TECHNICAL UNIVERSITY OF CRETE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

Machine Learning Techniques with Applications in Biomedical Signal Processing

by

Paris Karakasis

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE MASTERS OF SCIENCE OF

ELECTRICAL AND COMPUTER ENGINEERING

December 2019

THESIS COMMITTEE

Professor Athanasios P. Liavas, *Thesis Supervisor*
Professor Michael Zervakis
Professor George Karystinos

Abstract

Functional magnetic resonance imaging (fMRI) is one of the most popular methods for studying the human brain. The purpose of task-related fMRI data analysis is to determine which brain areas are activated when a specific task is performed, based on the BOLD signal analysis. Background BOLD signal reflects systematic fluctuations in regional brain activity that are attributed to the existence of resting-state brain networks. However, this structure usually is not taken into account in task-related fMRI data analysis.

A wide range of unsupervised multivariate statistical methods is being increasingly employed in fMRI data analysis. The main goal of these methods is to extract information from a dataset, often with no prior knowledge of the experimental conditions. Generalized canonical correlation analysis (gCCA) is a well known statistical method, that can be considered as the problem of estimating a linear subspace, which is "common" to multiple sets of random variables. We propose a new data generating model which takes into consideration the existence of common task-related and rest-related components. Moreover, we attempt to recover all task-related components via means of gCCA. We extensively test our theoretical results using both artificial and real-world fMRI data. We observe that our experimental findings corroborate our theoretical results, rendering our approach a very good candidate for multi-subject task-related fMRI processing.

Acknowledgements

A period of nine and a half years reaches its end. In this journey, I had the opportunity to meet many people, to make friends, and to get to know myself better. While I didn't know what to expect from this journey at its beginning, I had the good fortune to discover many interests and a way to become creative. It was the environment and the people of the School of Electrical and Computer Engineering at Technical University of Crete that made this possible and I truly feel blessed that I had the chance to be a student here. Many people here influenced me and provided me with valuable stimuli and knowledge, and I am grateful to all of them.

The educational process needs a student and a teacher in order to be defined. In my opinion, the most successful way, for a teacher, to motivate his/her students is to convince them that there is something beautiful and meaningful in what he/she perceives. For me, this person was Prof. Athanasios Liavas, who not only convinced me, but also guided me with enthusiasm, patience, and care, in the landscapes of research, mathematics, and engineering. I am greatly indebted to him and I will be always thankful for inspiring me to pursue an academic career.

In case where someone would look for the starting point of this work, then he/she would locate it in a conversation with Prof. Nikolaos Sidiropoulos about "common" subspaces. This conversation unstuck my attempts in finding commonalities across different samples and gave me the opportunity for collaboration with Prof. Athanasios Liavas, Prof. Nikolaos Sidiropoulos, and Prof. Panagiotis Simos on the problem of multi-subject fMRI data analysis. Crucial was also the help of Dr. Eleutherios Kavroulakis with the preprocessing of the fMRI data, as well as the useful critiques and comments from Prof. Dionisios Christopoulos and thesis committee members, Prof. Michalis Zervakis and Prof. Georgios Karystinos. Without the contribution of each one of them, this thesis would not be the same. I am thankful to all of them.

Each journey has its happy and sad moments. I had the pleasure to share many of these moments with family and friends, which celebrated with me the good moments and supported me through the bad ones. I am grateful to them for being there for me.

At last, because I would not like to forget anyone, I am grateful to everyone that helped and supported me.

Table of Contents

List of Figures

List of Abbreviations

BSS	Blind Source Separation
CCA	Canonical Correlation Analysis
gCCA	Generalized Canonical Correlation Analysis
PCA	Principal Components Analysis
ICA	Independent Component Analysis
LS	Least Squares
NNLS	Nonnegative Least Squares
PARAFAC or CP	Parallel Factor Analysis
fMRI	Functional Magnetic Resonance Imaging
BOLD	Blood Oxygenation Level Dependent
HRF	Hemodynamic Response Function
TR	Repetition Time

Chapter 1

Introduction

Functional magnetic resonance imaging (fMRI) is one of the most popular methods for studying the human brain. fMRI provides a non-invasive way to measure brain activity, by detecting local changes of blood oxygen level density (BOLD) in the brain, over time. More specifically, in fMRI data, a brain is represented by a finite set of volume elements (voxels). For each voxel, we have a time series that indicates the concentration of oxygen in this area over time. These time series are known as blood oxygenation level dependent (BOLD) signals. The purpose of task-based fMRI data analysis is to determine which brain areas are activated when a specific task is performed, based on the BOLD signal analysis. Hence, brain activation maps related to specific tasks can be obtained. This procedure is very useful for understanding how the human brain is functioning.

Spontaneous modulation of the BOLD signal, which cannot be attributed to the experimental paradigm or any other explicit input or output, is also present and is usually viewed as "noise" in task-related studies [3, 4]. However, in addition to physiological and magnetic noise, background BOLD signal reflects systematic fluctuations in regional brain activity. In particular, BOLD fluctuations are correlated between functionally related brain regions, forming resting-state brain networks. Moreover, this baseline activity continues during task performance, showing a similar neuro-anatomical distribution to that observed at rest [5, 6, 7, 8], while there are studies suggesting that measured neuronal responses represent an approximately linear superposition of task-evoked neuronal activity and ongoing spontaneous activity [5].
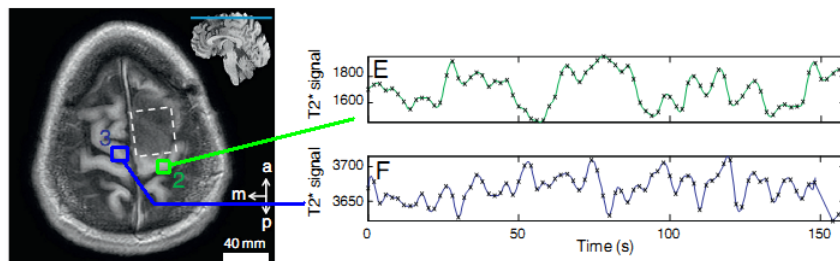
Figure 1.1: An example of fMRI data [1].

In order to isolate the signals that we are interested in, we can consider the problem of fMRI data analysis as a blind source separation problem (BSS). Blind source separation (BSS) refers to the problem of extracting a set of source signals from a set of mixed signals, without using prior information (or with very little prior information) about the source signals or the mixing process. The classical example of a source separation problem is the cocktail party problem, where a number of people are talking simultaneously in a room (for example, at a cocktail party), and a listener is trying to follow one of the discussions. The human brain can handle this sort of auditory source separation problem, but it is a difficult problem in digital signal processing. Hence, BSS aims in enhancing noisy speech in real world environments and the applications are not just limited to speech/audio processing but also include topics in astronomical, satellite, econometric, and biomedical signal processing.
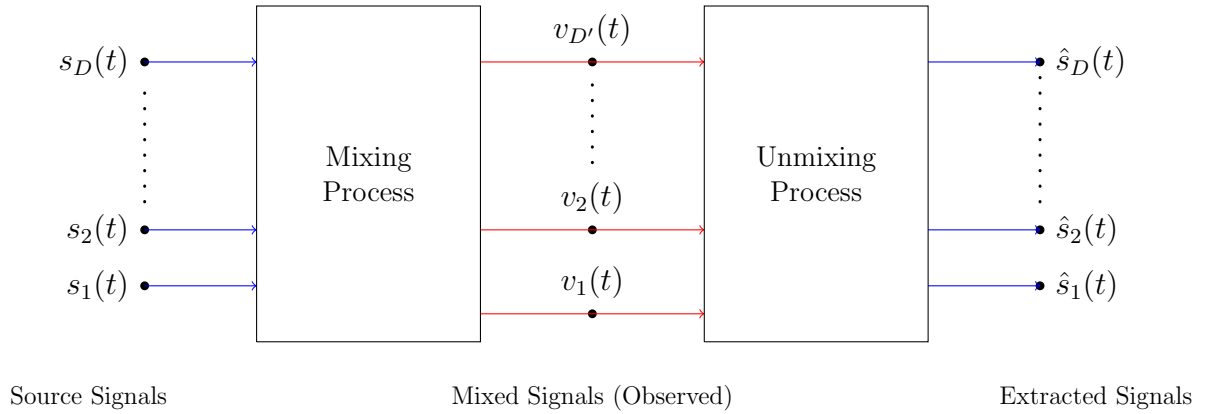
Figure 1.2: Abstract view of the BSS problem.

BSS problems are, in general, highly under-determined, since lack of prior knowledge (number of sources, characteristics of the source signals and the mixing procedure) may lead to a set of multiple solutions for the same problem. A variety of methods in addressing the BSS problem have been proposed in the literature. The most popular of them, among others, are principal component analysis (PCA), independent compontent analysis (ICA), nonnegative matrix factorization (NMF), as well as tensor factorizations models.

As a result, a wide range of unsupervised multivariate statistical methods is being increasingly employed in fMRI data analysis. The main goal of these data driven methods is to extract information from a dataset, often with no prior knowledge of the experimental conditions. In the context of fMRI, multivariate methods provide statistical inference on a whole-brain basis so as to describe brain responses in terms of spatial and temporal patterns. The most common multivariate methods are Principal Component Analysis

(PCA) [9, 10], Independent Component Analysis (ICA) [11, 12, 13, 14], and analysis via tensor factorization models [15, 16, 17, 18, 19, 20, 21, 22].

Canonical correlation analysis (CCA) is a well known statistical method, developed by Hotelling in 1936 [23]. Specifically, CCA can be seen as the problem of finding basis vectors for two sets of random variables (i.e., two random vectors) such that the correlation between the respective projections of the variables onto these basis vectors is mutually maximized [24]. After considering the subspace that is spanned from such a set of basis vectors, CCA can be also considered as the problem of estimating a linear subspace, which is "common" to these sets of variables [25].

Generalization of CCA to more than two random vectors dates back to [26, 27, 28, 29, 30]. Kettenring proposed five different formulations of the generalized CCA (gCCA) problem in [31]; all of them reduce to the classical CCA when the number of random vectors is two[32]. Among all different formulations of gCCA, the MAX-VAR formulation has attracted particular attention, since it enjoys a simple solution via eigen-decomposition, while scalable and structure promoting iterative algorithms for two of the gCCA formulations, MAX-VAR and SUM-COR, have been proposed in [33, 34, 35].

1.1 Problem Definition

We focus on the case where the task-related fMRI experiment of a session consists of only one type of stimulus. Our aim is to determine which brain areas are activated when the stimulus is applied, and construct the associated brain activation maps.

1.2 Related Work

CCA and constrained CCA based methods have been applied to local voxel neighborhoods to obtain adaptive subject-specific spatial filter kernels for noise reduction purposes [36, 37, 38, 39]. In [40], the authors used gCCA to separate different temporal sources in fMRI data. They assumed that there are some common temporal responses to external stimulation in the subjects being studied, and they showed that these may be explored using gCCA. In contrast, in [41], the underlying assumption is that there are multiple subjects that share an unknown spatial response (or spatial map) to the common experimental manipulation but may show different temporal responses to external stimulation. In addition, estimating "common" subspaces from multiple datasets via CCA and gCCA based methods have been considered in [42, 25].

1.3   Our Contribution

We adopt the assumptions of [41], with respect to the common spatial maps, and assume the existence of one common temporal component, which is related to the common experimental excitation. We propose a new data generating model which takes into consideration both the common task-related spatial component and the common rest-related spatial components. We use gCCA and estimate the subspace that is spanned by the common spatial components, both task- and rest-related. Based on this estimate, we compute the common task-related time component, using, again, gCCA. Finally, we use the estimated common task-related time component to derive an estimate of the associated common task-related spatial component and construct the respective activation map.

We extensively test our theoretical results using both artificial and real-world fMRI data. We observe that our experimental findings corroborate our theoretical results, rendering our approach a very good candidate for multi-subject task-related fMRI processing.

1.4   Notation

Scalars are denoted by small letters, vectors by small bold letters, and matrices by capital bold letters, for example, x, $\mathbf{x}$, $\mathbf{X}$. Sets are denoted by blackboard bold capital letters, for example, $\mathbb{U}$. $\mathbb{R}$ denotes the sets of real numbers. $\mathbb{R}^{I \times J}$ denotes the set of $(I \times J)$ real matrices. Inequality $\mathbf{A} \geq \mathbf{0}$ means that matrix $\mathbf{A}$ has nonnegative elements and $\mathbb{R}_+^{I \times J}$ denotes the set of $(I \times J)$ real matrices with nonnegative elements. $\|\cdot\|_2$ and $\|\cdot\|_F$ denote, respectively, the spectral and the Frobenius norm of the matrix argument. The transpose and the pseudoinverse of matrix $\mathbf{A}$ are denoted, respectively, by $\mathbf{A}^T$ and $\mathbf{A}^\dagger$. The linear space spanned by the columns of matrix $\mathbf{A}$ is denoted by $\text{col}(\mathbf{A})$. The orthogonal projection onto a subspace $\mathcal{S}$ is denoted by $\mathbf{P}_{\mathcal{S}}$. Finally, we introduce the Matlab style notations $\mathbf{A}(:, l)$ and $\mathbf{A}(k, :)$, which denote, respectively, the l-th column and the k-th row of matrix $\mathbf{A}$.

1.5   Structure

In Chapter 2, we introduce the reader to the problem of analysing fMRI data. In Chapter 3, we present the data generating model and our approach for estimating the common task-related temporal and spatial components, via two successive gCCA problems. We also discuss the estimation of the dimension of the common spatial subspace. In Chapter 4, we apply our approach to both artificial and real-world fMRI data. Finally, in Chapter 5, we conclude our work and we propose directions for future work.

Chapter 2

Functional Magnetic Resonance Imaging and Results

2.1 Introduction

Functional magnetic resonance imaging (fMRI) has, in less than two decades, become the most commonly used method for the study of human brain function. fMRI is a technique that uses magnetic resonance imaging to measure brain activity by measuring changes in the local oxygenation of blood, which in turn reflects the amount of local brain activity. The analysis of fMRI data is exceedingly complex, requiring the use of sophisticated techniques from signal and image processing and statistics in order to go from the raw data to the finished product, which is generally a statistical map showing which brain regions responded to a particular manipulation of mental or perceptual functions.

The most common method of fMRI takes advantage of the fact that, when neurons in the brain become active, the amount of blood flowing through that area is increased. This phenomenon has been known for more than 100 years, though the mechanisms that cause it remain only partly understood. What is particularly interesting is that the amount of blood that is sent to the area is more than is needed to replenish the oxygen that is used by the activity of the cells [43]. Thus, the activity related increase in blood flow caused by neuronal activity leads to a relative surplus in local blood oxygen. The signal measured in fMRI depends on this change in oxygenation and is referred to as the blood oxygenation level dependent (BOLD) signal.

For our purposes, it suffices to know that the BOLD signal is a measure of the ratio of oxygenated to deoxygenated hemoglobin. Hemoglobin is a molecule in the blood that carries oxygen from the lungs to all parts of the body. It has sites to bind up to four oxygen molecules. A key discovery that led eventually to BOLD fMRI was that hemoglobin molecules fully loaded with oxygen have different magnetic properties than hemoglobin molecules with empty binding sites [44].

The theory, which is not yet fully worked out, is that active brain areas consume more oxygen than inactive areas. When neural activity increases in an area, metabolic demands
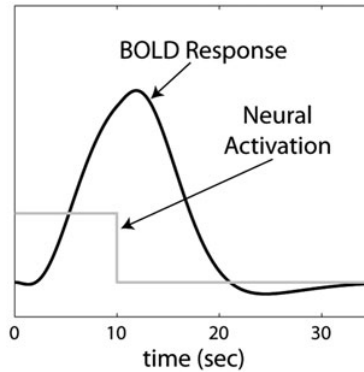
Figure 2.1: A hypothetical BOLD response (black curve) to a constant 10 sec neural activation (gray curve) [2].

rise and, as a result, the vascular system concentrates oxygenated hemoglobin into the area. An idealized example of this process is shown in Figure 2.1. The concentration of oxygenated hemoglobin into the area causes the ratio of oxygenated to deoxygenated hemoglobin (i.e., the BOLD signal) to rise quickly. As this happens, the vascular system over compensates, in the sense that the BOLD signal actually rises well above baseline to a peak at around 6 sec after the end of the neural activity that elicited these responses. Following this peak, the BOLD signal gradually decays back to baseline over a period of 20–25 sec.

2.2   Modeling the BOLD Response

The goal of almost all fMRI experiments is to obtain information about neural activity. However, the BOLD response measured in most fMRI experiments provides only an indirect measure of neural activation [45, 46]. Although it is commonly assumed that the BOLD signal increases with neural activation, it is known that the BOLD response is much more sluggish than the neural activation that is presumed to drive it. As a result, the peak of the BOLD signal lags considerably behind the peak neural activation (e.g., see Figure 2.1).

Almost all current applications of fMRI assume that the transformation from neural activation to BOLD response can be modeled as a linear, time-invariant system. Although it is becoming increasingly clear that the transformation is, in fact, nonlinear (e.g., [47, 48, 49]). It also appears that these departures from linearity are not severe so long as events are well separated in time (e.g., at least a few seconds apart) and brief exposure durations are avoided [49].

In the linear systems approach, one can conceive of the vascular system that responds

to a sudden oxygen debt as a black box. The input is neural activation and the output is the BOLD response. Let $N_i(t)$ denote the neural activation induced by this event at time t and let $B_i(t)$ denote the corresponding BOLD response. Then, from the systems theory perspective, the box represents the set of all mathematical transformations that convert the neural activation $N_i(t)$ into the BOLD response $B_i(t)$. For convenience, we will express this mathematical relationship as

$$B_i(t) = f(N_i(t)), \tag{2.1}$$

where the operator f symbolizes the workings of the black box.

A system of this type is said to be linear and time-invariant if and only if it satisfies the superposition principle, which is stated as follows:

$$
\begin{aligned}
&\text{If} \quad f(N_1(t)) = B_1(t) \quad \text{and} \quad f(N_2(t)) = B_2(t), \\
&\text{then} \quad f(\alpha_1 N_1(t) + \alpha_2 N_2(t)) = \alpha_1 B_1(t) + \alpha_2 B_2(t),
\end{aligned} \tag{2.2}
$$

for any constants α_1, α_2.

In other words, if we know what the BOLD response is to neural activation $N_1(t)$ and to neural activation $N_2(t)$, then we can determine exactly what the BOLD response will be to any weighted sum of these two neural activations by computing the same weighted sum of the component BOLD responses.

If the superposition principle holds, then there is a straightforward way to determine the BOLD response to any neural activation from the results of one simple experiment. All we need to do is to measure the BOLD response that occurs when the neural activation is an impulse. Denote the BOLD response in this idealized experiment by $h(t)$. In linear systems theory, the function $h(t)$ is called the impulse response function and is the response of the system to an impulse. In the fMRI literature, $h(t)$ is known as the hemodynamic response function (HRF).

If the relationship between neural activation and the BOLD response satisfies superposition, then, once we know the HRF, the BOLD response to any neural activation $N(t)$ can be computed exactly from the convolution integral:

$$B(t) = \int_0^t N(\tau) h(t - \tau) \, d\tau. \tag{2.3}$$

The convolution integral massively simplifies the analysis of fMRI data and, as a result, it forms the basis for the most popular methods of fMRI data analysis.
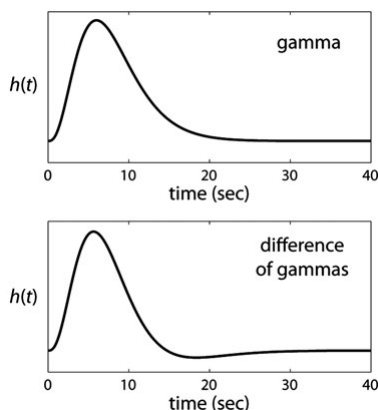
Figure 2.2: Two popular models of the HRF [2].

Given that the HRF plays such a critical role in analyzing fMRI data, the natural next question to ask is: how can we determine numerical values of the HRF? A popular method is to select a specific mathematical function for the HRF based on our knowledge of what we think this function should look like. For example, we know the HRF should peak at roughly 6 s and then slowly decay back to baseline. So, we could select a mathematical function with these properties and then just assume that this is a good model of the HRF. In fact, this is, by far, the most popular method for determining the HRF in fMRI data analysis. The most popular choices are a gamma function or the difference of two gamma functions. Examples of both of these models are shown in Figure 2.2.

2.3   Data Analysis

The size of acquired fMRI data greatly complicates its analysis. First, as mentioned above, a typical scanning session generates a huge amount of data. Second, fMRI data is characterized by substantial spatial and temporal correlations. For example, the sluggish nature of the BOLD response means that, if the BOLD response in some voxel is greater than the average on one TR, then it is also likely to be greater than the average on the ensuing TR. Similarly, because brain tissue in neighboring voxels will be supplied by a similar vasculature, a large response in one voxel increases the likelihood that a large response will also be observed at neighboring voxels. A third significant challenge to fMRI data analysis is the noisy nature of fMRI data. Typically the signal that the data analysis techniques are trying to find is less than 2 or 3 % of the total BOLD response.

The analysis of fMRI BOLD data is broken down into two general stages - preprocessing and task-related analysis. Preprocessing includes a number of steps that are required

to prepare the data for task-related analysis. This includes, for example, aligning the functional and structural scans, correcting for any possible head movements that might have occurred during the functional run, and various types of smoothing (to reduce noise). Typically, the same preprocessing steps are always completed, regardless of the particular research questions that the study was designed to address. In contrast, the task-related analyses include all analyses that are directed at these questions.

2.3.1 Task-Related Data Analysis

At the preprocessing stages, the quality of the fMRI data is improved. After that, statistical analysis is attempted to determine which voxels are activated by the stimulation. Most fMRI studies are established upon the correlation of hemodynamic response function with stimulation. Activation defines the local intensity changes in the voxels. These methods can be grouped into two broad categories: the univariate methods (hypothesis testing methods), and the multivariate methods (exploratory methods) [50].

The univariate methods attempt to define which voxels can be characterized as activated given one signal model. This allows the parametrization of the response and then the estimation of the model parameters. The univariate methods are widely used to analyze brain data obtained from fMRI. In these methods, signal estimation and the presence or the absence of activation are defined by the statistical test. One of the most popular methods is the generalized linear model (GLM), which is based upon the hypothesis of linear correlation between neuro-activities and the tasks [51].

Multivariate methods are also applied to fMRI data analysis, which extract information from dataset, often with any prior knowledge of the experimental conditions. They use some structural properties, such as decorrelation, independence, similarity measures, that can discriminate between features of interest present in the data. Unlike the univariate methods which carry out voxel-wise statistical analysis, multivariate methods provide statistical inference about the whole brain so as to describe brain responses in terms of spatial patterns [12]. A wide range of multivariate statistical methods is being increasingly employed to analyse the fMRI time series. fMRI data are essentially multivariate in nature, since information about thousands of measured locations (voxels) are being impacted in each scan [52]. Those methods aim at summarizing the spatial and temporal structures of the data. As the distribution of brain regions are involved in a task, it seems to be desirable to use the spatially distributed information from different areas to understand a brain function. So the multivariate approaches seem to be interesting in this case to consider the spatially distributed information. The most common multivariate methods, among others,

are Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Multi-Voxel Pattern Analysis (MVPA).

Tensor factorization models have been also proposed in processing fMRI data. In contrast to PCA and ICA, tensor factorization models are unique under mild conditions, without enforcing any constraints to source signals. In Andersen and Rayens (2004) it was demonstrated how the PARAFAC (CP) model is useful in the analysis of neuroimaging data such as fMRI [15]. However, degeneration of the unconstrained PARAFAC model is a frequent problem in the analysis of fMRI data. Additional applications of multilinear (multiway) modeling in fMRI include the PICA model [16], where the ICA model is extended for tensors. Also, the Shifted PARAFAC model (SCP) has been proposed in processing fMRI data, since time shifts occur naturally in fMRI data. For instance, these could be due to hemodynamic delay [53] or they could arise in stimuli studies [54], where delays play a particularly important role. Finally, the PARAFAC model with orthogonality constraints over the spatial mode has been proposed in processing fMRI data in [55], in order to reduce cross-talk between spatial components and eliminate degeneration phenomena.

CCA and constrained CCA based methods has been applied to local voxel neighborhoods to obtain adaptive subject-specific spatial filter kernels for noise reduction purposes [36, 37, 38, 39]. In [40], the authors used gCCA to separate different temporal sources in fMRI data. They assumed that there are some common temporal responses to external stimulation in the subjects being studied, and they showed that these may be explored using gCCA. In contrast, in [41], the underlying assumption is that there are multiple subjects that share an unknown spatial response (or spatial map) to the common experimental manipulation but may show different temporal responses to external stimulation. In addition, estimating "common" subspaces from multiple datasets via CCA and gCCA based methods have been considered in [42, 25].

Chapter 3

Multi-subject fMRI data analysis

3.1 Data Model

Let $\{\mathbf{X}_k\}_{k=1}^{K}$ be a set of matrices, where $\mathbf{X}_k \in \mathbb{R}^{N \times M}$ denotes the data of the k-th subject, N denotes the number of voxels, and M denotes the number of time points (note that, in general, $N \gg M$). Let R be a positive integer smaller than M. For each matrix $\mathbf{X}_k$, for $k = 1, \ldots, K$, we adopt the model

$$\mathbf{X}_k = \lambda_k \mathbf{a}\mathbf{s}^T + \mathbf{A}\mathbf{S}_k^T + \mathbf{E}_k, \tag{3.1}$$

where:

1. $\mathbf{a} \in \mathbb{R}_+^N$ and $\mathbf{s} \in \mathbb{R}^M$ denote, respectively, the common, to all subjects, task-related spatial and temporal component, and $\lambda_k \in \mathbb{R}_+$ denotes the intensity of the common rank-one term for the k-th subject;

2. $\mathbf{A} \in \mathbb{R}_+^{N \times (R-1)}$, whose columns are the common, to all subjects, spatial components related with the spontaneous fMRI activity;

3. $\mathbf{S}_k \in \mathbb{R}^{M \times (R-1)}$, whose columns are the temporal components, which are associated with the spontaneous fMRI activity and, in general, vary across subjects;

4. $\mathbf{E}_k \in \mathbb{R}^{N \times M}$ denotes the "unmodeled fMRI signal" of the k-th subject and can be considered as (strong) additive noise. We assume that terms $\mathbf{E}_k$ are statistically independent from each other.

We propose model (3.1) based on both the existing literature [3, 4, 5, 6, 7, 8, 14, 41, 12] and the detailed examination of our real-world data. Our aim is to obtain an accurate estimate of the common spatial term $\mathbf{a}$, which will lead to a precise activation brain map and by extension to the localization of the stimulated brain areas.

In order to use simpler notation, we define the matrix of the common spatial components

$$\mathbf{W} := [\mathbf{a} \ \mathbf{A}] \in \mathbb{R}_+^{N \times R}, \tag{3.2}$$

and the matrices of the temporal components

$$\mathbf{Z}_k := [\lambda_k \mathbf{s} \ \mathbf{S}_k] \in \mathbb{R}^{M \times R}, \quad \text{for } k = 1, \dots, K. \tag{3.3}$$

We further assume that matrices $\mathbf{W}$ and $\mathbf{Z}_k$, for $k = 1, \dots, K$, are full-column rank.

Using this notation, matrix $\mathbf{X}_k$, defined in (3.1), can be expressed as

$$\mathbf{X}_k = \mathbf{W}\mathbf{Z}_k^T + \mathbf{E}_k. \tag{3.4}$$

One way to achieve our aim, that is, to estimate $\mathbf{a}$, is to concatenate matrices $\mathbf{X}_k$, for $k = 1, \dots, K$, and construct matrix $\mathbf{X}$ as

$$\mathbf{X} := [\mathbf{X}_1 \ \dots \ \mathbf{X}_K] = \mathbf{W}\mathbf{Z}^T + \mathbf{E}, \tag{3.5}$$

where matrices $\mathbf{Z}$ and $\mathbf{E}$ are defined similarly to $\mathbf{X}$, and use matrix factorization techniques to obtain estimates of $\mathbf{W}$ and $\mathbf{Z}$. However, without imposing stringent constraints on $\mathbf{W}$ and $\mathbf{Z}$, their estimation is a difficult task. This happens because, if $(\hat{\mathbf{W}}, \hat{\mathbf{Z}})$ is a pair that achieves the best rank-R approximation of matrix $\mathbf{X}$, then, for any invertible matrix $\mathbf{H} \in \mathbb{R}^{R \times R}$, the pair $(\tilde{\mathbf{W}} = \hat{\mathbf{W}}\mathbf{H}, \tilde{\mathbf{Z}} = \hat{\mathbf{Z}}\mathbf{H}^{-T})$ attains the same approximation error. Without imposing stringent constraints on $\mathbf{W}$ and $\mathbf{Z}$, we can only obtain estimates of the subspaces $\text{col}(\mathbf{W})$ and $\text{col}(\mathbf{Z})$.

3.2   On the Estimation of Common Subspaces and Common Components

In this section, we describe our approach for the estimation of the common spatial factor $\mathbf{a}$, which consists of three stages:

1. we use $\mathbf{X}_k$, for $k = 1, \dots, K$, and obtain an orthonormal basis for an estimate of the common spatial subspace, $\text{col}(\mathbf{W})$, by solving a gCCA problem;

2. using the solution of the first stage, we obtain an estimate of the unique common time component $\mathbf{s}$, by solving a second gCCA problem;

3. using the estimate of $\mathbf{s}$, we obtain an estimate of $\mathbf{a}$ (we propose two different estimates).

3.2.1   Common Spatial Subspace Estimation via gCCA

We assume that the dimension, R, of the common spatial subspace, $\mathrm{col}(\mathbf{W})$, is known; we shall say more on this important topic later.

In order to estimate an orthonormal basis for the common spatial subspace, $\mathrm{col}(\mathbf{W})$, we solve the following optimization problem, which arises from the MAXVAR formulation of the gCCA [28]

$$\min_{\{\mathbf{Q}_k\}_{k=1}^K, \mathbf{G}} \sum_{k=1}^K \|\mathbf{X}_k \mathbf{Q}_k - \mathbf{G}\|_F^2 \tag{3.6}$$
$$\text{s.t.} \quad \mathbf{G}^T \mathbf{G} = \mathbf{I}_R,$$

where $\mathbf{Q}_k \in \mathbb{R}^{M \times R}$, for $k = 1, \dots, K$, and $\mathbf{G} \in \mathbb{R}^{N \times R}$.

The solution $\mathbf{Q}_k^o$, for $k = 1, \dots, K$, and $\mathbf{G}^o$ of problem (3.6) can be computed as follows. For a fixed $\mathbf{G}$, the optimal $\mathbf{Q}_k$ can be expressed as $\mathbf{Q}_k(\mathbf{G}) = \mathbf{X}_k^\dagger \mathbf{G}$, for $k = 1, \dots, K$. If we substitute this expression into problem (3.6), then the problem becomes

$$\max_{\mathbf{G}^T \mathbf{G} = \mathbf{I}_R} \mathrm{Tr} \left(\mathbf{G}^T \left(\sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^\dagger \right) \mathbf{G} \right). \tag{3.7}$$

If we define

$$\mathbf{M} := \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^\dagger, \tag{3.8}$$

with eigenvalue decomposition given by

$$\mathbf{M} = \mathbf{U}_M \mathbf{\Lambda}_M \mathbf{U}_M^T, \tag{3.9}$$

then an optimal solution $\mathbf{G}^o$ is given by [34]

$$\mathbf{G}^o = \mathbf{U}_M (:, 1:R). \tag{3.10}$$

If the fMRI data matrices $\mathbf{X}_k$ were noiseless, in the sense that $\mathbf{E}_k = \mathbf{0}$, for $k = 1, \dots, K$, then the solution of problem (3.6) would result to $\mathbf{G}^o$, such that (see (3.4))

$$\mathrm{col}(\mathbf{G}^o) = \mathrm{col}(\mathbf{W}). \tag{3.11}$$

Furthermore, in this case and for all $k \in \{1, \dots, K\}$, matrices $\mathbf{Q}_k^o$ and $\mathbf{Z}_k$ would span the

same subspace, namely

$$\mathrm{col}(\mathbf{Q}_k^o) = \mathrm{col}\left(\mathbf{Z}_k\right). \tag{3.12}$$

This holds because, if $\mathbf{W} = \mathbf{G}^o\mathbf{P}$, then $\mathbf{Q}_k^o = \mathbf{X}_k^{\dagger}\mathbf{G}^o = \left(\mathbf{Z}_k^T\right)^{\dagger}\mathbf{W}^{\dagger}\mathbf{G}^o = \mathbf{Z}_k\mathbf{F}$, where $\mathbf{F} := \left(\mathbf{Z}_k^T\mathbf{Z}_k\right)^{-1}\mathbf{P}^{-1}$. The fact that $\mathbf{E}_k$, for $k = 1, \ldots, K$, are nonzero makes (3.11) and (3.12) approximate and not exact equalities.

In the sequel, we shall compute an estimate of $\mathbf{s}$ by assuming that (3.12) is exact. We shall check the accuracy of our arguments and the effectiveness of our approach in the section with the experimental results.

3.2.2   Estimation of the Common Time Component

We have no reason to believe that, in general, the spontaneous fMRI components $\mathbf{S}_k$, for $k = 1, \ldots, K$, generate the same or even close subspaces. Thus, we assume that (see (3.3))

$$\bigcap_{k=1}^{K} \mathrm{col}\left(\mathbf{Z}_k\right) = \mathrm{col}\left(\mathbf{s}\right), \tag{3.13}$$

which, using (3.12), leads to

$$\bigcap_{k=1}^{K} \mathrm{col}\left(\mathbf{Q}_k^o\right) = \mathrm{col}\left(\mathbf{s}\right). \tag{3.14}$$

We obtain an estimate of $\mathbf{s}$ by solving the following MAXVAR problem

$$\min_{\{\mathbf{d}_k\}_{k=1}^K, \mathbf{g}} \sum_{k=1}^{K} \left\|\mathbf{Q}_k^o\mathbf{d}_k - \mathbf{g}\right\|_2^2 \tag{3.15}$$

$$\text{s.t. } \|\mathbf{g}\| = 1.$$

If we denote the optimal $\mathbf{g}$ in (3.15) by $\mathbf{g}^o$, we have that

$$\mathbf{g}^o = \pm\frac{\mathbf{s}}{\|\mathbf{s}\|_2}. \tag{3.16}$$

Since (3.12) defines a family of approximate equalities, equality (3.14) and, therefore, (3.16) are approximate.

3.2.3 Estimation of the common spatial component a

Having obtained the estimate $\mathbf{g}^o$ of the common temporal component, $\mathbf{s}$, we can estimate the common spatial component, $\mathbf{a}$, by using various approaches. First, we consider the constrained optimization problem

$$\min_{\boldsymbol{\lambda} \geq 0} \sum_{k=1}^{K} \|\mathbf{X}_k - \lambda_k \mathbf{a} \mathbf{g}^{o^T}\|_F^2 \tag{3.17}$$

$$\text{s.t. } \mathbf{a} \in \text{col}\left(\mathbf{G}^o\right).$$

Let $\tilde{\mathbf{X}}_k = \mathbf{P}_{\text{col}(\mathbf{G}^o)} \mathbf{X}_k$, for $k = 1, \ldots, K$, denote the data matrices after projecting them onto the subspace spanned by the columns of matrix $\mathbf{G}^o$. Then, one can easily show that the optimization problem (3.17) is equivalent to the problem

$$\min_{\boldsymbol{\lambda} \geq 0} \sum_{k=1}^{K} \|\tilde{\mathbf{X}}_k - \lambda_k \mathbf{a} \mathbf{g}^{o^T}\|_F^2. \tag{3.18}$$

Next, we consider problem (3.18) with additional nonnegativity constraints on $\mathbf{a}$, namely

$$\min_{\mathbf{a} \geq \mathbf{0}, \boldsymbol{\lambda} \geq \mathbf{0}} \sum_{k=1}^{K} \|\tilde{\mathbf{X}}_k - \lambda_k \mathbf{a} \mathbf{g}^{o^T}\|_F^2. \tag{3.19}$$

The solution $\mathbf{a}^o$ of this problem is an estimate of the common spatial component $\mathbf{a}$.

A second approach emerges by assuming that $\lambda_k = 1$, for $k = 1, \ldots, K$. Then, we can estimate $\mathbf{a}$ by computing the average inner product between $\mathbf{g}^o$ and each voxel's time-series, namely

$$\hat{\mathbf{a}} = \left(\frac{1}{K} \sum_{k=1}^{K} \tilde{\mathbf{X}}_k \mathbf{g}^o \right)_+ . \tag{3.20}$$

3.3 On the Dimension of Common Spatial Subspace

In Subsection 3.2.1, we assumed that we know the true dimension, R, of the common spatial subspace, $\text{col}(\mathbf{W})$, and derived the optimal orthonormal basis $\mathbf{G}^o$ of $\text{col}(\mathbf{W})$. Of course, in general, the value of R is not known, thus, we must estimate it. As we shall see in the section with the experimental results, the estimate of the common time component, $\mathbf{g}^o$, is quite "stable" if the value of the assumed rank belongs to a certain set of values. Thus, since the focus, in this work, is on the estimation of $\mathbf{a}$, which is achieved via $\mathbf{g}^o$,

we may argue that an accurate common spatial subspace dimension estimation is not of foremost importance. However, in the sequel, we provide a procedure which gives us very useful information about the value of R.

Let the assumed dimension of the common spatial subspace be $\hat{R} = R$. Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be a partition of the set of the subjects $\{1, \ldots, K\}$. In the noiseless case, if we solve the problem corresponding to (3.6) twice, for $k \in \mathcal{K}_1$ and for $k \in \mathcal{K}_2$, and call the resulting orthonormal bases $\mathbf{G}_1^o$ and $\mathbf{G}_2^o$, respectively, then we must have

$$\operatorname{col}(\mathbf{G}_1^o) = \operatorname{col}(\mathbf{G}_2^o). \tag{3.21}$$

That is, the common spatial subspaces, in the two cases, coincide. If we start adding noise and repeat the process, then we will obtain $\operatorname{col}(\mathbf{G}_1^o)$ and $\operatorname{col}(\mathbf{G}_2^o)$ which will be "close" to each other. One way to measure the distance between subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ is to compute their gap, defined as [56]

$$\rho_{g,2}\left(\mathcal{S}_1, \mathcal{S}_2\right) := \|\mathbf{P}_{\mathcal{S}_1} - \mathbf{P}_{\mathcal{S}_2}\|_2. \tag{3.22}$$

Thus, if $\hat{R} = R$ and $\|\mathbf{E}_k\|_2 = O(\epsilon)$, for $k = 1, \ldots, K$, where ϵ is a small positive number, then we expect that

$$\|\mathbf{P}_{\operatorname{col}(\mathbf{G}_1^o)} - \mathbf{P}_{\operatorname{col}(\mathbf{G}_2^o)}\|_2 = O(\epsilon). \tag{3.23}$$

If $\hat{R} > R$, then, by solving (3.6), besides the R-dimensional common subspace, $\operatorname{col}(\mathbf{W})$, we shall try to model "common" noise subspace. Since the noise terms $\mathbf{E}_k$ are independent across subjects and $N \gg M$, we do not expect to find any common noise subspace in the two subsets, $\mathcal{K}_1$ and $\mathcal{K}_2$. Thus, in this case, we expect that

$$\|\mathbf{P}_{\operatorname{col}(\mathbf{G}_1^o)} - \mathbf{P}_{\operatorname{col}(\mathbf{G}_2^o)}\|_2 \approx 1. \tag{3.24}$$

At last, if $\hat{R} < R$ and the rank-one terms that constitute the product $\mathbf{W}\mathbf{Z}_k^T$ are of almost "equal" strength then, we expect that

$$O(\epsilon) \leq \|\mathbf{P}_{\operatorname{col}(\mathbf{G}_1^o)} - \mathbf{P}_{\operatorname{col}(\mathbf{G}_2^o)}\|_2 \lessgtr 1, \tag{3.25}$$

because, $\operatorname{col}(\mathbf{G}_1^o)$ and $\operatorname{col}(\mathbf{G}_2^o)$ will "randomly" capture $\hat{R}$ out of R dimensions of the common spatial subspace.

Accurate expressions for the distance between subspaces $\operatorname{col}(\mathbf{G}_1^o)$ and $\operatorname{col}(\mathbf{G}_2^o)$ lie beyond the scope of this manuscript, require tools from matrix perturbation theory, and pose

stringent assumptions on the size of the noise, which may not be fulfilled in our case. We will check the usefulness of our claims experimentally later.

3.4 Data Compression

Let $\mathbf{X}_k \in \mathbb{R}^{V \times T}$, for $k = 1, \ldots, K$, be a set of full column rank K matrices. As we showed in the section III.A., in order to solve that MAXVAR problem, we have to calculate and decompose a $V \times V$ matrix, which may be prohibitive on a standard computer for large values of V. In this section we show that, when $KT << V$, there is a way for circumvent this problem.

Let matrix $\mathbf{Y} \in \mathbb{R}^{V \times KT}$ denote the matrix that emerges after the concatenation of matrices $\mathbf{X}_k$ along the second dimension, namely

$$\mathbf{Y} = [\mathbf{X}_1 \ldots \mathbf{X}_K] \tag{3.26}$$

and consider a factorization of $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{U}_Y \mathbf{V}_Y^T, \tag{3.27}$$

such that $\mathbf{U}_Y \in \mathbb{R}^{V \times KT}$ with $\mathbf{U}_Y^T \mathbf{U}_Y = \mathbf{I}_{KT}$ and $\mathbf{V}_Y \in \mathbb{R}^{KT \times KT}$. Then, it holds that

$$\mathrm{col}\,(\mathbf{X}_k) \subseteq \mathrm{col}\,(\mathbf{Y}) = \bigcup_{k=1}^{K} \mathrm{col}\,(\mathbf{X}_k). \tag{3.28}$$

Furthermore, we have that

$$\mathrm{col}\,(\mathbf{Y}) \subseteq \mathrm{col}\,(\mathbf{U}_Y). \tag{3.29}$$

Therefore, we can conclude that for each matrix $\mathbf{X}_k$ there exists a matrix $\mathbf{H}_k \in \mathbb{R}^{KT \times T}$ such that

$$\mathbf{X}_k = \mathbf{U}_Y \mathbf{H}_k \tag{3.30}$$

for $k = 1, \ldots, K$ and since $\mathbf{U}_Y^T \mathbf{U}_Y = \mathbf{I}_{KT}$, we also have

$$\mathbf{H}_k = \mathbf{U}_Y^T \mathbf{X}_k \tag{3.31}$$

for $k = 1, \ldots, K$.

Recall that matrix $\mathbf{M}$ is defined as

$$\mathbf{M} = \sum_{k=1}^{K} \mathbf{X}_k \mathbf{X}_k^{\dagger} = \sum_{k=1}^{K} \mathbf{X}_k \left(\mathbf{X}_k^T \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T. \tag{3.32}$$

Due to relation (3.30), we have that

$$\begin{aligned}
\mathbf{M} &= \sum_{k=1}^{K} \mathbf{X}_k \left(\mathbf{X}_k^T \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T \\
&= \mathbf{U}_Y \left(\sum_{k=1}^{K} \mathbf{H}_k \left(\mathbf{H}_k^T \mathbf{H}_k \right)^{-1} \mathbf{H}_k^T \right) \mathbf{U}_Y^T \\
&= \mathbf{U}_Y \tilde{\mathbf{M}} \mathbf{U}_Y^T,
\end{aligned} \tag{3.33}$$

where $\tilde{\mathbf{M}} = \sum_{k=1}^{K} \mathbf{H}_k \left(\mathbf{H}_k^T \mathbf{H}_k \right)^{-1} \mathbf{H}_k^T \in \mathbb{R}^{KT \times KT}$. Consider the eigenvalue decompositions of matrices $\mathbf{M}$ and $\tilde{\mathbf{M}}$

$$\mathbf{M} = \mathbf{U}_M \mathbf{\Lambda}_M \mathbf{U}_M^T \quad \text{and} \quad \tilde{\mathbf{M}} = \mathbf{U}_{\tilde{M}} \mathbf{\Lambda}_{\tilde{M}} \mathbf{U}_{\tilde{M}}^T. \tag{3.34}$$

Since $\mathbf{M} = \mathbf{U}_Y \tilde{\mathbf{M}} \mathbf{U}_Y^T$ and $\mathbf{U}_Y^T \mathbf{U}_Y = \mathbf{I}_{KT}$, we have that

$$\mathbf{\Lambda}_M = \mathbf{\Lambda}_{\tilde{M}} \quad \text{and} \quad \mathbf{U}_M = \mathbf{U}_Y \mathbf{U}_{\tilde{M}}. \tag{3.35}$$

Concluding, since we are interested on R first eigenvectors of matrix $\mathbf{M}$, $[\mathbf{U}_M]_{:,1:R}$ and the following is true

$$[\mathbf{U}_M]_{:,1:R} = \mathbf{U}_Y [\mathbf{U}_{\tilde{M}}]_{:,1:R}, \tag{3.36}$$

it suffices to solve the MAXVAR problem for matrices $\mathbf{H}_k$, calculate $[\mathbf{U}_{\tilde{M}}]_{:,1:R}$, and via $\mathbf{U}_Y$, obtain $\mathbf{G}^o$ without direct calculation of matrix $\mathbf{M}$.

Chapter 4

Experiments

4.1 Synthetic Data

4.1.1 Description of the Datasets

Four categories of simulated datasets were generated according to the data model used in (3.1). For each category, the proposed method was tested on a set of 100 randomly generated datasets. Specifically, vector $\mathbf{a}$ and matrix $\mathbf{A}$ were generated randomly with independent and identically distributed (i.i.d.) elements, taking values uniformly at random in the interval [0,1], while vector $\mathbf{s}$, matrices $\mathbf{S}_k$, and $\mathbf{E}_k$, for $k = 1, \ldots, K$, were generated with i.i.d. elements with $\mathbf{s}_m$, $\mathbf{S}_k(m, r)$, and $\mathbf{E}_k(n, m) \sim \mathcal{N}(0, 1)$. Next, we quote the factor dimensions for each category.

	category (i)	category (ii)	category (iii)
N	1,000	5,000	100,000
M	75	100	100
K	10	20	25
R	5	10	30

In Figures 4.1–4.9, we present the mean absolute correlation coefficients for factors $\boldsymbol{\lambda}$, $\mathbf{a}$, and $\mathbf{s}$, respectively, for categories (i), (ii), and (iii), under different SNR values. In our experiments, we used the following SNR definition

$$\text{SNR} = \frac{\sum_{k=1}^{K} \left\| \lambda_k \mathbf{a s}^T \right\|_F^2}{\sum_{k=1}^{K} \left\| \mathbf{A S}_k^T + \mathbf{E}_k \right\|_F^2}. \tag{4.1}$$

For each realization, we controlled the SNR value by appropriately tuning the coefficient β in the following relation

$$\mathbf{X}_k = \lambda_k \mathbf{a s}^T + \beta \left(\mathbf{A S}_k^T + \mathbf{E}_k \right). \tag{4.2}$$

Finally, three methods were tested for each synthetic dataset. The first method was to directly attempt to retrieve factors $\boldsymbol{\lambda}$, $\mathbf{a}$, and $\mathbf{s}$ from matrices $\{\mathbf{X}_k\}_{k=1}^{K}$, via solving the

following problem

$$\min_{\boldsymbol{\lambda}\geq0,\mathbf{a}\geq0,\mathbf{s}} \sum_{k=1}^{K} \|\mathbf{X}_k - \lambda_k \mathbf{a}\mathbf{s}^T\|_F^2. \tag{4.3}$$

in an alternating fashion (ALS minimization). The second method was to estimate the common spatial subspace $\mathbf{G}^o$ first and then to estimate factors $\boldsymbol{\lambda}$, $\mathbf{a}$, and $\mathbf{s}$ from matrices $\left\{\tilde{\mathbf{X}}_k\right\}_{k=1}^{K}$, as solutions of

$$\min_{\boldsymbol{\lambda}\geq\mathbf{0},\mathbf{a}\geq\mathbf{0},\mathbf{s}} \sum_{k=1}^{K} \|\tilde{\mathbf{X}}_k - \lambda_k \mathbf{a}\mathbf{s}^T\|_F^2. \tag{4.4}$$

As for the third method, both common spatial subspace $\mathbf{G}^o$ and common temporal component $\mathbf{g}^o$ were estimated first and then factors $\boldsymbol{\lambda}$ and $\mathbf{a}$ were estimated via ALS minimization of the problem in (3.19). Since ALS minimization is sensitive to local minima, each optimization problem was solved for 3 different initializations and only the solution that achieved the lowest value of the cost function was kept.
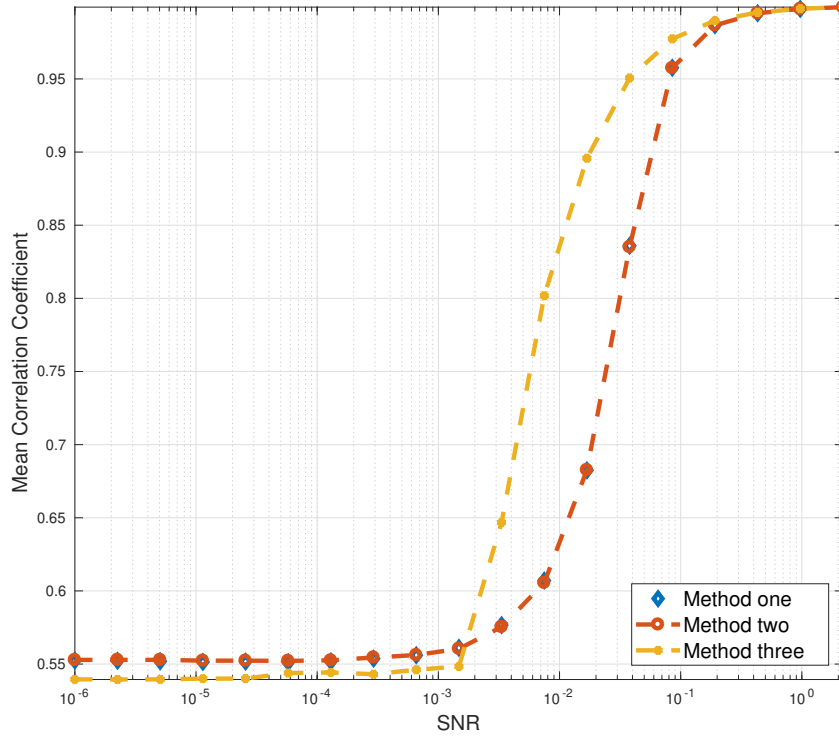

Figure 4.1: Mean correlation coefficient between $\boldsymbol{\lambda}_{true}$ and $\boldsymbol{\lambda}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (i).
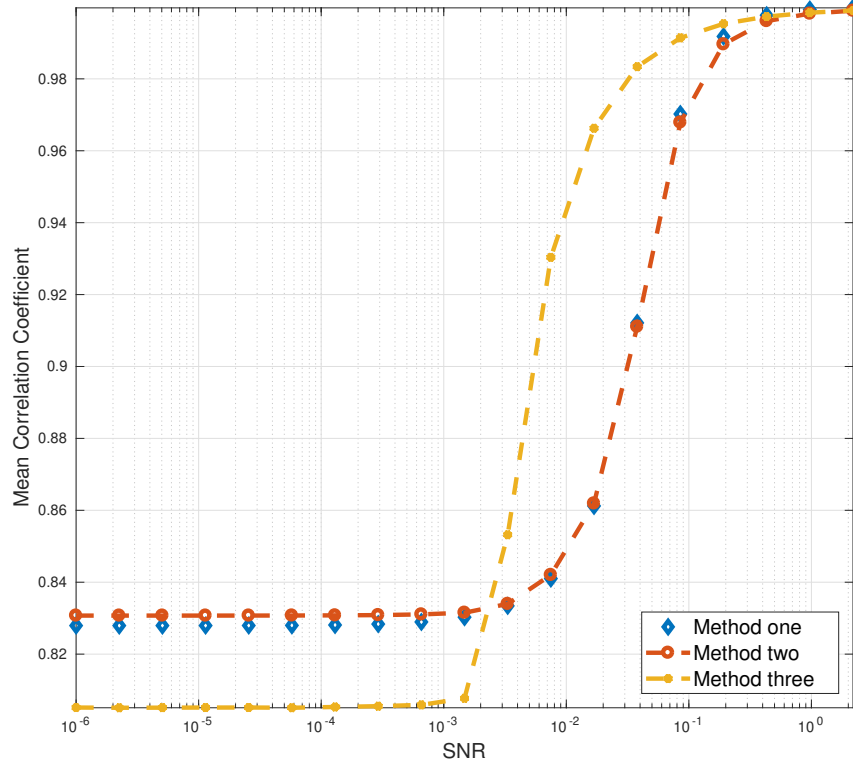
Figure 4.2: Mean correlation coefficient between $\mathbf{a}_{true}$ and $\mathbf{a}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (i).
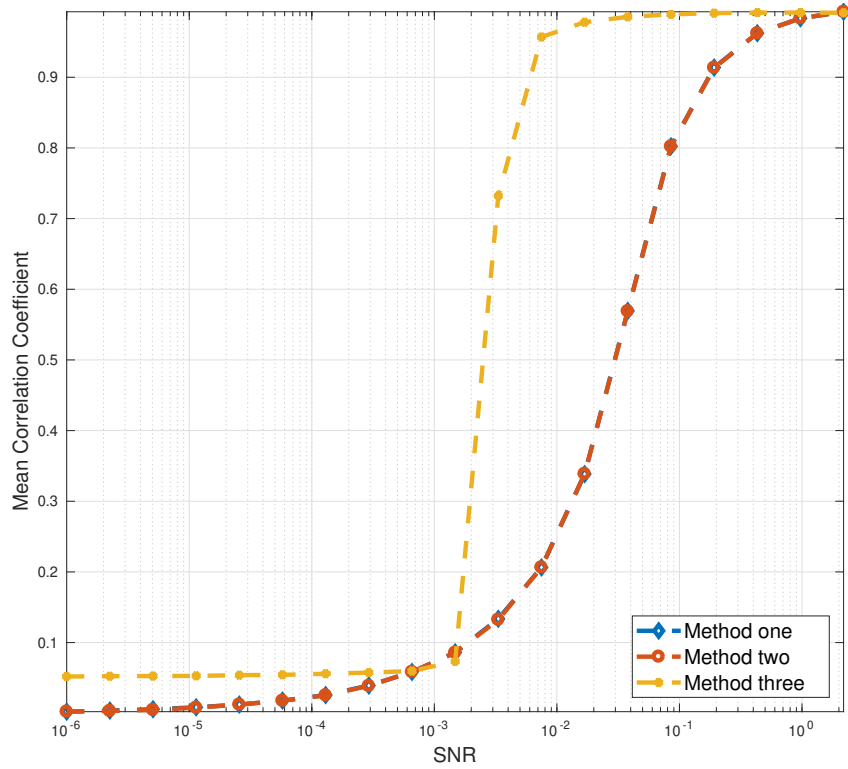

Figure 4.3: Mean correlation coefficient between $\mathbf{s}_{true}$ and $\mathbf{s}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (i).
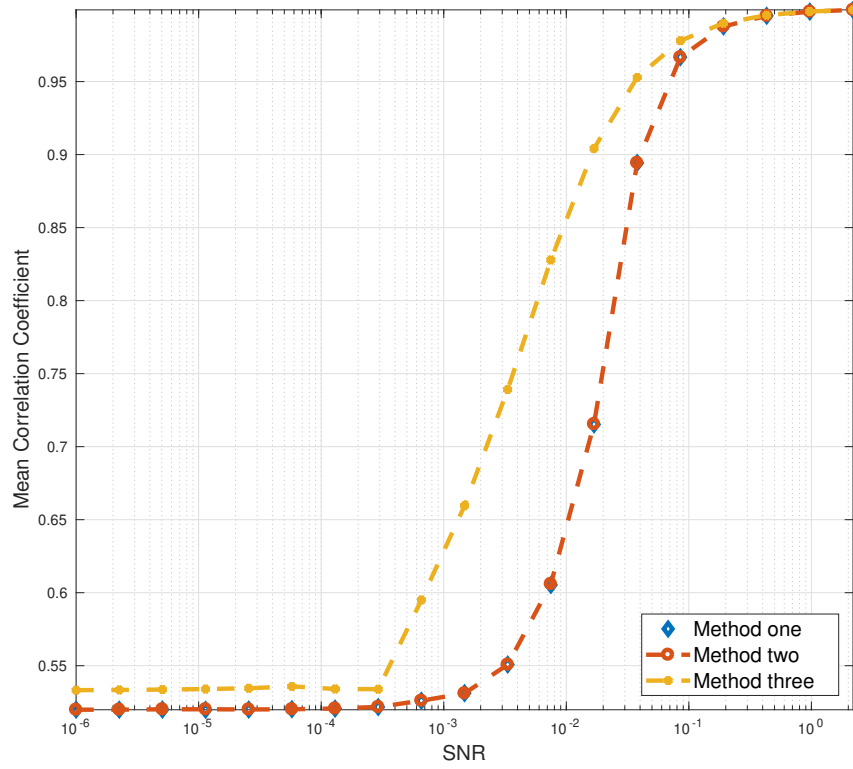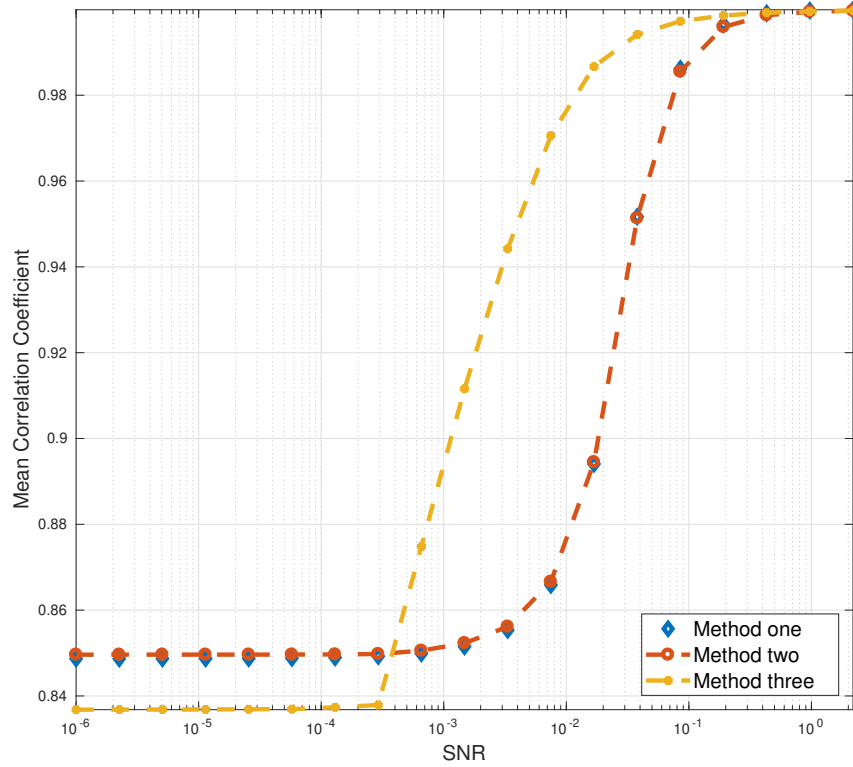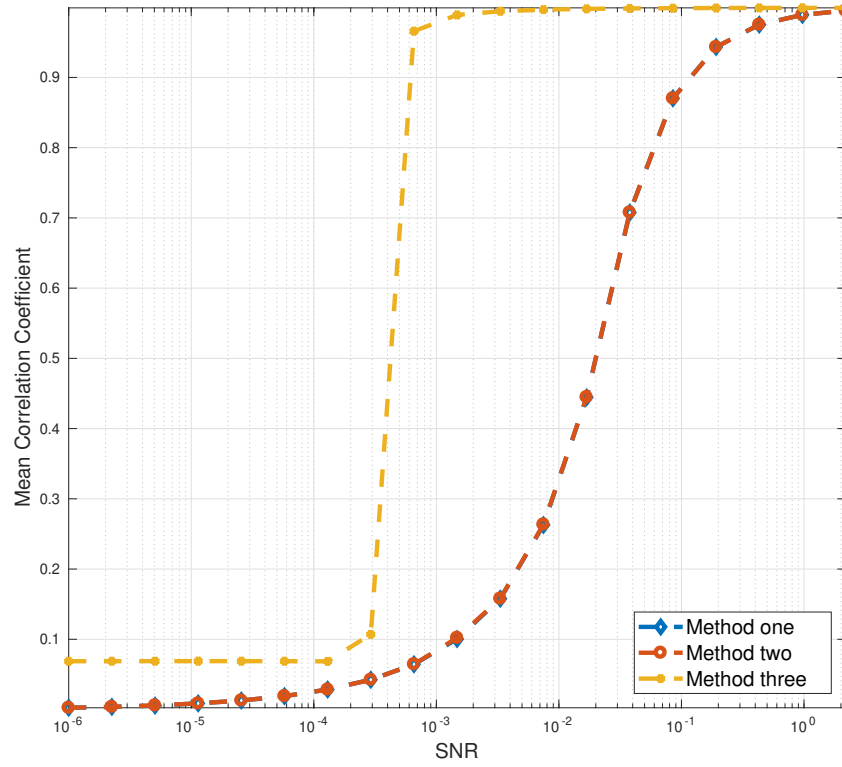
Figure 4.4: Mean correlation coefficient between $\boldsymbol{\lambda}_{true}$ and $\boldsymbol{\lambda}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (ii).

Figure 4.5: Mean correlation coefficient between $\mathbf{a}_{true}$ and $\mathbf{a}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (ii).

Figure 4.6: Mean correlation coefficient between $\mathbf{s}_{true}$ and $\mathbf{s}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (ii).
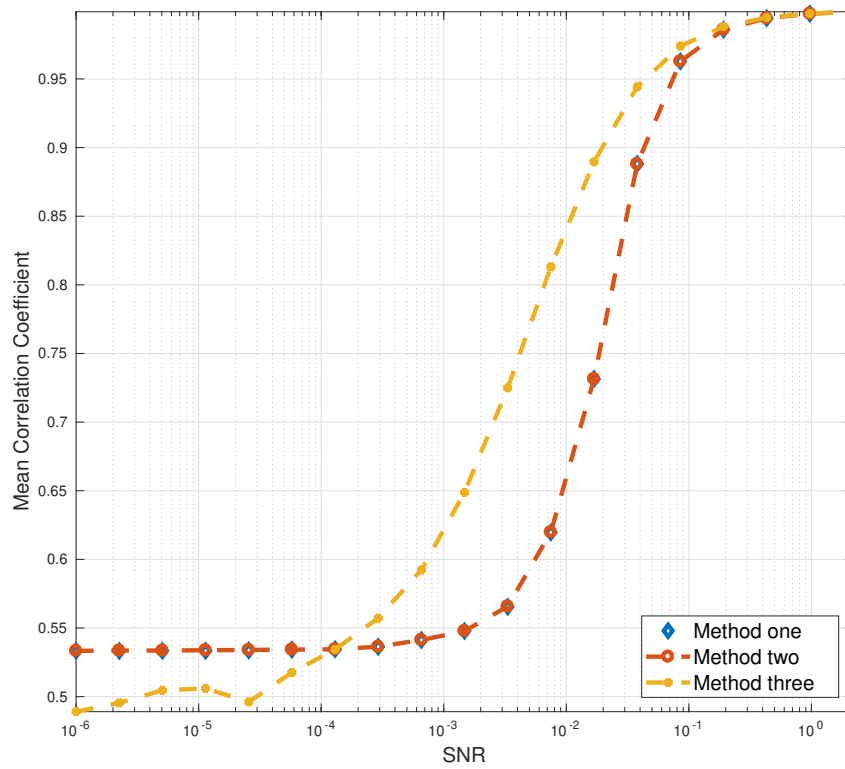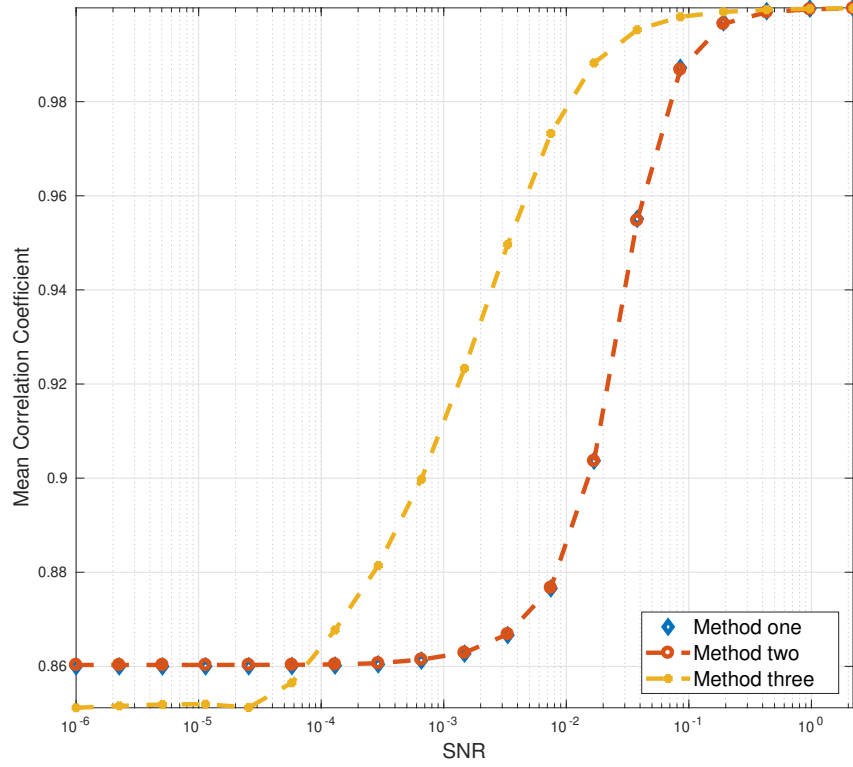

Figure 4.7: Mean correlation coefficient between $\boldsymbol{\lambda}_{true}$ and $\boldsymbol{\lambda}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (iii).

Figure 4.8: Mean correlation coefficient between $\mathbf{a}_{true}$ and $\mathbf{a}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (iii).
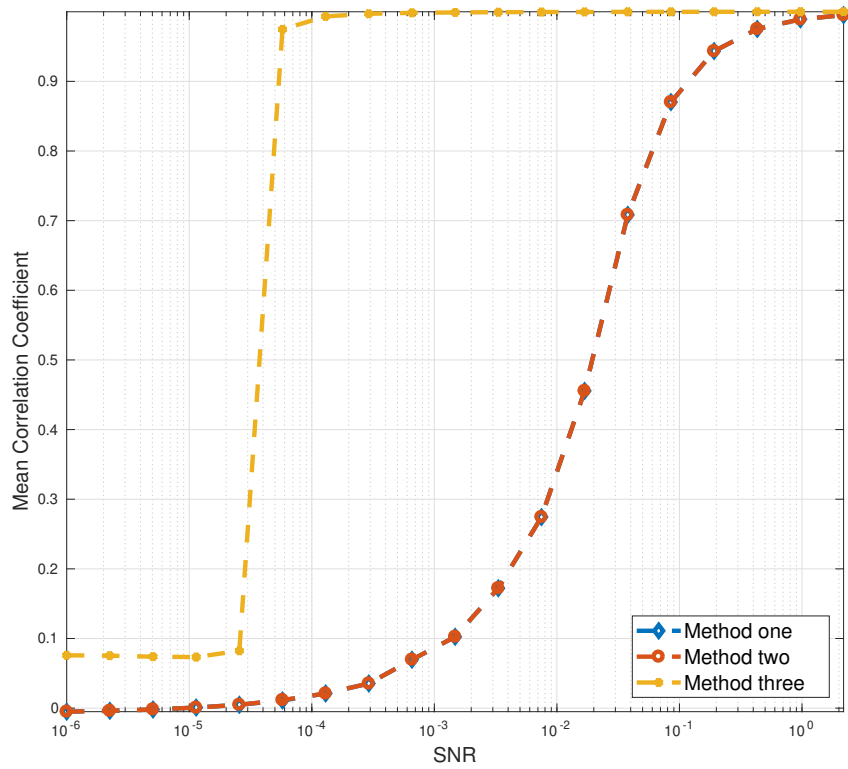

Figure 4.9: Mean correlation coefficient between $\mathbf{s}_{true}$ and $\mathbf{s}_{est}$, over 100 runs and different SNR values, for the three considered methods in category (iii).

4.2 Real World Data

In this section, we test our approach using real-world task-related fMRI data. Specifically, we process four datasets, recorded at the University of Crete General Hospital, from a group of 25 healthy adults performing four visual tasks which were identical in all but one aspect (the precise kinematics of an observed person-directed action). First, we quote some information regarding the experiment design and the preprocessing pipeline that was applied to the data and, then, we present the results obtained by analyzing the data using our method.

4.2.1 Experiment design

The fMRI block design consists of four action observation conditions, each involving four "active" 35 sec blocks alternating with four 35 sec baseline blocks. Indicative specifications are presented below. A video clip illustrating a two-movement action sequence was presented 6 times within each "active" block. The stimulus set-up was identical across blocks and conditions, consisting of a fixed red spot at the center of the display, presenting a female person sitting behind a table. A white tea cup was positioned on the table and a ceramic bowl 30 cm in diameter was located on a smaller table right next to the person's head. The data employed in the main analyses reported here were derived from four experimental conditions examining the effects of an action with the same goal but different kinematics:

(i) Fast to cup - Slow to person: It consists of a rapid grasping movement toward the tea cup (time duration equal to 700 ms and average velocity equal to 0.64 m/sec), followed by a much slower movement that brings the cup to the person's mouth (time duration equal to 3300 ms and average velocity equal to 0.14 m/sec).

(ii) Slow to cup - Fast to person: It consists of a slow grasping movement toward the tea cup (time duration equal to 3300 ms and average velocity equal to 0.14 m/sec), followed by a much faster movement that brings the cup to the person's mouth (time duration equal to 700 ms and average velocity equal to 0.64 m/sec). The stimulus layout is identical to (i).

(iii) Fast to cup - Slow to bowl: It consists of a rapid grasping movement toward the tea cup (time duration equal to 700 *ms* and average velocity equal to 0.64 m/sec), followed by a much slower movement that brings the cup over the edge of the bowl

(time duration equal to 3300 ms and average velocity equal to 0.14 m/sec) which is located in the position of the person's head in conditions (i) and (ii).

(iv) Aimless action: The proximal and distal tables seen in conditions (i-iii) constitute the visual background in a 5 sec video clip depicting an extended hand executing a two-step motion toward the center of the proximal table and then toward (but not reaching) the distal table). Peak motion velocity was the same for the first and second step of the action, which was repeated 6 times within each "active"' block.

Each recording session consisted of 4 35 sec "active" blocks alternating with 4 35 sec "baseline" blocks. The baseline was identical in the two conditions and consisted of the repeated presentation of a static hand over the same background setting displayed in the "active" blocks. A sample sequence of frames is shown in Figure 4.10.
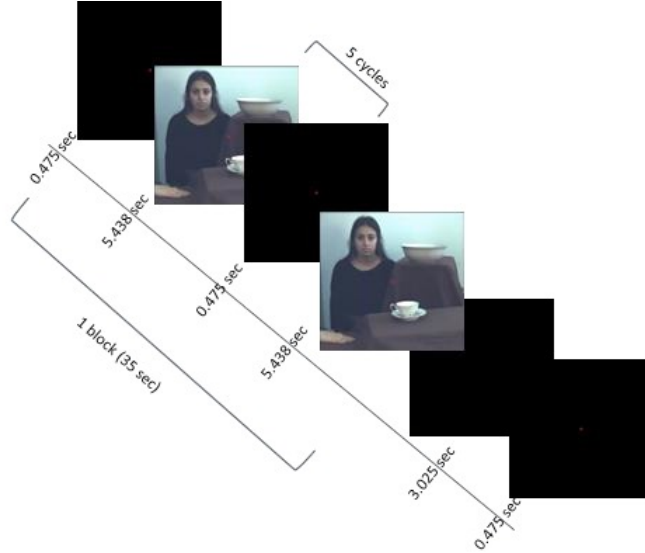

Figure 4.10: A sample sequence of frames from the described experiment

4.2.2   Image acquisition and pre-processing

For the BOLD-fMRI, a $T2^*$-weighted, fat-saturated $2D$-FID-EPI sequence was used with the following parameters: repetition time (TR) 3500 ms, echo time (TE) 50 ms, field of view (FOV) $192 \times 192 \times 108$ (x, y, z), acquisition voxel size $3 \times 3 \times 3$ mm. Whole brain scans consisted of 36 transverse slices with 3.0-mm slice thickness and no interslice gap. The time series recorded in each condition comprised 80 volumes (time points), with 40 volumes recorded during observation of repeated person-directed action and 40 volumes recorded

during observation of repeated presentation of the static hand. In all analyses the first 5 volumes of each time series were ignored as is customary in fMRI studies. Additionally, high resolution anatomical images were acquired sagittally, using a $3D$ magnetization-prepared rapid acquisition gradient echo sequence ($3D$-MPRAGE) with the following parameters: TR 9.8 ms, TE 4.6 ms, flip angle 8 deg, inversion time (TI) 922 ms, FOV 180×230 (x, z), with acquisition voxel size of 0.98×0.98 (x, z) and slice thickness of 1 mm.

For both datasets, image preprocessing was performed in SPM8 (Statistical Parametric Mapping software, SPM: Welcome Department of Imaging Neuroscience, London, UK; available at: http://www.fil.ion.ucl.ac.uk/spm/). Initially, EPI scans were spatially re-aligned to the first image of the first time-series using second-degree B-spline interpolation algorithms and motion-corrected through rigid body transformations (three translations and three rotations about each axis). Next, images were spatially normalized to a common brain space (MNI template) and smoothed using an isotropic Gaussian filter (FWHM=8 mm). At last, the SPM platform is able to provide a time response component, based on the activation onsets and offsets, which is expected to appear in the activated brain voxels.

From now on, we denote this expected response as $\mathbf{s}_{\mathrm{exp}}$, which will be the same for both conditions, since, as we mentioned above, the stimulus layout, between the two conditions, is identical.

(a) Retrieved $\mathbf{g}^o$ for varying common subspace dimensions, from 10 to 40, for condition (i). The signal depicted with blue stars is the $\mathbf{s}_{exp}$.

(b) Absolute correlation coefficient between $\mathbf{s}_{exp}$ and $\mathbf{g}^o$ across different common subspace dimensions.
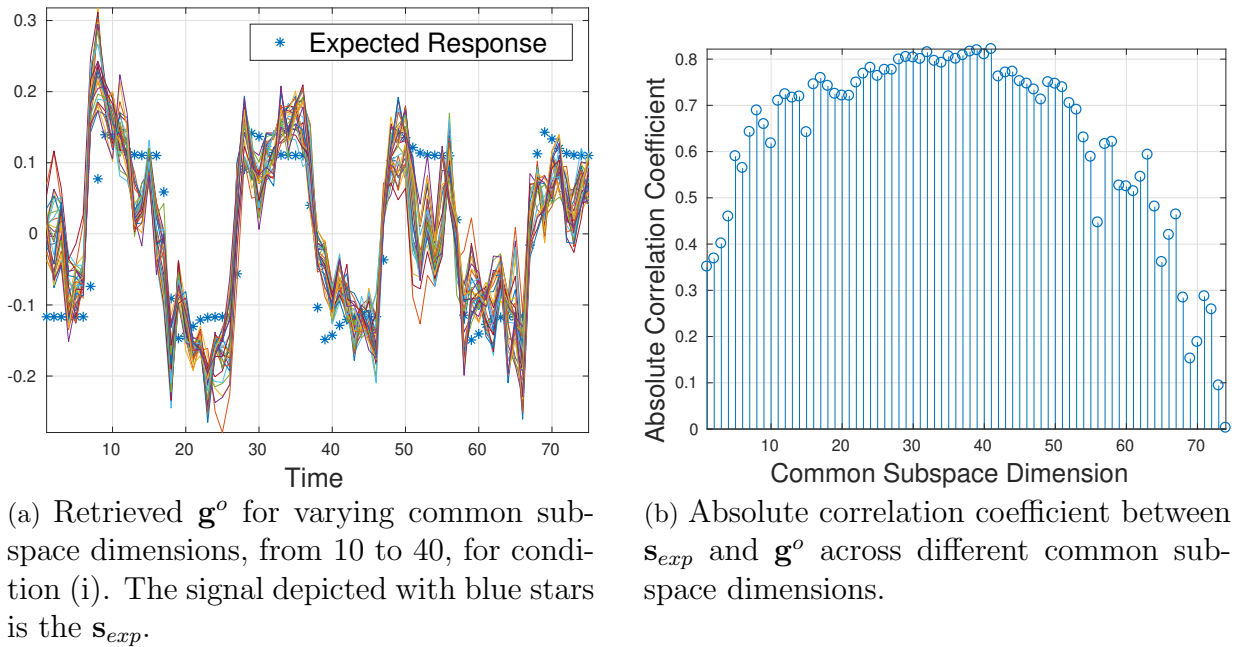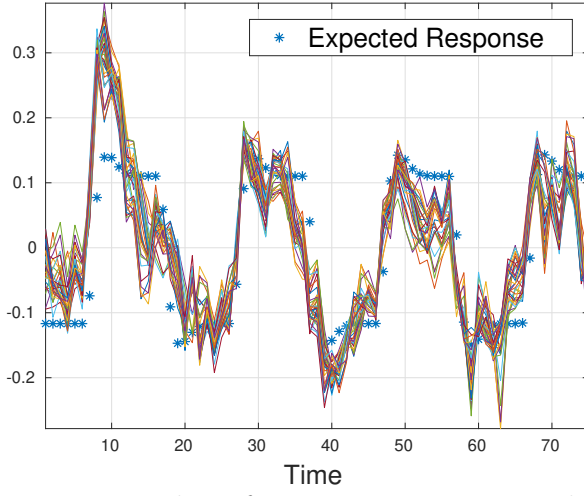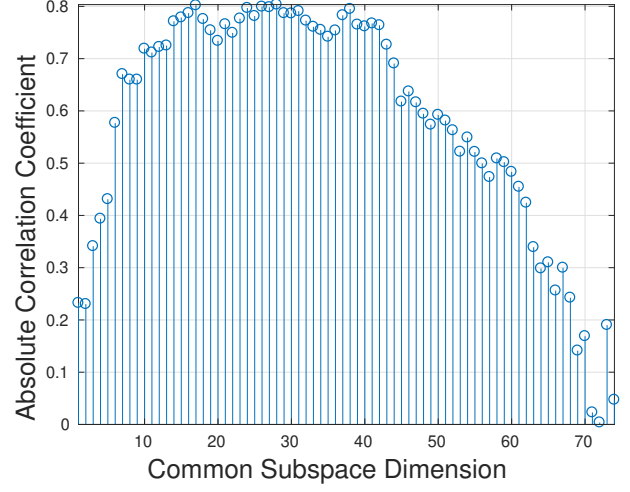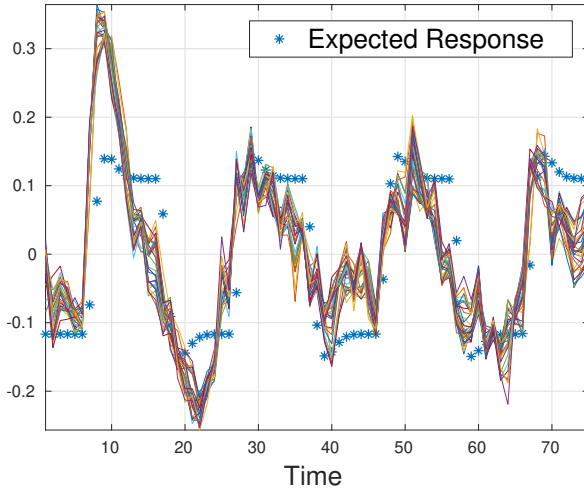
Figure 4.11: Results from condition (i).

(a) Retrieved $\mathbf{g}^o$ for varying common subspace dimensions, from 10 to 40, for condition (ii). The signal depicted with blue stars is the $\mathbf{s}_{exp}$.
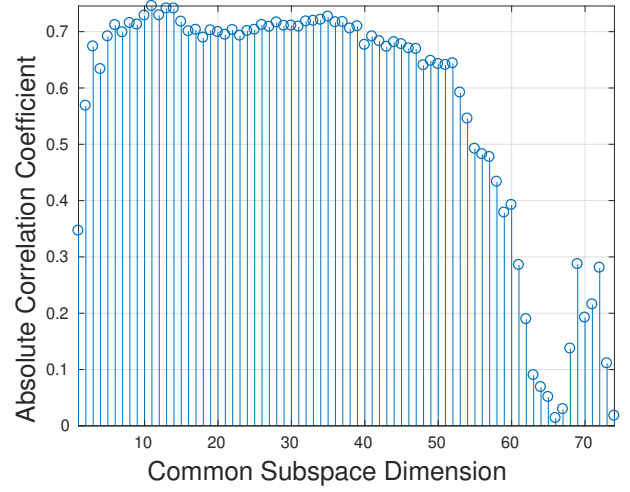
(b) Absolute correlation coefficient between $\mathbf{s}_{exp}$ and $\mathbf{g}^o$ across different common subspace dimensions.
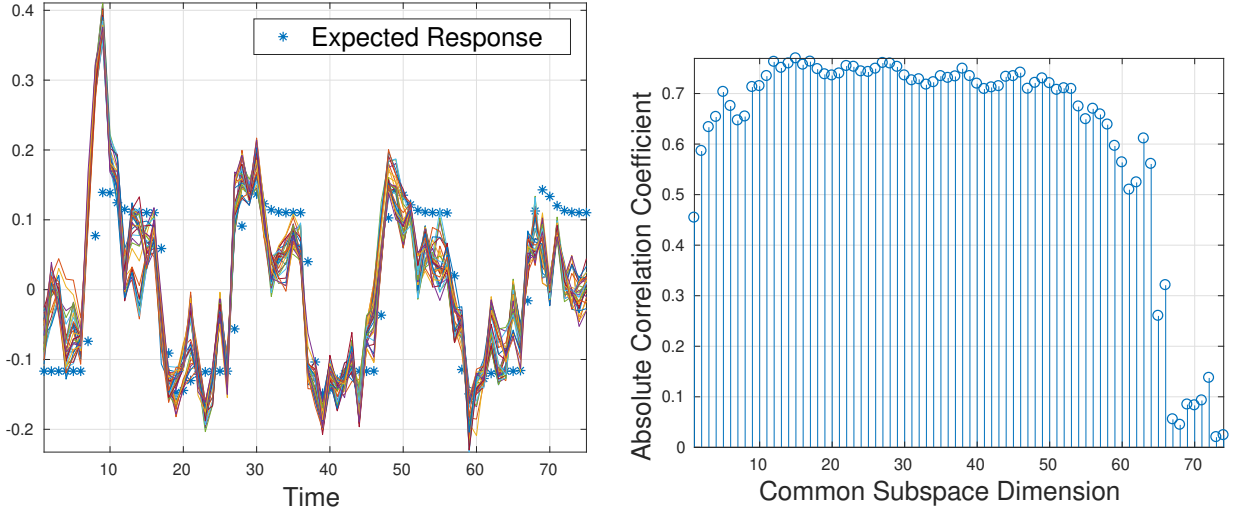
Figure 4.12: Results from condition (ii).

(a) Retrieved $\hat{\mathbf{s}}_0$ for varying common subspace dimensions, from 10 to 40, for condition (iii). The signal depicted with blue stars is the $\mathbf{s}_{exp}$.

(b) Absolute correlation coefficient between $\mathbf{s}_{exp}$ and $\hat{\mathbf{s}}_0$ across different common subspace dimensions.

Figure 4.13: Results from condition (iii).

(a) Retrieved $\hat{\mathbf{s}}_0$ for varying common subspace dimensions, from 10 to 40, for condition (iv). The signal depicted with blue stars is the $\mathbf{s}_{exp}$.

(b) Absolute correlation coefficient between $\mathbf{s}_{exp}$ and $\hat{\mathbf{s}}_0$ across different common subspace dimensions.

Figure 4.14: Results from condition (iv).

4.2.3 Results

Next, we present the results from our gCCA based analysis of the datasets from all conditions (i)-(iv). More specifically, in Figures 4.11 - 4.14, we illustrate for each dataset, respectively,

- (left) the extracted common temporal component $\mathbf{g}^o$ that emerged for various common spatial subspace dimensions, as well as the normalized, to unit 2-norm, expected response $\mathbf{s}_{exp}$.

- (right) the absolute correlation coefficients between $\mathbf{s}_{exp}$ and the extracted common temporal component $\mathbf{g}^o$, that emerged for all possible common subspace dimensions.

For all conditions, we observe that:

- the estimated common temporal components $\mathbf{g}^o$, for different common spatial subspace dimensions, are very much alike; this implies that our estimate is *not* sensitive to the common subspace dimension, which is unknown, in general. Thus, we can get useful results over a wide range of values of the dimension of the common subspace.

- the estimated common temporal components $\mathbf{g}^o$ are quite similar to the expected signal $\mathbf{s}_{exp}$. Their correlation coefficients take values at about 0.7-0.8, over a wide

range of dimensions, while they outreach 0.8 in some cases.

Consequently, we conclude that our method effectively estimates the common temporal component, without using prior knowledge about its shape.

Regarding the rank determination procedure for the datasets from conditions (i) and (ii), in Figures 4.18, 4.22, 4.26, and 4.30, we depict the gap function between two estimates of the "common" subspace, $\mathcal{S}_1$ and $\mathcal{S}_2$, for all possible dimensions, starting from 1 to 74 (length of each voxel time series minus one). We observe that the value of the gap function becomes approximately equal to 1, for all dimensions larger than a dimension of around 20. On the other hand, for lower dimension the gap function takes values between zero and one. This observation indicates the appropriateness of a low rank model, while it enable us to obtain an estimate for the rank of the model.

As for the extracted spatial components, in Figures 4.15, 4.17, 4.19, 4.21, 4.23, 4.25, 4.27, and 4.29, we present the spatial maps that emerged from the two proposed approaches, for all conditions, while in Figures 4.16, 4.20, 4.24, and 4.28, we present the intensities across subjects of the rank one term that emerged from applying the second method.

For both conditions, we can deduce that voxels from the occipital and parietal lobes present a strong activation, across almost all subjects. Furthermore, we can observe that voxels from the premotor cortex (PMv area) and inferior frontal gyrus also present activation. Both observations are consistent with the nature of the experiment.
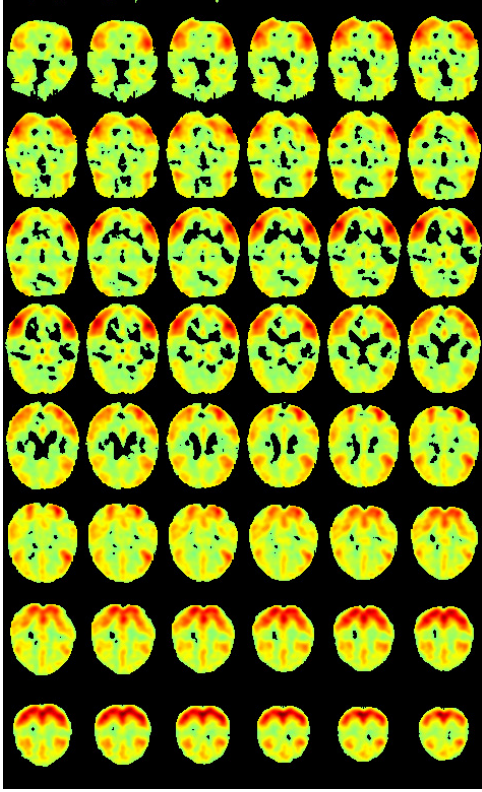
Figure 4.15: Nonnegative map $\mathbf{a}^o$ calculated for common subspace dimension equal to 23, for condition (i).
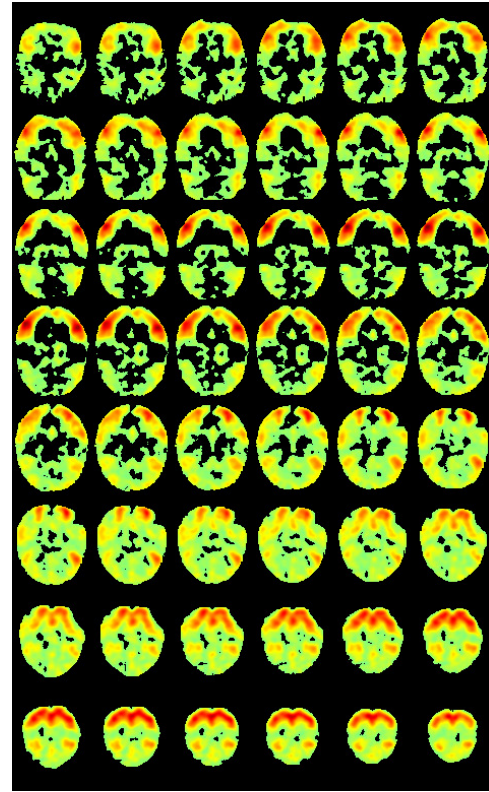

Figure 4.17: Projected mean map $\hat{\mathbf{a}}$ calculated for common subspace dimension equal to 23, for condition (i).
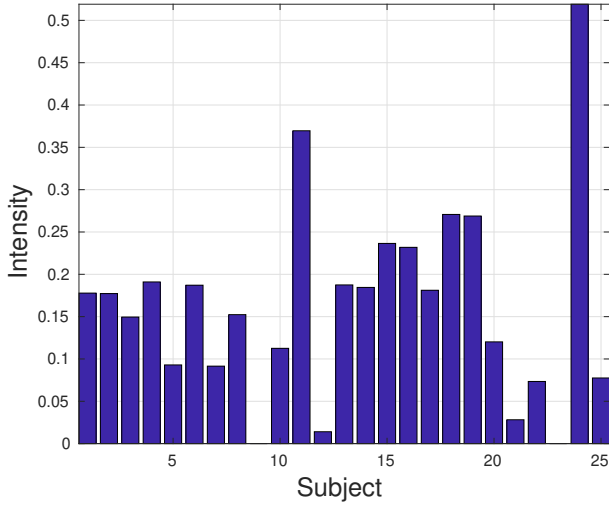

Figure 4.16: Intensities of the rank one term $\boldsymbol{\lambda}$ across subjects for the nonnegative case and "common" subspace dimension equal to 23, for condition (i).
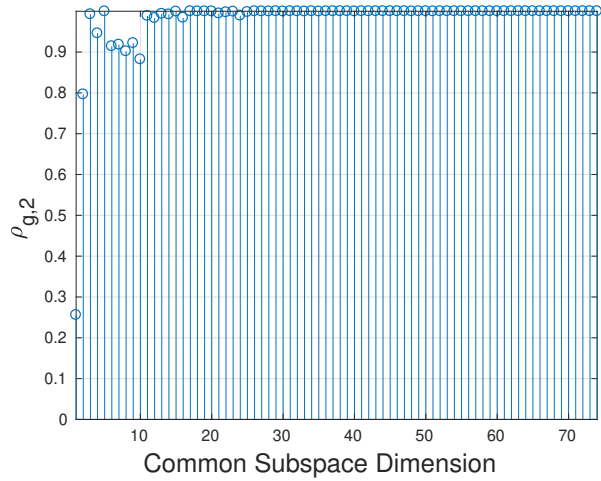

Figure 4.18: Gap function $\rho_{g,2}\left(\mathcal{S}_1, \mathcal{S}_2\right)$ evaluated for varying dimension of $\mathcal{S}_1$ and $\mathcal{S}_2$, for condition (i).
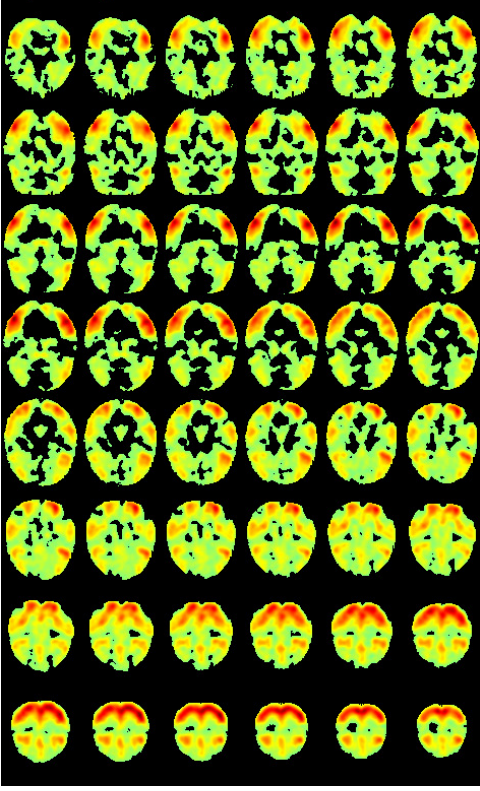
Figure 4.19: Nonnegative map $\mathbf{a}^o$ calculated for common subspace dimension equal to 23, for condition (ii).
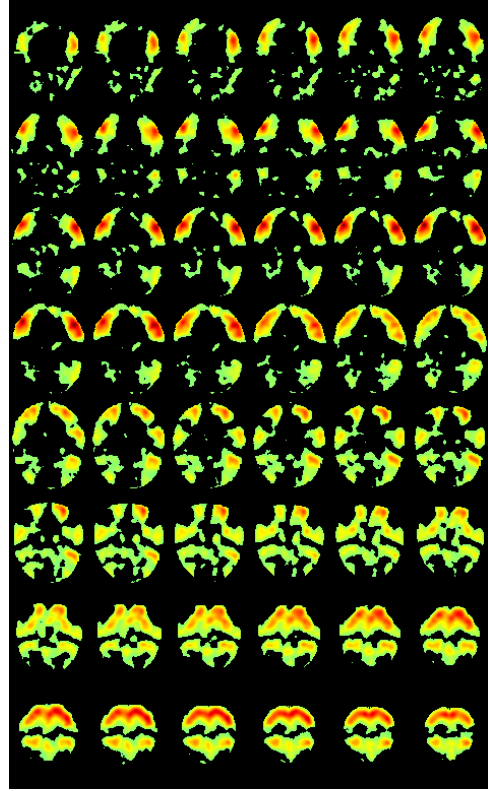

Figure 4.21: Projected mean map $\hat{\mathbf{a}}$ calculated for common subspace dimension equal to 23, for condition (ii).

Figure 4.20: Intensities of the rank one term $\boldsymbol{\lambda}$ across subjects for the nonnegative case and "common" subspace dimension equal to 23, for condition (ii).

Figure 4.22: Gap function $\rho_{g,2}\left(\mathcal{S}_1, \mathcal{S}_2\right)$ evaluated for varying dimension of $\mathcal{S}_1$ and $\mathcal{S}_2$, for condition (ii).

Figure 4.23: Nonnegative map $\mathbf{a}^o$ calculated for common subspace dimension equal to 23, for condition (iii).

Figure 4.25: Projected mean map $\hat{\mathbf{a}}$ calculated for common subspace dimension equal to 23, for condition (iii).
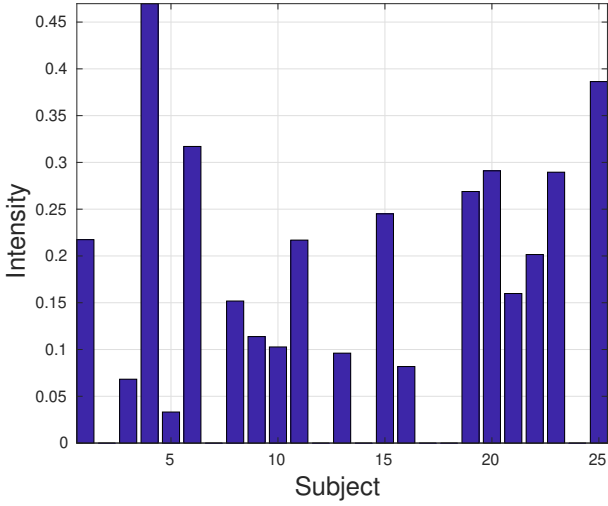

Figure 4.24: Intensities of the rank one term across subject for the nonnegative case and "common" subspace dimension equal to 23, for condition (iii).
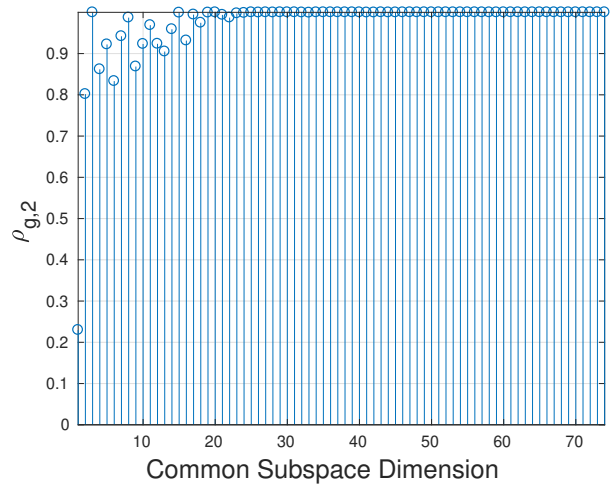

Figure 4.26: Gap function $\rho_{g,2}(\mathcal{S}_1, \mathcal{S}_2)$ evaluated for varying dimension of $\mathcal{S}_1$ and $\mathcal{S}_2$, for condition (iii).
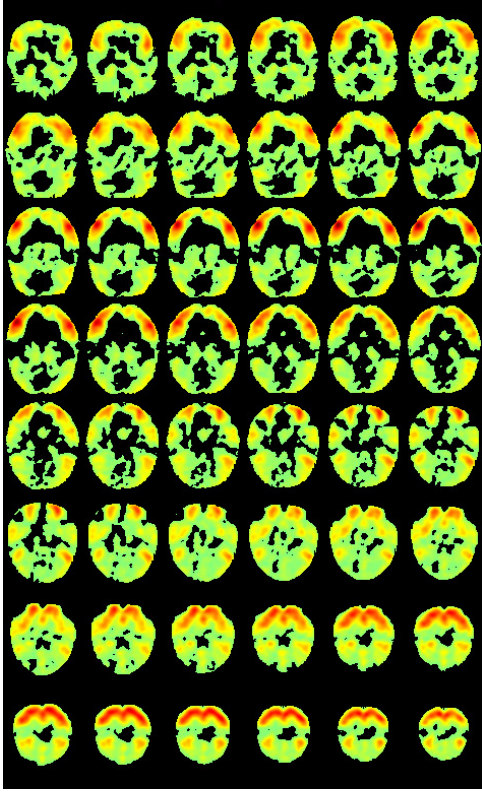
Figure 4.27: Nonnegative map $\mathbf{a}^o$ calculated for common subspace dimension equal to 23, for condition (iv).
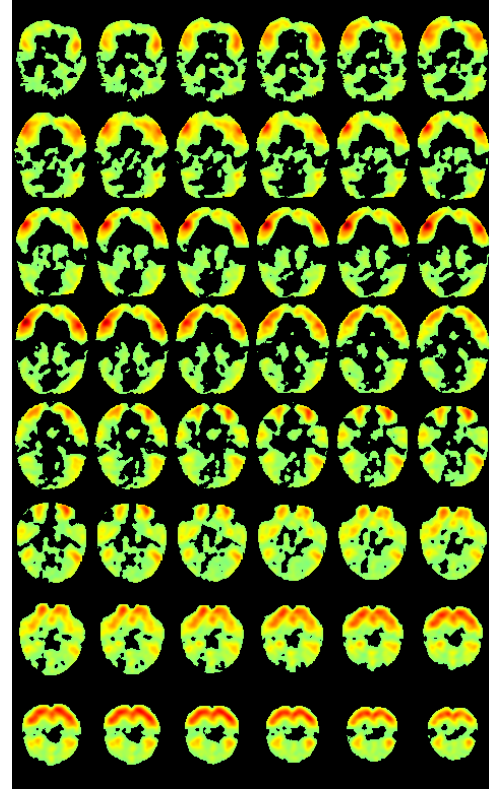

Figure 4.29: Projected mean map $\hat{\mathbf{a}}$ calculated for common subspace dimension equal to 23, for condition (iv).
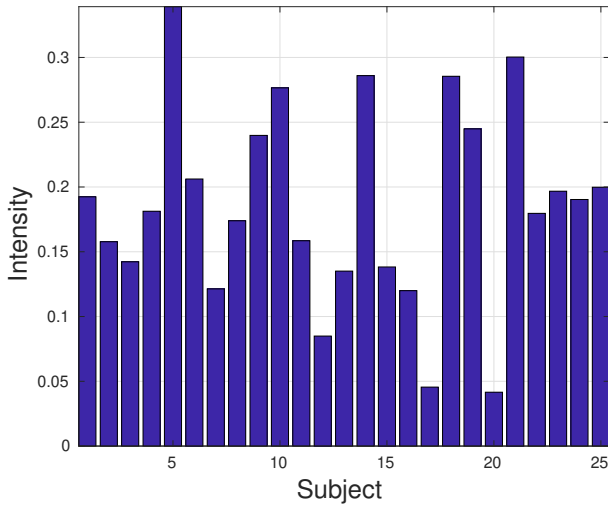

Figure 4.28: Intensities of the rank one term across subject for the nonnegative case and "common" subspace dimension equal to 23, for condition (iv).
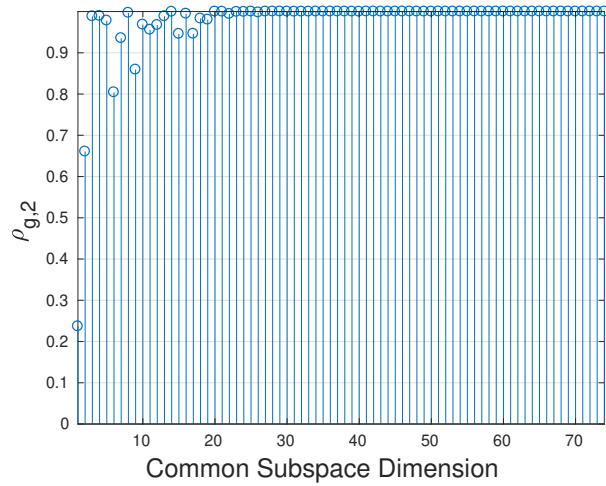

Figure 4.30: Gap function $\rho_{g,2}\left(\mathcal{S}_1, \mathcal{S}_2\right)$ evaluated for varying dimension of $\mathcal{S}_1$ and $\mathcal{S}_2$, for condition (iv).
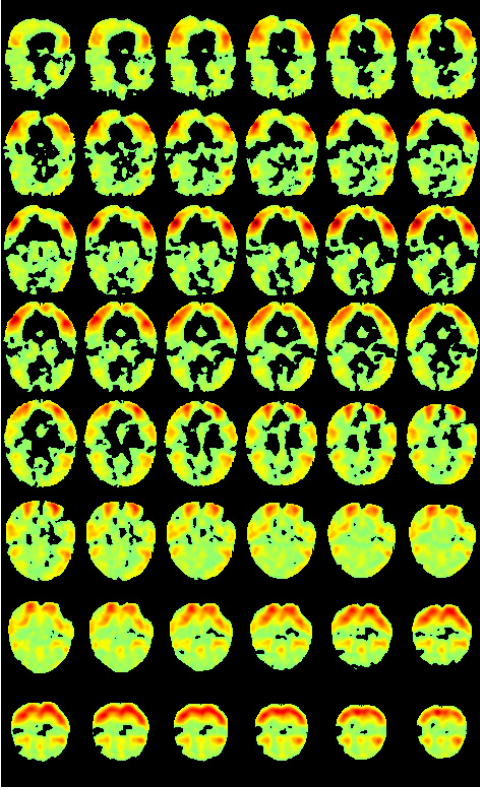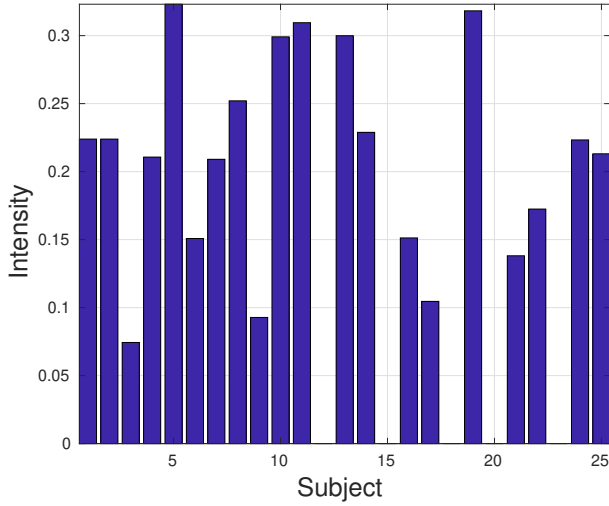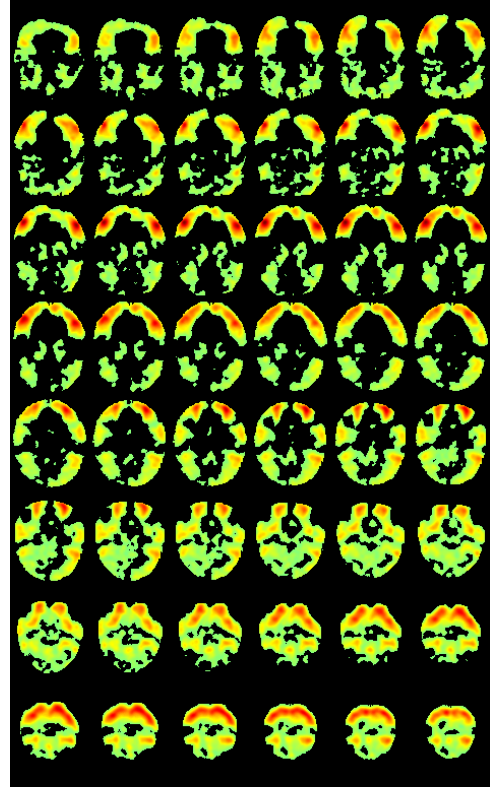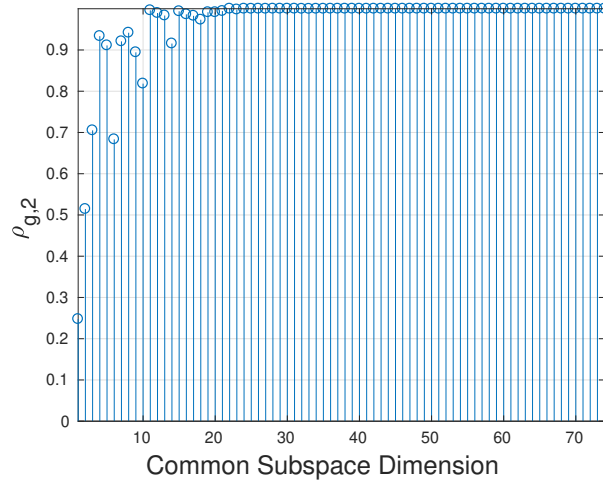
Chapter 5

Conclusion & Future Work

5.1 Conclusion

In this work, we considered the problem of multi-subject task-related fMRI data analysis with one type of stimulus. We proposed a data generating model that takes into account both task-related and resting-state components. We used two successive gCCAs and computed, blindly, an estimate of the common spatial subspace, the common temporal component, and the common activation map. Moreover, we proposed a method for the rank determination problem. At last, we tested our approach, using both synthetic and real-world data.

5.2 Future Work

Experimental designs with multiple types of stimuli are popular in the fMRI area. Therefore, extending our methods for the case of multiple types of stimuli/tasks is challenging. In the proposed model, besides the common activation term, there are structured terms that express the brain activity at rest. Recovering the resting-state components is also challenging due to the ambiguities of the matrix factorization problem.

Besides fMRI, many other techniques have been employed in the endeavor of deciphering the human brain, like EEG, MEG, and others. Testing our methods in data analysis from other modalities and exploiting gCCA in concurrent data analysis from different modalities is intriguing.

At last, tensor decomposition models for multidimensional data analysis have attracted intense research interest over the last years. Combining and adapting gCCA with/to popular tensor models like PARAFAC, Tucker, BTD, and others, should be taken into consideration. Finally, devising parallel algorithms for gCCA based analysis of large collections of data should be considered.

Bibliography

[1] A. Rayshubskiy, T. J. Wojtasiewicz, C. B. Mikell, M. B. Bouchard, D. Timerman, B. E. Youngerman, R. A. McGovern, M. L. Otten, P. Canoll, G. M. McKhann *et al.*, "Direct, intraoperative observation of˜ 0.1 hz hemodynamic oscillations in awake human cortex: implications for fmri," *Neuroimage*, vol. 87, pp. 323–331, 2014.

[2] F. G. Ashby, *An introduction to fMRI.* Springer, 2015.

[3] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nature reviews neuroscience*, vol. 8, no. 9, p. 700, 2007.

[4] D. A. Fair, B. L. Schlaggar, A. L. Cohen, F. M. Miezin, N. U. Dosenbach, K. K. Wenger, M. D. Fox, A. Z. Snyder, M. E. Raichle, and S. E. Petersen, "A method for using blocked and event-related fMRI data to study resting state functional connectivity," *Neuroimage*, vol. 35, no. 1, pp. 396–405, 2007.

[5] M. D. Fox, A. Z. Snyder, J. M. Zacks, and M. E. Raichle, "Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses," *Nature neuroscience*, vol. 9, no. 1, p. 23, 2006.

[6] M. D. Greicius and V. Menon, "Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation," *Journal of cognitive neuroscience*, vol. 16, no. 9, pp. 1484–1492, 2004.

[7] K. Arfanakis, D. Cordes, V. M. Haughton, C. H. Moritz, M. A. Quigley, and M. E. Meyerand, "Combining independent component analysis and correlation analysis to probe interregional connectivity in fMRI task activation datasets," *Magnetic resonance imaging*, vol. 18, no. 8, pp. 921–930, 2000.

[8] P. Fransson, "How default is the default mode of brain function?: Further evidence from intrinsic BOLD signal fluctuations," *Neuropsychologia*, vol. 44, no. 14, pp. 2836–2845, 2006.

[9] A. H. Andersen, D. M. Gash, and M. J. Avison, "Principal component analysis of the dynamic response measured by fMRI: a generalized linear systems framework," *Magnetic Resonance Imaging*, vol. 17, no. 6, pp. 795–815, 1999.

[10] L. K. Hansen, J. Larsen, F. Å. Nielsen, S. C. Strother, E. Rostrup, R. Savoy, N. Lange, J. Sidtis, C. Svarer, and O. B. Paulson, "Generalizable patterns in neuroimaging: How many principal components?" *NeuroImage*, vol. 9, no. 5, pp. 534–544, 1999.

[11] M. J. McKeown and T. J. Sejnowski, "Independent component analysis of fMRI data: examining the assumptions," *Human brain mapping*, vol. 6, no. 5-6, pp. 368–372, 1998.

[12] V. Calhoun, T. Adali, G. Pearlson, and J. Pekar, "Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms," *Human brain mapping*, vol. 13, no. 1, pp. 43–53, 2001.

[13] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "ICA of functional MRI data: an overview," in *in Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation.* Citeseer, 2003.

[14] E. B. Erhardt, S. Rachakonda, E. J. Bedrick, E. A. Allen, T. Adali, and V. D. Calhoun, "Comparison of multi-subject ICA methods for analysis of fMRI data," *Human brain mapping*, vol. 32, no. 12, pp. 2075–2095, 2011.

[15] A. H. Andersen and W. S. Rayens, "Structure-seeking multilinear methods for the analysis of fMRI data," *NeuroImage*, vol. 22, no. 2, pp. 728–739, 2004.

[16] C. F. Beckmann and S. M. Smith, "Tensorial extensions of independent component analysis for multisubject fMRI analysis," *Neuroimage*, vol. 25, no. 1, pp. 294–311, 2005.

[17] A. Stegeman, "Comparing independent component analysis and the parafac model for artificial multi-subject fMRI data," *Unpublished Technical Report, Univ. of Groeningen*, 2007.

[18] M. Mørup, L. K. Hansen, S. M. Arnfred, L.-H. Lim, and K. H. Madsen, "Shift-invariant multilinear decomposition of neuroimaging data," *NeuroImage*, vol. 42, no. 4, pp. 1439–1450, 2008.

[19] M. Mørup, L. K. Hansen, and K. H. Madsen, "Modeling latency and shape changes in trial based neuroimaging data," in *2011 Conference Record of the Forty Fifth Asilomar*

Conference on Signals, Systems and Computers (ASILOMAR). IEEE, 2011, pp. 439–443.

[20] C. Chatzichristos, E. Kofidis, Y. Kopsinis, M. M. Moreno, and S. Theodoridis, "Higher-order block term decomposition for spatially folded fMRI data," in *International Conference on Latent Variable Analysis and Signal Separation.* Springer, 2017, pp. 3–15.

[21] C. Chatzichristos, E. Kofidis, and S. Theodoridis, "PARAFAC2 and its block term decomposition analog for blind fMRI source unmixing," in *2017 25th European Signal Processing Conference (EUSIPCO).* IEEE, 2017, pp. 2081–2085.

[22] K. H. Madsen, N. W. Churchill, and M. Mørup, "Quantifying functional connectivity in multi-subject fMRI data using component models," *Human brain mapping*, vol. 38, no. 2, pp. 882–899, 2017.

[23] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3-4, pp. 321–377, 1936.

[24] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural computation*, vol. 16, no. 12, pp. 2639–2664, 2004.

[25] M. S. Ibrahim and N. D. Sidiropoulos, "Cell-edge interferometry: Reliable detection of unknown cell-edge users via canonical correlation analysis," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).* IEEE, 2019, pp. 1–5.

[26] B. Vinograde, "Canonical positive definite matrices under internal linear transformations," *Proceedings of the American Mathematical Society*, vol. 1, no. 2, pp. 159–161, 1950.

[27] R. G. Steel *et al.*, "Minimum generalized variance for a set of linear functions," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 456–460, 1951.

[28] P. Horst, "Generalized canonical correlations and their applications to experimental data," *Journal of Clinical Psychology*, vol. 17, no. 4, pp. 331–347, 1961.

[29] ——, "Relations among m sets of measures," *Psychometrika*, vol. 26, no. 2, pp. 129–149, 1961.

[30] J. D. Carroll, "Generalization of canonical correlation analysis to three or more sets of variables," in *Proceedings of the 76th annual convention of the American Psychological Association*, vol. 3, 1968, pp. 227–228.

[31] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[32] N. A. Asendorf, "Informative data fusion: Beyond canonical correlation analysis," 2015.

[33] X. Fu, K. Huang, E. E. Papalexakis, H.-A. Song, P. P. Talukdar, N. D. Sidiropoulos, C. Faloutsos, and T. Mitchell, "Efficient and distributed algorithms for large-scale generalized canonical correlations analysis," in *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, 2016, pp. 871–876.

[34] X. Fu, K. Huang, M. Hong, N. D. Sidiropoulos, and A. M.-C. So, "Scalable and flexible multiview MAX-VAR canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4150–4165, 2017.

[35] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and M. Hong, "Structured SUMCOR multiview canonical correlation analysis for large-scale data," *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 306–319, 2018.

[36] O. Friman, J. Cedefamn, P. Lundberg, M. Borga, and H. Knutsson, "Detection of neural activity in functional MRI using canonical correlation analysis," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 45, no. 2, pp. 323–330, 2001.

[37] X. Zhuang, Z. Yang, K. R. Sreenivasan, V. R. Mishra, T. Curran, R. Nandy, and D. Cordes, "Multivariate group-level analysis for task fMRI data with canonical correlation analysis," *NeuroImage*, vol. 194, pp. 25–41, 2019.

[38] Z. Yang, X. Zhuang, K. Sreenivasan, V. Mishra, T. Curran, R. Byrd, R. Nandy, and D. Cordes, "3D spatially-adaptive canonical correlation analysis: Local and global methods," *NeuroImage*, vol. 169, pp. 240–255, 2018.

[39] X. Zhuang, Z. Yang, T. Curran, R. Byrd, R. Nandy, and D. Cordes, "A family of locally constrained CCA models for detecting activation patterns in fMRI," *NeuroImage*, vol. 149, pp. 63–84, 2017.

[40] Y.-O. Li, W. Wang, T. Adali, and V. D. Calhoun, "CCA for joint blind source separation of multiple datasets with application to group fMRI analysis," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 1837–1840.

[41] B. Afshin-Pour, G.-A. Hossein-Zadeh, S. C. Strother, and H. Soltanian-Zadeh, "Enhancing reproducibility of fMRI statistical maps using generalized canonical correlation analysis in NPAIRS framework," *NeuroImage*, vol. 60, no. 4, pp. 1970–1981, 2012.

[42] G. Zhou, A. Cichocki, Y. Zhang, and D. P. Mandic, "Group component analysis for multiblock data: Common and individual feature extraction," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 11, pp. 2426–2439, 2015.

[43] R. A. Poldrack, J. A. Mumford, and T. E. Nichols, *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.

[44] L. Pauling and C. D. Coryell, "The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin," *Proceedings of the National Academy of Sciences*, vol. 22, no. 4, pp. 210–216, 1936.

[45] S. Ogawa, T.-M. Lee, A. R. Kay, and D. W. Tank, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proceedings of the National Academy of Sciences*, vol. 87, no. 24, pp. 9868–9872, 1990.

[46] S. Ogawa, T.-M. Lee, A. S. Nayak, and P. Glynn, "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields," *Magnetic resonance in medicine*, vol. 14, no. 1, pp. 68–78, 1990.

[47] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger, "Linear systems analysis of functional magnetic resonance imaging in human v1," *Journal of Neuroscience*, vol. 16, no. 13, pp. 4207–4221, 1996.

[48] R. B. Buxton and L. R. Frank, "A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation," *Journal of Cerebral Blood Flow & Metabolism*, vol. 17, no. 1, pp. 64–72, 1997.

[49] A. L. Vazquez and D. C. Noll, "Nonlinear aspects of the bold response in functional mri," *NeuroImage*, vol. 7, no. 2, pp. 108–118, 1998.

[50] M. Behroozi, M. R. Daliri, and H. Boyaci, "Statistical analysis methods for the fmri data," *Basic and Clinical Neuroscience*, vol. 2, no. 4, pp. 67–74, 2011.

[51] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: a general linear approach," *Human brain mapping*, vol. 2, no. 4, pp. 189–210, 1994.

[52] E. Formisano, F. De Martino, and G. Valente, "Multivariate analysis of fmri time series: classification and regression of brain responses using machine learning," *Magnetic resonance imaging*, vol. 26, no. 7, pp. 921–934, 2008.

[53] R. B. Buxton, E. C. Wong, and L. R. Frank, "Dynamics of blood flow and oxygenation changes during brain activation: the balloon model," *Magnetic resonance in medicine*, vol. 39, no. 6, pp. 855–864, 1998.

[54] M. I. Sereno, A. Dale, J. Reppas, K. Kwong *et al.*, "Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging," *Science*, vol. 268, no. 5212, p. 889, 1995.

[55] B. Sen and K. K. Parhi, "Extraction of common task signals and spatial maps from group fmri using a parafac-based tensor decomposition technique," in *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*.   IEEE, 2017, pp. 1113–1117.

[56] G. Stewart and J. Sun, "Computer science and scientific computing. matrix perturbation theory," 1990.