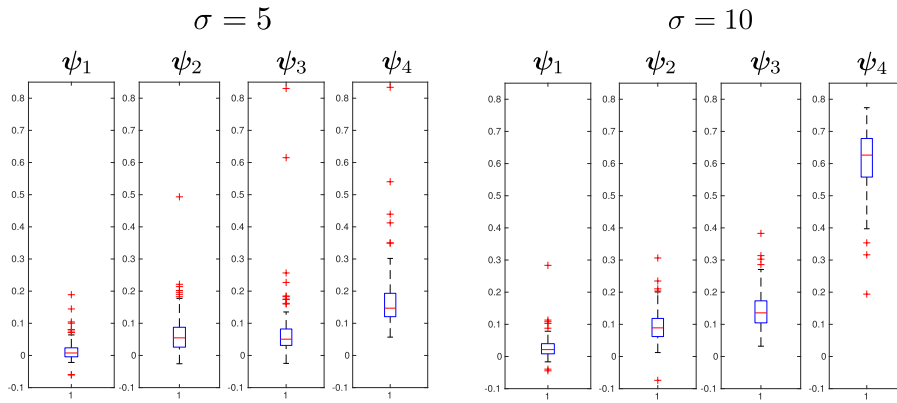


Figure 6. On the left, a summary of the results in a medium signal-to-noise ratio regime. On the right, a summary of the results in a low signal-to-noise ratio regime. Each boxplot displays the paired differences of the estimation errors $E(\psi_r, \hat{\psi}_r)$ between the estimates of the two steps naive method and those obtained by applying algorithm 1. A paired difference greater than 0 indicates that, for the dataset in question, algorithm 1 has performed better than the two steps naive approach.

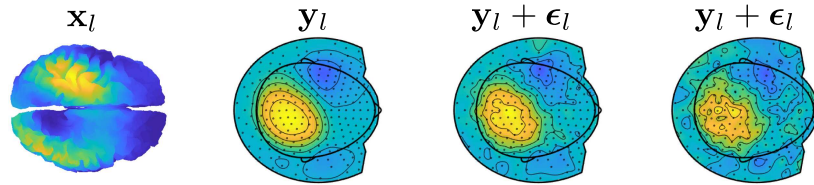


Figure 7. From left to right, the energy map of a generated function \mathbf{x}_l , the associated signal \mathbf{y}_l on the sensors space with respectively no additional error, Gaussian error of standard deviation $\sigma = 5$ and Gaussian error of standard deviation $\sigma = 10$.

241 detectors on the sensors space, and the head model, to compute a forward operator K_i for the i th subject. The 8k nodes of each subject's triangular mesh are in correspondence with the 8k nodes of the template mesh $\mathcal{M}_{\mathcal{T}}$. This allows the model to be defined on the template $\mathcal{M}_{\mathcal{T}}$.

As in the previous section, we construct four $L^2(\mathcal{M}_{\mathcal{T}})$ orthonormal functions $\{\psi_r = (\psi_{r,1}, \psi_{r,2}, \psi_{r,3}) : r = 1, \dots, 4\}$. The energy maps of $\{\psi_r\}$ are shown in figure 9. We generate synthetic data from model (6) as follows:

$$C_i = \sum_{r=1}^4 z_{ir}^2 \psi_r \otimes \psi_r = \sum_{r=1}^4 z_{ir}^2 \begin{bmatrix} \psi_{r,1} \otimes \psi_{r,1} & \psi_{r,1} \otimes \psi_{r,2} & \psi_{r,1} \otimes \psi_{r,3} \\ \psi_{r,2} \otimes \psi_{r,1} & \psi_{r,2} \otimes \psi_{r,2} & \psi_{r,2} \otimes \psi_{r,3} \\ \psi_{r,3} \otimes \psi_{r,1} & \psi_{r,3} \otimes \psi_{r,2} & \psi_{r,3} \otimes \psi_{r,3} \end{bmatrix},$$

where z_{i1}, \dots, z_{i4} are i.i.d realizations of the four independent random variables $\{z_r \sim N(0, \gamma_r) : r = 1, \dots, 4\}$, with $\gamma_1 = 3^2$, $\gamma_2 = 2.5^2$, $\gamma_3 = 2^2$ and $\gamma_4 = 1$. 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

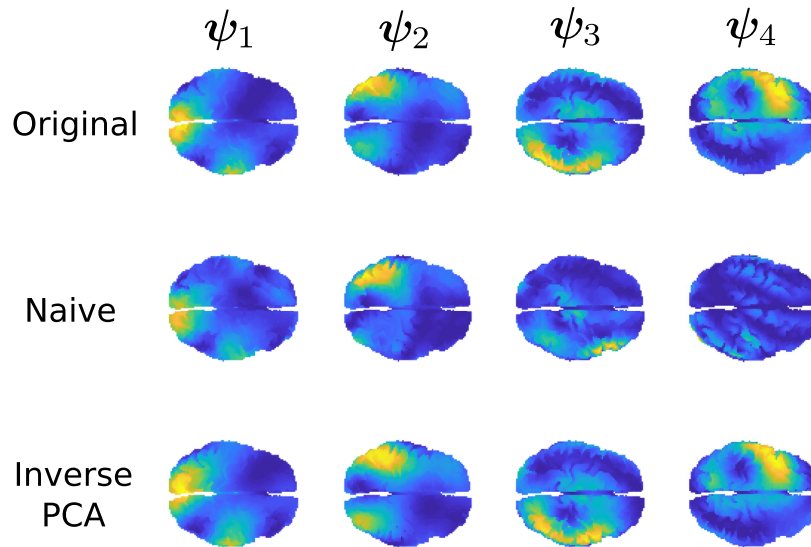


Figure 8. On the first row the energy maps of the true four PC components to be estimated, on the second row the estimations given by the two steps naive method, and on the third row the reconstructions obtained by applying algorithm 1.

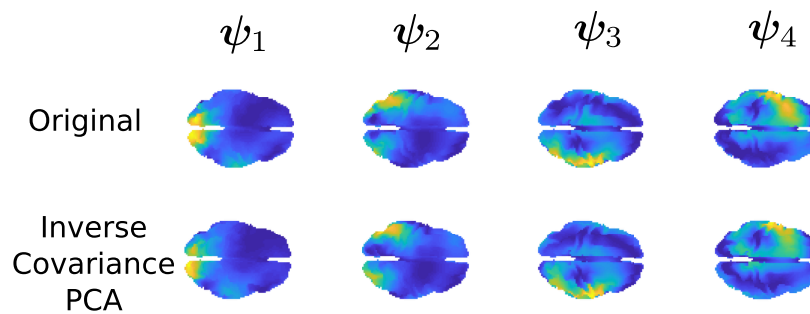


Figure 9. On the top row, the energy maps of ψ_1, \dots, ψ_4 . On the bottom row the energy maps of the estimates $\hat{\psi}_1, \dots, \hat{\psi}_4$ obtained by applying algorithm 2.

defined as

$$S_i = K_i C_i K_i^T + E_i^T E_i, \quad i = 1, \dots, n.$$

The term $E_i^T E_i$ is an error term, where E_i is an $s \times s$ matrix with each entry that is an independent sample from a Gaussian distribution with mean zero and standard deviation 5. We then apply algorithm 2 with 15 iterations, feeding in input $\{S_i\}$. The results are shown in figure 9, in terms of energy maps of the reconstructed functions $\{\hat{\psi}_r\}$. These are a close approximation of the underlying functions $\{\psi_r\}$. The fidelity measure $\sum_{q=1}^3 \|\nabla_{\mathcal{M}}(\psi_{r,q} - \hat{\psi}_{r,q})\|^2$ of such estimates is 6.8×10^{-2} , 6.1×10^{-1} , 6.8×10^{-1} and 7.4×10^{-1} , for ψ_1, \dots, ψ_4 respectively, which is comparable in term of order of magnitude to the results obtained in the case of PCs

of indirectly observed functions. Across the generation of multiple datasets, results are stable, with the exception of few situations where the cross-validation approach suggests a penalization coefficient 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 many other options have been proposed in the inverse problems literature (Vogel 2002). Some of these, however, involve visual inspection.

6. Application

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

6.1. Pre-processing

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

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

Moreover, a surface based registration algorithm has been applied to register each of the extracted cortical surfaces to a triangulated template cortical surface, which is shown in figure 10. Post registration, the triangulated template cortical surface is sub-sampled to a 8k nodes surface. Moreover, the nodes on the cortical surface of each subject are also sub-sampled to a set of 8k nodes in correspondence to the 8k nodes of the template. For each subject, a $248 \times 24k$ matrix, representing the forward operator, has been computed with FieldTrip (Oostenveld *et al* 2011) from its head surface, cortical surface and sensors position. Such a 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 min resting state MEG scans have been performed, of which one session is used in our analysis. During the 6 min, 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

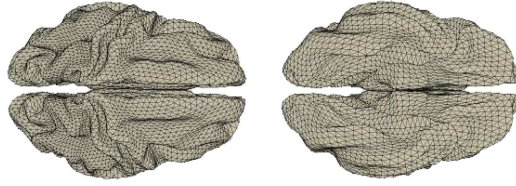


Figure 10. Top side and bottom side views of the template triangular mesh \mathcal{M}_T composed of 8k nodes.

found in the HCP MEG reference manual. Moreover, we apply a band pass filter, limiting the spectrum of the signal to the [12.5, 29] Hz, also known as the beta waves. For the signal of each channel we compute its amplitude envelope (see figure B.1) which describes the evolution of the signal amplitude. The measure of connectivity between channels that we adopt in this work is the covariance of the amplitude envelopes. Other connectivity metrics, such as phase-based metrics, have been proposed in the literature (see Colclough *et al* 2016, and references therein).

6.2. Analysis

Here we apply the population model introduced in section 4.2 to the HCP MEG data. The first part of the analysis focuses on studying dynamic functional connectivity of a specific subject. For this purpose, we subdivide the 6 min session in $n = 40$ consecutive intervals. Each of these segments is used to compute a covariance matrix in the sensors space, resulting in n covariance matrices S_1, \dots, S_n . In this setting, we have one forward operator $K = K_1 = \dots = K_n$. The aim is understanding the main modes of variation of the functional connectivity on the brain space of the subject. Thus, algorithm 2, with 20 iterations, is applied to S_1, \dots, S_n to find the PC covariance functions.

A regularization parameter λ common to all the PC components is chosen by inspecting the plot of the regularity of the first $R = 10$ PC covariance functions ($\sum_{r=1}^R \int_{\mathcal{M}} \|\nabla \psi_r\|^2$) versus the residual norm, for different choices of the parameter. This is a version of the L-curve plot (Hansen 2000) and is shown on the left panel of figure B.2. Here we show the results for $\lambda = 10^2$, in the appendices we show the results for $\lambda = 10$. The energy maps of the estimated $\hat{\psi}_1, \hat{\psi}_2$ and $\hat{\psi}_3$ resulting from the analysis are shown in figure 11. These are associated with the first three PC covariance functions $\hat{\psi}_1 \otimes \hat{\psi}_1, \hat{\psi}_2 \otimes \hat{\psi}_2$ and $\hat{\psi}_3 \otimes \hat{\psi}_3$. High intensity areas, in yellow, indicate which areas present high average interconnectivity, either by means of positive or negative correlation in time.

In figure 12, we show the plot of variances associated with each time segment, describing the variation in time of the PC covariance functions, hence the variation in interconnectivity. The variance can be either defined on the sensors space, by normalizing the PC covariance functions $\{K\hat{\psi}_r\}$, with K the forward operator, or on the brain space, by normalizing the PC covariance functions on the brain space $\{\hat{\psi}_r\}$. Due to the presence of invisible dipoles, which are dipoles that display zero magnetic field on the sensors space, the two norms can be quite different, leading to different average variances for each PC covariance function. Due to the high sensitivity of the source space variances on the choice of the regularization parameter, we focus on the estimated variances on the sensors space.

We have also applied our model to the covariances obtained by subdividing the MEG session in $n = 80$ segments. As expected the PC covariance functions, shown in figure B.5 are very similar. However, the variances, in figure B.4, show higher variability in time, which can be

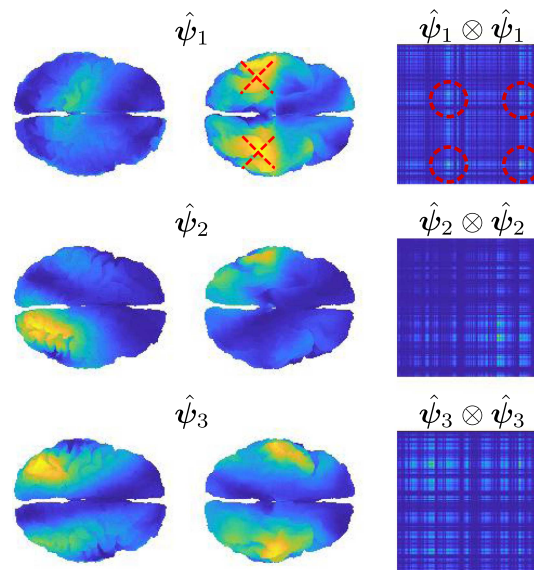


Figure 11. Top side and bottom side views of the estimated energy maps $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying algorithm 2 to the covariance matrices computed from the MEG resting state data of a single subject on $n = 40$ consecutive time intervals. On the right panel, the covariance functions associated with these energy maps. On the top right panel we highlight with red circles the areas with high average interconnectivity, which correspond to the neighborhoods of the red crossed vertices in the plot of the energy map of ψ_1 .

partially explained by the fact that shorter time segments lead to covariance estimates that have higher variability.

The second part of the analysis focuses on applying the proposed methodology to a multi-subject setting. Specifically, $n = 40$ different subjects are considered. For each subject, the 6 min scan is used to compute a covariance matrix, resulting in n covariance matrices S_1, \dots, S_n . The template geometry in figure 10 is used as a model of the brain space. Algorithm 2 is then applied to find the PC covariance functions on the template brain, associated with S_1, \dots, S_n . We run the algorithm for 20 iterations, and choose the regularizing parameter to be $\lambda = 10^2$ by inspecting the L-curve plot in the right panel of figure B.2. The results for $\lambda = 10$ are shown in the appendices. The energy maps of the estimated functions $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ and the associated first three covariance functions $\hat{\psi}_1 \otimes \hat{\psi}_1$, $\hat{\psi}_2 \otimes \hat{\psi}_2$ and $\hat{\psi}_3 \otimes \hat{\psi}_3$, are shown in figure 13. High intensity areas, in yellow, indicate which areas present high average connectivity. In figure 14, we show the subject-specific associated variances, both in the sensors space and the brain space.

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 in both the single and multi-subject settings, the areas with high interconnectivity, displayed in yellow in figures 11 and 13, seem to be at least partially overlapping with the brain's default network (Buckner *et al* 2008, Yeo *et al* 2011). The brain's default network consists of the brain regions known to have highly correlated hemodynamic activity (i.e. highest functional connectivity levels), and to be most active, when the subject is not performing

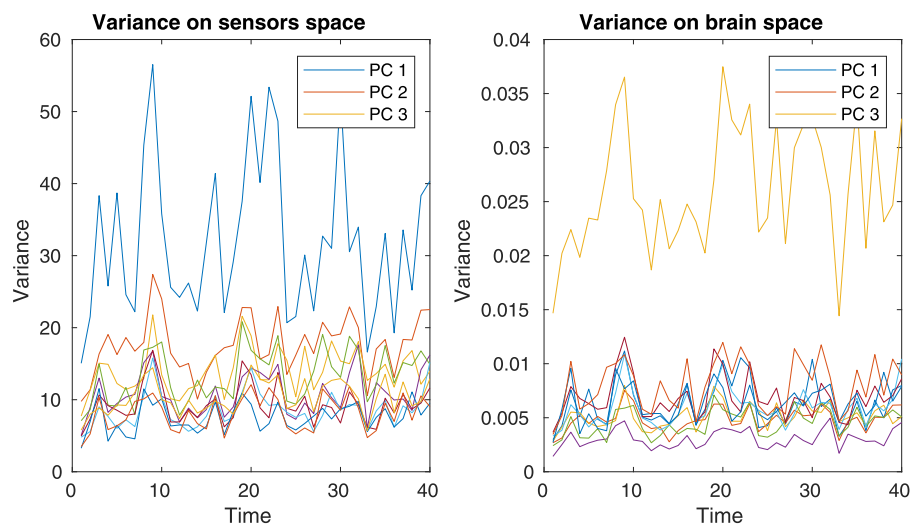


Figure 12. Plots of the segment-specific variances of the first $R = 10$ PC covariance functions. On the left, the estimated variances on the sensors space, on the right, the estimated variances on the brain space.

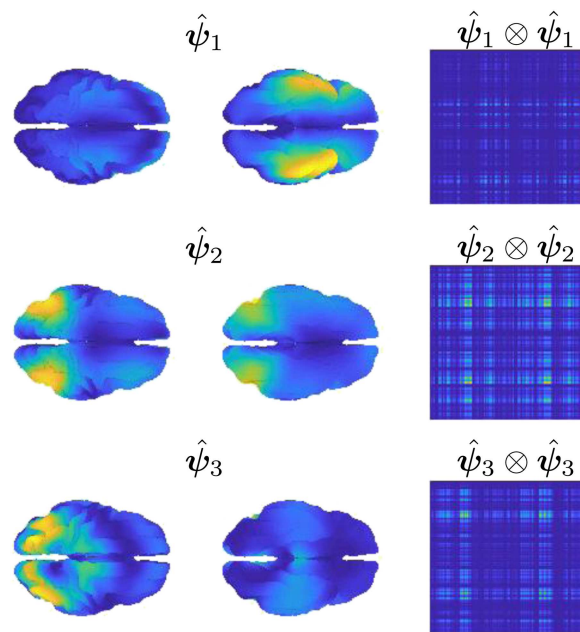


Figure 13. Top side and bottom side views of the estimated energy maps $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying algorithm 2 to the covariance matrices computed from the MEG resting state data of $n = 40$ different subjects. On the right panel, the covariance functions associated with these energy maps.

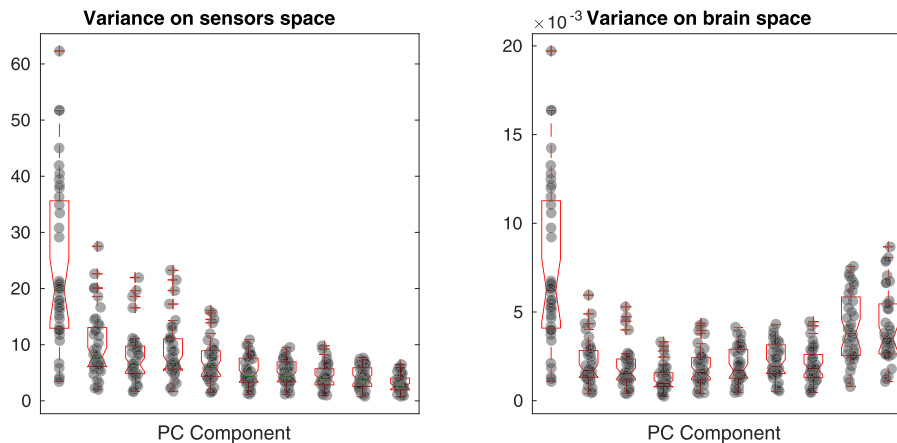


Figure 14. Plots of the subject-specific variances associated with the first $R = 10$ PC covariance functions. On the left, the estimated variances on the sensors space, on the right, the estimated variances on the brain space.

any specific task. An image of the spatial configuration of the default network can be found, for instance, in figure 2 of Buckner *et al* (2008). From the plots of the associated variances in the sensors space (left panel of figures 12 and 14) we can see that these areas are also the ones that show high variability in connectivity across time or across subjects. This might suggest that the brain's default network is also the brain region that shows among the highest levels of spontaneous variability in connectivity.

The plots of the variances on the brain space (right panel of figure 14), when compared to those on the sensors space (left panel of figure 14), demonstrate that these type of studies are highly sensitive to the choice of the regularization, not only in terms of spatial configuration of the results, but also in terms of estimated variances on the brain space. With a naive 'first reconstruct and then analyze' approach, where the reconstructed data on the brain space replace those observed on the sensors space, this issue could go unnoticed, as the variability that does not fit the chosen model is implicitly discarded in the reconstruction step and does not appear in the subsequent analysis. Also, importantly, our analysis deals with statistical samples that are entire covariances, overcoming the limitations of seed-based approaches, where prior spatial information is required to choose the seed. Seed locations are usually informed by fMRI studies and this comes with the risk of biasing the analysis when comparing electrophysiological networks (MEG) and hemodynamic networks (fMRI).

In general, care should be taken when drawing conclusions from MEG studies. Establishing static and dynamic functional connectivity from MEG data remains challenging, due to the strong ill-posedness of the inverse problem. It is known that other variables, such as the choice of the frequency band or the choice of the connectivity metric can influence the analysis. While the choice of the neural oscillatory frequency band could be seen as an additional parameter in MEG functional connectivity studies, there is no general agreement on the choice of the connectivity metrics (Gross *et al* 2013). It is important to highlight that in this paper we focus on methodological contributions to the specific problem of reconstructing and representing indirectly observed functional images and covariance functions.

7. Discussion

In this work we introduce a general framework for the reconstruction and representation 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 constraint 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 perspective, it is of interest to apply them to different settings, not necessarily involving neuroimaging, where studying second order information has been so far prohibitive. Direct examples are second order analysis of the dynamics of meteorological observations, such as temperature. Another possible application is the study of the dynamics of ocean currents, where the irregularity of the spatial domain, and its complex boundaries, can be easily accounted for thanks to the manifold representation approach in our models.

From a modeling point of view, it is of interest to take a step further toward the integration of the inverse problems literature with the approach we adopt in this paper. For instance, penalization terms that have been shown to be successful in the inverse problems literature, e.g. total variation penalization, could be introduced in our models.

Acknowledgments

The authors would like to thank the anonymous reviewers and the member of the Editorial Board for their useful and constructive comments.

Appendix A. Discrete solutions

Proof of proposition 1. 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 (7):

$$\begin{aligned} & \sum_{l=1}^m \|\mathbf{y}_l - z_l K_l \Psi f\|^2 + \lambda \mathbf{z}^T \mathbf{z} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f \\ & \propto (\Psi f)^T \left(\sum_{l=1}^m z_l^2 K_l^T K_l \right) \Psi f - 2(\Psi f)^T \left(\sum_{l=1}^m z_l K_l^T \mathbf{y}_l \right) + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \end{aligned} \quad (21)$$

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

$$(\Psi \varphi)^T \left(\sum_{l=1}^m z_l^2 K_l^T K_l \right) \Psi \hat{f} + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}} \varphi \Delta_{\mathcal{M}} \hat{f} = (\Psi \varphi)^T \left(\sum_{l=1}^m z_l K_l^T \mathbf{y}_l \right) \quad (22)$$

for every $\varphi \in H^2(\mathcal{M})$ (see Braess 2007, chapter 2). Moreover, such minimizer is unique if $A(\varphi, f) = (\Psi \varphi)^T \left(\sum_{l=1}^m z_l^2 K_l^T K_l \right) \Psi f + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}} \varphi \Delta_{\mathcal{M}} f$ is positive definite. Given that for a closed manifold \mathcal{M} , $\int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f = 0$ iff f is a constant function (Dziuk and Elliott 2013), the

positive definiteness condition is equivalent to assuming that $\ker(\sum_{l=1}^m z_l^2 K_l^T K_l)$, the kernel of $\sum_{l=1}^m z_l^2 K_l^T K_l$, does not contain the subspace of p -dimensional constant vectors.

Moreover, we can reformulate equation (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 $\hat{f}, g \in V$

$$\begin{cases} (\Psi\varphi)^T \left(\sum_{l=1}^m z_l^2 K_l^T K_l \right) \Psi \hat{f} + \lambda \int_{\mathcal{M}} \nabla_{\mathcal{M}} \varphi \cdot \nabla_{\mathcal{M}} g = (\Psi\varphi)^T \left(\sum_{l=1}^m z_l K_l^T \mathbf{y}_l \right) \\ \int_{\mathcal{M}} \nabla_{\mathcal{M}} \hat{f} \cdot \nabla_{\mathcal{M}} w - \int_{\mathcal{M}} g w = 0 \end{cases} \quad (23)$$

for all $\varphi, w \in V$. The operator $\nabla_{\mathcal{M}}$ is the gradient operator on the manifold \mathcal{M} . The gradient operator $\nabla_{\mathcal{M}}$ is such that $(\nabla_{\mathcal{M}} w)(v)$, for w a smooth real function on \mathcal{M} and $v \in \mathcal{M}$, takes value on the tangent space at v . We denote with \cdot the scalar product on the tangent space.

We recall here the definition of the $\kappa \times \kappa$ matrices to be $(M)_{jj'} = \int_{\mathcal{M}_T} \phi_j \phi_{j'}$ and $(A)_{jj'} = \int_{\mathcal{M}_T} \nabla_{\mathcal{M}_T} \phi_j \cdot \nabla_{\mathcal{M}_T} \phi_{j'}$. Note that requiring (23) to hold for all $\varphi, w \in V$ is equivalent to requiring that (23) holds for all φ, w that are basis elements of V , thus exploiting the basis expansion formula (11) we can characterize (23) with the solution of the linear system

$$\begin{bmatrix} \sum_{l=1}^m z_l^2 K_l^T K_l & \lambda A \\ A & -M \end{bmatrix} \begin{bmatrix} \hat{\mathbf{c}} \\ \hat{\mathbf{q}} \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^m z_l K_l^T \mathbf{y}_l \\ \mathbf{0} \end{bmatrix}, \quad (24)$$

where $\hat{\mathbf{c}}$ and $\hat{\mathbf{q}}$ are the basis coefficients of $f \in V$ and $g \in V$, respectively. Solving (24) in $\hat{\mathbf{c}}$ leads to

$$\left(\sum_{l=1}^m z_l^2 K_l^T K_l + \lambda A M^{-1} A \right) \hat{\mathbf{c}} = \sum_{l=1}^m z_l K_l^T \mathbf{y}_l. \quad (25)$$

□

Proof of proposition 2. 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 (14):

$$\begin{aligned} & \sum_{i=1}^n \|S_i^{1/2} - \mathbf{z}_i (K_i \Psi f)^T\|^2 + \lambda \sum_{i=1}^n \|\mathbf{z}_i\|^2 \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f \\ & \propto (\Psi f)^T \left(\sum_{i=1}^n \|\mathbf{z}_i\|^2 K_i^T K_i \right) \Psi f - 2(\Psi f)^T \sum_{i=1}^n K_i^T S_i^T \mathbf{z}_i. \end{aligned} \quad (26)$$

Comparing (26) with (21) it is evident that by following the same steps of the proof of proposition 1 we obtain the desired result, which is

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

□

Appendix B. Application—additional material

Here we present further material complementing the analysis in section 6. In figure B.1 we show the amplitude envelope computed from a filtered version of a signal detected by an MEG sensor. The covariance of the amplitude envelopes across different sensors is the measure of connectivity used in this work.

In figure B.2 we show the L-curve plots associated with the PC covariance models applied to the dynamic and multi-subject functional connectivity studies.

In figures B.3 and B.4 we show respectively the plots of the estimated PC covariance functions and associated variances from the dynamic functional connectivity study on $n = 40$ segments with regularization parameter $\lambda = 10$.

In figures B.5 and B.6 we show the estimated PC covariance functions and associated variances from the dynamic functional connectivity study on $n = 80$ time segments with regularization parameter $\lambda = 10^2$.

In figures B.7 and B.8 we show the estimated PC covariance functions and associated variances from the multi-subject functional connectivity study on $n = 40$ subjects with regularization parameter $\lambda = 10$.

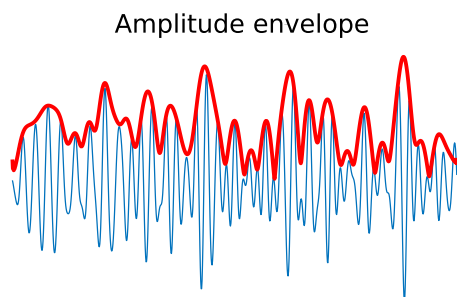


Figure B.1. Amplitude envelope (in red) of the filtered signal (in blue) detected by an MEG sensor.

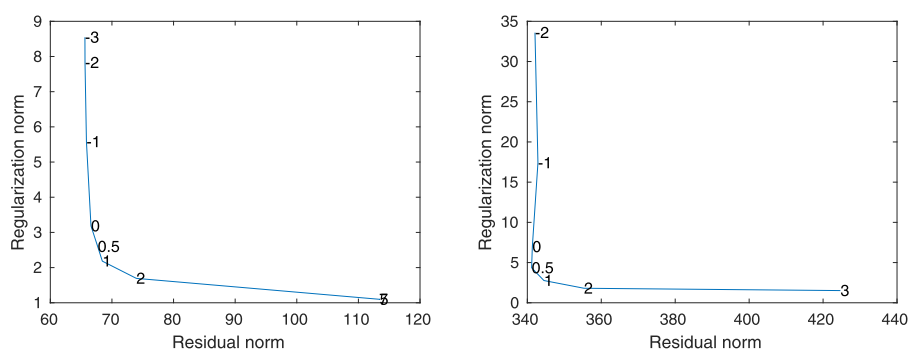


Figure B.2. Plots of the regularity of the first $R = 10$ PC covariance functions, measured as $\sum_{r=1}^{10} \int_{\mathcal{M}} \|\nabla \psi_r\|^2$ versus the residual norm in the data, for different choices of $\log(\lambda)$. On the left panel, the plot refers to the dynamic connectivity study, on the right panel the plot of the multi-subject connectivity study.

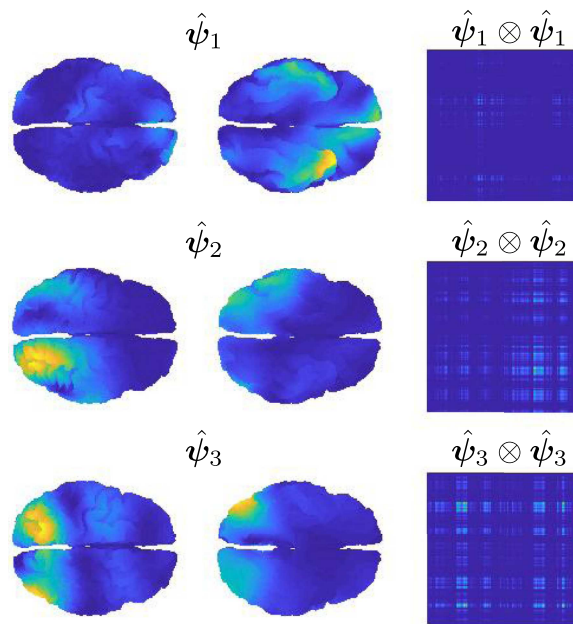


Figure B.3. Energy maps of the estimated $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying algorithm 2, with lower regularization ($\lambda = 10$), to the covariance matrices computed from the MEG resting state data of a single subject on $n = 40$ consecutive time intervals. On the right panel, the covariance functions associated with these energy maps.

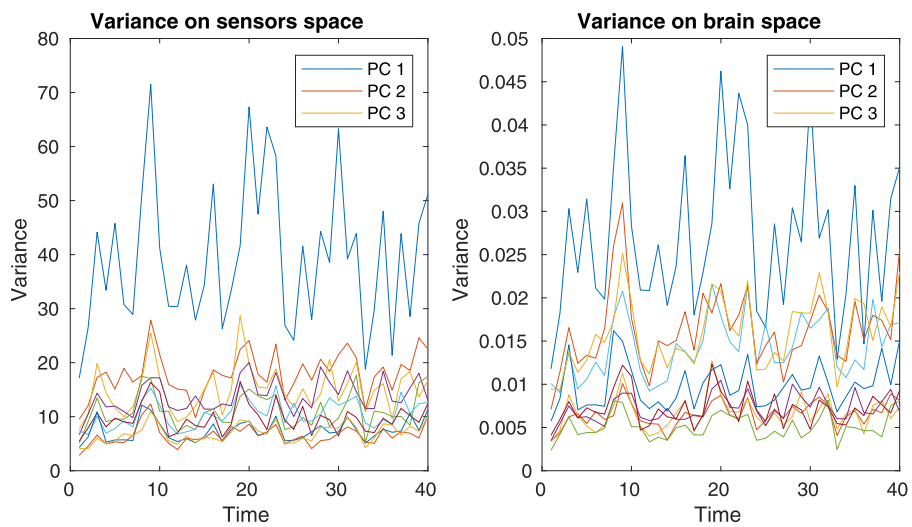


Figure B.4. Plots of the segment-specific variances of the first $R = 10$ PC covariance functions in time when a smaller regularization parameter is chosen ($\lambda = 10$).

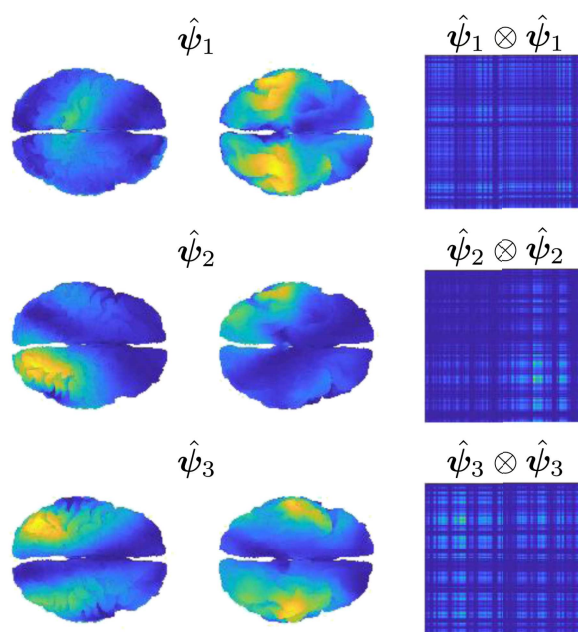


Figure B.5. Energy maps of the estimated $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying algorithm 2, with $\lambda = 10^2$, to the covariance matrices computed from the MEG resting state data of a single subject on $n = 80$ consecutive time intervals. On the right panel, the covariance functions associated with these energy maps.

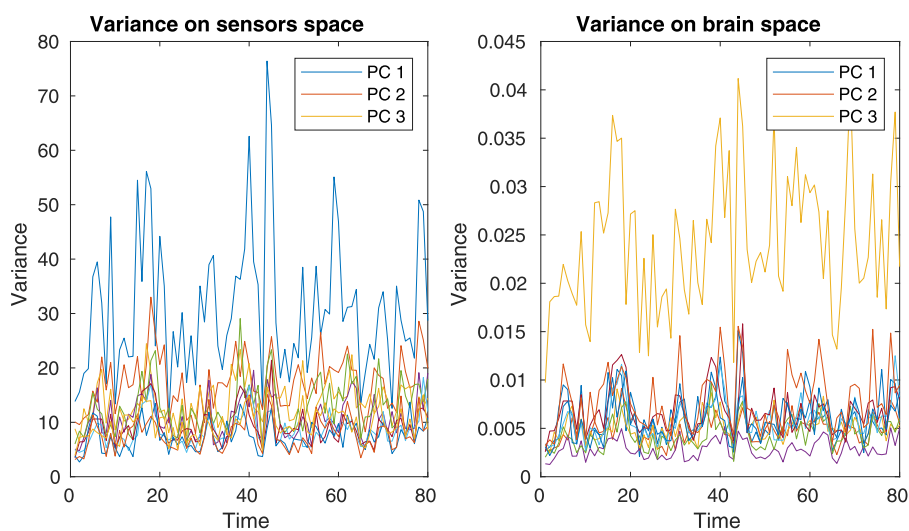


Figure B.6. Plots of the segment-specific variances of the first $R = 10$ PC covariance functions in time, with $\lambda = 10^2$, when the MEG resting state data is split into $n = 80$ consecutive time intervals.

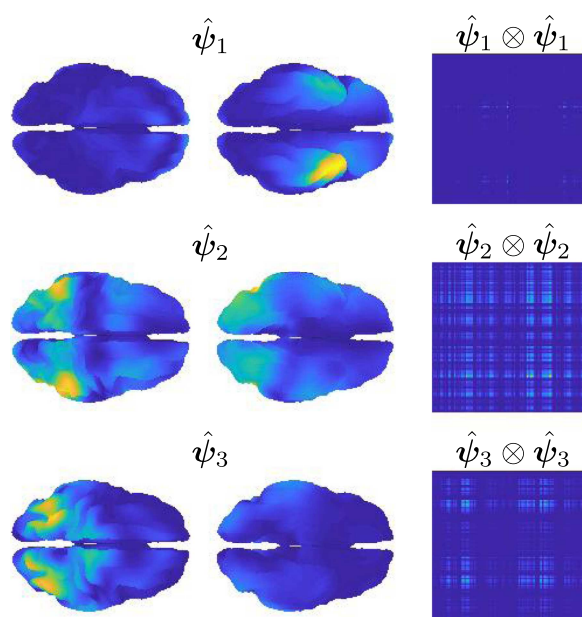


Figure B.7. Energy maps of the estimated $\hat{\psi}_1$, $\hat{\psi}_2$ and $\hat{\psi}_3$ obtained by applying algorithm 2, with lower regularization ($\lambda = 10$), to the covariance matrices computed from the MEG resting state data of $n = 40$ different subjects. On the right panel, the covariance functions associated with these energy maps.

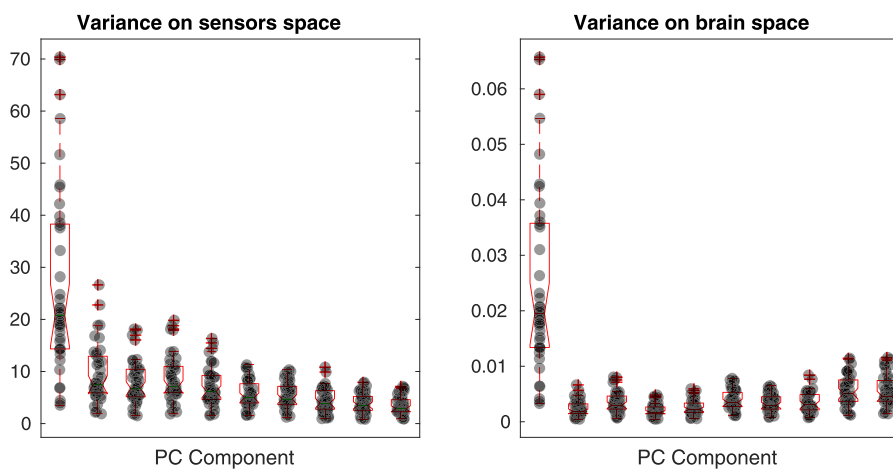


Figure B.8. Plots of the subject-specific variances associated with the first $R = 10$ PC covariance functions computed from $n = 40$ subject, with regularization parameter $\lambda = 10$.

ORCID iDs

Eardi Lila  <https://orcid.org/0000-0002-9199-392X>

Simon Arridge  <https://orcid.org/0000-0003-1292-0210>

References

- Adorf H M 1995 Hubble space telescope image restoration in its fourth year *Inverse Problems* **11** 639–53
- Amini A A and Wainwright M J 2012 Sampled forms of functional PCA in reproducing kernel Hilbert spaces *Ann. Stat.* **40** 2483–510
- Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–93
- Arridge S R, Kaipio J P, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography *Inverse Problems* **22** 175–95
- Azzimonti L, Nobile F, Sangalli L M and Secchi P 2014 Mixed finite elements for spatial regression with PDE penalization *SIAM/ASA J. Uncertain. Quantification* **2** 305–35
- Beck A and Teboulle M 2009 A fast iterative shrinkage-thresholding algorithm for linear inverse problems *SIAM J. Imaging Sci.* **2** 183–202
- Benko M, Härdle W and Kneip A 2009 Common functional principal components *Ann. Stat.* **37** 1–34
- Boyd S, Parikh N, Chu E, Peleato B and Eckstein J 2010 Distributed optimization and statistical learning via the alternating direction method of multipliers *Found. Trends Mach. Learn.* **3** 1–122
- Braess D 2007 *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics* (Cambridge: Cambridge University Press)
- Buckner R L, Andrews-Hanna J R and Schacter D L 2008 The Brain’s default network *Ann. New York Acad. Sci.* **1124** 1–38
- Bunea F and Xiao L 2015 On the sample covariance matrix estimator of reduced effective rank population matrices, with applications to fPCA *Bernoulli* **21** 1200–30
- Burger M, Sawatzky A and Steidl G 2016 First order algorithms in variational image processing *Splitting Methods in Communication, Imaging, Science, and Engineering* eds Glowinski R, Osher S and Yin W (Cham: Springer) 345–407
- Calderón A 1980 On an inverse boundary value problem *Seminar on Numerical Analysis and Its Applications to Continuum Physics* eds W H Meyer and M A Raupp (Rio de Janeiro: Soc. Brasileira de Matematica) pp 65–73
- Calvetti D and Somersalo E 2007 *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing* (Surveys and Tutorials in the Applied Mathematical Sciences vol 2) (New York, NY: Springer)
- Cavalier L 2008 Nonparametric statistical inverse problems *Inverse Problems* **24** 034004
- Chambolle A and Pock T 2011 A first-order primal-dual algorithm for convex problems with applications to imaging *J. Math. Imaging Vis.* **40** 120–45
- Chambolle A and Pock T 2016 An introduction to continuous optimization for imaging *Acta Numer.* **25** 161–319
- Colclough G, Woolrich M, Tewarie P, Brookes M, Quinn A and Smith S 2016 How reliable are MEG resting-state connectivity metrics? *NeuroImage* **138** 284–93
- 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 *J. Mach. Learn. Res.* **6** 883–904
- Dobriban E, Leeb W and Singer A 2017 Optimal prediction in the linearly transformed spiked model arXiv:1709.03393
- Dryden I L, Koloydenko A and Zhou D 2009 Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging *Ann. Appl. Stat.* **3** 1102–23
- Dziuk G and Elliott C M 2013 Finite element methods for surface PDEs *Acta Numer.* **22** 289–396
- Eggebrecht A T, Ferradal S L, Robichaux-Viehoever A, Hassanpour M S, Dehghani H, Snyder A Z, Hershey T and Culver J P 2014 Mapping distributed brain function and networks with diffuse optical tomography *Nat. Photon.* **8** 448–54
- Ferrari M and Quaresima V 2012 A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application *NeuroImage* **63** 921–35

- Flury B N 1984 Common principal components in k groups *J. Am. Stat. Assoc.* **79** 892–8
- Fransson P, Åden U, Blennow M and Lagercrantz H 2011 The functional architecture of the infant brain as revealed by resting-state fMRI *Cereb. Cortex* **21** 145–54
- Fried I and Malkus D S 1975 Finite element mass matrix lumping by numerical integration with no convergence rate loss *Int. J. Solids Struct.* **11** 461–6
- Friedman J, Hastie T and Tibshirani R 2008 Sparse inverse covariance estimation with the graphical lasso *Biostatistics* **9** 432–41
- Geman D 1990 Random fields and inverse problems in imaging *École d'Été de Probabilités de Saint-Flour XVIII—1988* (Lecture Notes in Mathematics) ed P-L Hennequin (Berlin: Springer) pp 115–93
- Golub G H and van Loan C F 1980 An analysis of the total least squares problem *SIAM J. Numer. Anal.* **17** 883–93
- Gross J et al 2013 Good practice for conducting and reporting MEG research *NeuroImage* **65** 349–63
- Gutta S, Bhatt M, Kalva S K, Pramanik M and Yalavarthy P K 2019 Modeling errors compensation with total least squares for limited data photoacoustic tomography *IEEE J. Sel. Top. Quantum Electron.* **25** 1–14
- Hansen P C 2000 The L-curve and its use in the numerical treatment of inverse problems *Computational Inverse Problems in Electrocardiology* (Advances in Computational Bioengineering) ed P Johnston (Southampton: WIT Press) pp 119–42
- Hartley R and Zisserman A 2004 *Multiple View Geometry in Computer Vision* (Cambridge: Cambridge University Press)
- Hu Y and Jacob M 2012 Higher degree total variation (HDTV) regularization for image recovery *IEEE Trans. Image Process.* **21** 2559–71
- Huang J Z, Shen H and Buja A 2008 Functional principal components analysis via penalized rank one approximation *Electron. J. Stat.* **2** 678–95
- Hutchison R M et al 2013 Dynamic functional connectivity: promise, issues, and interpretations *NeuroImage* **80** 360–78
- Jin K H, McCann M T, Froustey E and Unser M 2017 Deep convolutional neural network for inverse problems in imaging *IEEE Trans. Image Process.* **26** 4509–22
- Johnstone I M and Silverman B W 1990 Speed of estimation in Positron emission tomography and related inverse problems *Ann. Stat.* **18** 251–80
- Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (Applied Mathematical Sciences vol 160) (New York, NY: Springer)
- Katsevich E, Katsevich A and Singer A 2015 Covariance matrix estimation for the cryo-EM heterogeneity problem *SIAM J. Imaging Sci.* **8** 126–85
- Kluth T and Maass P 2017 Model uncertainty in magnetic particle imaging: nonlinear problem formulation and model-based sparse reconstruction *Int. J. Magn. Part. Imaging* **3** 1707004
- Lee H-L, Zahneisen B, Hugger T, LeVan P and Hennig J 2013 Tracking dynamic resting-state networks at higher frequencies using MR-encephalography *NeuroImage* **65** 216–22
- Lefkimiatis S, Bourquard A and Unser M 2012 Hessian-based norm regularization for image restoration with biomedical applications *IEEE Trans. Image Process.* **21** 983–95
- Lehikoinen A, Finsterle S, Voutilainen A, Heikkinen L M, Vauhkonen M and Kaipio J P 2007 Approximation errors and truncation of computational domains with application to geophysical tomography *Inverse Problems Imaging* **1** 371–89
- Li K, Guo L, Nie J, Li G and Liu T 2009 Review of methods for functional brain connectivity detection using fMRI *Comput. Med. Imaging Graph.* **33** 131–9
- Lila E, Aston J A D and Sangalli L M 2016 Smooth principal component analysis over two-dimensional manifolds with an application to neuroimaging *Ann. Appl. Stat.* **10** 1854–79
- Liu X and Zhang N 2019 Sparse inverse covariance matrix estimation via the l_0 -norm with Tikhonov regularization *Inverse Problems* **35** 115010
- Lustig M, Donoho D, Santos J and Pauly J 2008 Compressed sensing MRI *IEEE Signal Process. Mag.* **25** 72–82
- Mathé P and Pereverzev S V 2006 Regularization of some linear ill-posed problems with discretized random noisy data *Math. Comput.* **75** 1913–29
- McCann M T, Jin K H and Unser M 2017 Convolutional neural networks for inverse problems in imaging: a review *IEEE Signal Process. Mag.* **34** 85–95

- Mosher J, Leahy R and Lewis P 1999 EEG and MEG: forward solutions for inverse methods *IEEE Trans. Biomed. Eng.* **46** 245–59
- Nissinen A, Heikkinen L M, Kolehmainen V and Kaipio J P 2009 Compensation of errors due to discretization, domain truncation and unknown contact impedances in electrical impedance tomography *Meas. Sci. Technol.* **20** 105504
- 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 *Comput. Intell. Neurosci.* **2011** 1–9
- Petersen A and Müller H-G 2019 Fréchet regression for random objects with Euclidean predictors *Ann. Stat.* **47** 691–719
- Pigoli D, Aston J A, Dryden I L and Secchi P 2014 Distances and inference for covariance operators *Biometrika* **101** 409–22
- Ramsay J and Silverman W B 2005 *Functional Data Analysis* (Springer Series in Statistics) (New York, NY: Springer)
- Repetti A, Pereyra M and Wiaux Y 2019 Scalable Bayesian uncertainty quantification in imaging inverse problems via convex optimization *SIAM J. Imaging Sci.* **12** 87–118
- Riesz F and Szokefalvi-Nagy B 1955 *Functional Analysis* (New York, NY: F. Ungar Pub. Co.)
- Silverman B W 1996 Smoothed functional principal components analysis by choice of norm *Ann. Stat.* **24** 1–24
- Singh H *et al* 2014 Mapping cortical haemodynamics during neonatal seizures using diffuse optical tomography: a case study *NeuroImage: Clinical* **5** 256–65
- Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
- Tenorio L 2001 Statistical regularization of inverse problems *SIAM Rev.* **43** 347–66
- Tian T S, Huang J Z, Shen H and Li Z 2012 A two-way regularization method for MEG source reconstruction *Ann. Appl. Stat.* **6** 1021–46
- Van Essen D *et al* 2012 The human connectome project: a data acquisition perspective *NeuroImage* **62** 2222–31
- Vogel C R 2002 *Computational Methods for Inverse Problems* (Philadelphia, PA: Society for Industrial and Applied Mathematics)
- Yao F, Müller H-G and Wang J-L 2005 Functional data analysis for sparse longitudinal data *J. Am. Stat. Assoc.* **100** 577–90
- Ye J C, Tak S, Jang K E, Jung J and Jang J 2009 NIRS-SPM statistical parametric mapping for near-infrared spectroscopy *NeuroImage* **44** 428–47
- Yeo B T T *et al* 2011 The organization of the human cerebral cortex estimated by intrinsic functional connectivity *J. Neurophysiol.* **106** 1125–65
- Zhdanov M 2002 *Inverse Theory and Applications in Geophysics* (Amsterdam: Elsevier)
- Zhu H, Leus G and Giannakis G B 2011 Sparsity-cognizant total least-squares for perturbed compressive sampling *IEEE Trans. Signal Process.* **59** 2002–16
- Zienkiewicz O, Taylor R and Zhu J Z 2013 *The Finite Element Method: Its Basis and Fundamentals* 7th edn (Amsterdam: Elsevier)