Temporal-Spatial Modeling for fMRI Data

Ping Bai

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2007

Approved by

Advisor: Dr. Young Truong

Co-Advisor: Dr. Richard L. Smith

Reader: Dr. Haipeng Shen

Reader: Dr. Xuemei Huang

Reader: Dr. Chuanshu Ji

ABSTRACT

PING BAI: Temporal-Spatial Modeling for fMRI Data
(Under the direction of Dr. Young Truong)

By generating high quality "movies" of the brain in action, functional Magnetic Resonance Imaging (fMRI) helps us examine which parts of the human brains are activated by different task performances. Many techniques for fMRI analysis have been developed in the last decade. Independent component analysis (ICA) is an effective data-driven method to explore spatio-temporal features in fMRI data. It has been especially successful to recover brain-function-related signals from recorded mixtures of unrelated signals. Due to the high sensitivity of MR scanners, spikes are commonly observed in fMRI data sets and they deteriorate the analysis. No particular method exists yet to address this problem. In the first part of this work, we introduce a supervised singular value decomposition technique into the data reduction step of ICA. The proposed method improves the robustness of ICA against spikes and makes the computation more efficient by using the particular fMRI experiment designs to guide the fully data-driven ICA. The advantages are demonstrated using a simulation study as well as a real data analysis.

ICA aims to separate blind source signals from their linear mixture signals based on the assumptions of the statistical independence and non-Gaussian distributions of the source signals. The second part of this work studies the methodology of some most popular ICA algorithms and propose to evaluate some of the algorithms by assessing the variability of the estimates of the mixing matrix through a nonparametric bootstrap procedure. Two maximum likelihood ICA algorithms are studied in detail through a simulation study.

Another popular category of statistical techniques for fMRI analysis consists of model-

driven strategies. Among them, the most widely used approach is statistical parametric mapping (SPM), where the key technique is general linear model (GLM) and the temporal characteristic of the expected response is usually modeled by the convolution of the experiment stimulus and a predefined hemodynamic response function (HRF). However, the subjective assumptions of the form of HRF introduce estimation biases and subsequently reduce the detection power of activation. In the third part of this work, we propose a new nonparametric method to model the time component adaptively in the context of SPM. The idea is to start from an initial time component obtained from general SPM procedure and then apply a penalized smoothing technique to update the shape of the hemodynamic response in an adaptive way. The nice performance of the proposed method is illustrated through a simulation study as well as a real fMRI data analysis.

Event-related fMRI (ER-fMRI) has played an important role in many recent brain imaging studies to explore the relationship between recorded fMRI signals and neural activity. Different from traditional block-design fMRI, ER-fMRI is very good at estimating the timing and waveform of the hemodynamic response. Various methods have been proposed in the literature to model the HRF. However, most of them have a number of limitations. In the last part of this work, we propose a novel regression approach to estimate the HRF directly. The approach is based on point processes modeling to account for the event-related designs. Compared to the existing methods, the proposed procedure yields simultaneously the nonparametric estimate of the HRF and a test for the linearity assumption. To illustrate its usefulness and the scientific implications, we applied this procedure to study the spatial variation of the HRF, and the extent to which the linear relationship holds in various regions of interest for Parkinson's disease patients.

ACKNOWLEDGMENTS

The writing of this dissertation has been one of the most significant academic challenges I have ever had. I owe my deepest gratitude to all those people who have made this dissertation possible and because of whom my graduate study in the past five years has been a most enjoyable, rewarding and unforgettable journey in my life.

First of all, I would like to gratefully and sincerely thank my advisor, Dr. Young Truong, for his guidance, patience and most importantly, his friendship during my graduate study at UNC. I have always been feeling fortunate to have an advisor who not only guided me in the wonderful research world, but also taught me the philosophy of life in the real world. I deeply appreciate his confidence and belief in me. Through out my graduate work, he continually stimulated my analytical thinking and greatly assisted me with scientific writing.

A special thanks goes to Dr. Haipeng Shen, who has been a friend and mentor. He taught me how to write academic papers and brought out the good ideas in me. He was always there to meet and talk about my ideas, to proofread and mark up my papers, and to ask me good questions to help me think through my problems.

I am grateful to Dr. Xuemei Huang for her encouragement, patience and all the thought-provocative discussion. I am also thankful to her for allowing me to use the computing facilities in her lab, reading my papers, commenting on my views and helping me better understand my research from the point of view of a neurologist. I am also indebted to the members of Dr. Huang's fMRI lab with whom I have worked during the course of my graduate study. In particular, I would like to acknowledge Dr. Mechelle

Lewis and Dr. Suman Sen for the many valuable discussions that helped me understand my research area better. My sincere thanks to Cara Slagle, Andrew Smith and Roxanne Poole for their patience and all the data-related assistance whenever I needed.

I would like to acknowledge Dr. Richard Smith, who reviewed my work and gave insightful comments, and Dr. Chuanshu Ji, who has given his time to read this manuscript and gave me guidance in the early stages of my graduate study.

Many friends have helped me a lot through these years. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value and appreciate their friendship.

Finally, and most importantly, I would like to express my heart-felt gratitude to my family, who has been a constant source of love, concern, support and strength all these years. I thank my parents for educating me to be what I am, for unconditional support and belief in me no matter what. Also, I thank my two sisters, the greatest in the world, for sharing their experience of the graduate study endeavor with me, for listening to my complaints and frustrations, and for always believing in me.

CONTENTS

LIST OF FIGURES

ABBREVIATIONS

A^T transpose of A

BOLD   Blood-Oxygenation-Level-Dependent

ER-fMRI   Event-related fMRI

fMRI   functional Magnetic Resonance Imaging

GLM   General Linear Model

HRF   Hemodynamic Response Function

ICA   Independent Component Analysis

SPM   Statistical Parametric Map

SSVD   Supervised Singular Value Decomposition

SVD   Singular Value Decomposition

CHAPTER 1

Introduction

Functional Magnetic Resonance Imaging (fMRI) is a set of noninvasive techniques for functional brain mapping. By generating high quality "movies" of the brain in action, it helps neuroscientists to study brain functions in vivo (Jezzard *et al.*, 2001; Huettel *et al.*, 2004). Since early 1990s, it has gained growing popularity in both clinical and basic neuroscience researches, and has influenced our understanding of the neurobiology of human behavior.

The most popular fMRI technique makes use of *blood-oxygenation-level-dependent* (BOLD) contrast, which is based on the differing magnetic properties of oxygenated (diamagnetic) and deoxygenated (paramagnetic) blood. Simply speaking, increases in local brain activity increase the local levels of blood oxygen. This in turn causes the measured fMRI signal to increase. In a typical fMRI experiment, functional images are recorded every few seconds while the subject is performing a specific task sequence or receiving a series of stimuli. Because the images are taken using an magnetic resonance (MR) sequence which is sensitive to changes in local blood oxygenation level, parts of the images taken during a certain activation or stimulation would show different intensity. And the parts of the images which show different intensity should correspond to the parts of the brain which are activated by the certain stimulation. Through the BOLD mechanism we can use fMRI data sets to answer two basic questions: (1) which brain regions are responsible for certain cognitive functions of interest? (2) how do these brain regions respond to the interesting cognitive functions? That is, what's the pattern of the

blood changes over time within an active brain region?

Since all measurements in the natural world are subject to random errors and an image is a measurement, the images are subject to random errors too. This makes it natural to involve fMRI studies with statistical analysis, which is concerned with making inference about underlying features in data that contain a large amount of random errors. Nowadays, statistics is playing a more and more active role in brain imaging science. To help improve the overall quality of the design and analysis of fMRI experiments, statistical techniques are required at almost all steps of fMRI analysis.

To date, a large variety of statistical procedures has been proposed for the analysis of BOLD fMRI data. These methods include the most common univariate analysis and multivariate methods. There are also parametric and nonparametric models considered. From the point of view of time series analysis, there are time-domain and spectral-domain methods. Although it has been demonstrated that some statistical procedures outperform others in certain contexts, there does not exist a single, globally optimal statistical method for the analysis of any particular fMRI study (Lange *et al.*, 1999).

In this work, we focus on the study and improvement of three statistical techniques that are most dominantly used in fMRI study. The first part of this work answers the first question mentioned earlier, that is, which brain regions are activated by certain stimuli. The key technique we use is independent component analysis (ICA) and we introduce a supervised singular value decomposition (SSVD) method into the ICA procedure to improve the robustness of ICA for fMRI data analysis against spikes which are common in fMRI data.

In the second part, we study some popular ICA algorithms in the literature and propose a bootstrap procedure to assess the variability of ICA estimates, which has not been done in the literature. Using this proposed procedure, we can compare different ICA algorithms as regard to reducing the variance of estimates.

The third part of this work refines a popular model-driven method for fMRI analysis,

statistical parametric mapping (SPM), to improve the estimation accuracy of the temporal characteristic of the expected responses, and subsequently, the detection power of activation.

The last part of this work handles the estimation of hemodynamic response by modeling the hemodynamic response function (HRF). We propose a novel approach to detect nonperiodic activations, estimate the HRF, and test the linearity assumption at the same time.

The rest of this report is organized as follows. Chapter 2 is an overview of the fMRI process, including the basic background of fMRI technique, the description of the fMRI data structure and the role that statistics play in this particular field. More specifically, detailed review of several most popular strategies developed for the analysis of BOLD fMRI data is also provided.

In Chapter 3, we introduce our robust ICA procedure, which is developed to deal with the spikes in fMRI data sets. Due to the high sensitivity of the MR scanner, spikes are inevitable in acquiring the fMRI data, while they cause misleading effects for the analysis. Currently in the literature, no particular methods are available for this issue. Our method is proven to be powerful and advantageous for handling this spike-situation.

In Chapter 4, we present the bootstrap procedure for assessing the variability of ICA estimates. We study two popular ICA algorithms and compare them through a simulation study, using the proposed bootstrap procedure.

Chapter 5 describes the proposed adaptive SPM method to estimate the temporal characteristic of the expected responses nonparametrically.

Chapter 6 introduces our novel approach for event-related fMRI study to model the HRF nonparametrically and test the linearity assumption at the same time.

More thoughts about the future work in fMRI studies are discussed in Chapter 7 to close this report.

CHAPTER 2

Overview of fMRI Analysis

2.1 Background

Back in 1890, the physiologist Arthur Sherrington originally demonstrated that local neuronal activity is related to changes in brain metabolism and blood supply (Lange, 1996; Jezzard *et al.*, 2001). This idea provides the basis for today's *blood-oxygenation-level-dependent* (BOLD) fMRI technique.

BOLD fMRI records signal contrast arising from the changes in magnetic susceptibility of oxygenated and deoxygenated blood. The magnetic field applied in an fMRI scanner is distorted to a different extent when it interacts with a different material. Since oxygenated blood is diamagnetic and deoxygenated blood is paramagnetic, the measured fMRI signal would increase when increases in local brain activity increase the local levels of blood oxygen. This mechanism is how we can use the data collected from fMRI experiments to localize specific areas of the brain that are activated by cognitive functions of interest. For more detailed physics and physiology of BOLD fMRI, please refer to two excellent introduction books Jezzard *et al.* (2001) and Huettel *et al.* (2004).

The fMRI techniques began to grow rapidly since the early 1990s, due to the increased prevalence of MRI scanners and other related techniques. The range of applications of fMRI to neuroscience is expanding rapidly as well. Because of its noninvasive nature, the widespread availability of the technology and the high spatial and temporal resolutions provided by its high quality images, BOLD fMRI has been dominant over other functional

brain mapping techniques.

2.2   General fMRI Process

An fMRI study usually starts with a question or hypothesis brought up by neuroscientists. According to the question or hypothesis, proper fMRI experiments are designed and implemented. Various related analyses are then carried out on the MR images recorded during the experiments, to answer the question or test the hypothesis.

2.2.1   fMRI Experiment

There are some different strategies for fMRI experimental design available. The earliest and most straightforward approach for comparing brain response to different tasks during the imaging experiment is the "block design". According to the practical goals of the experiment, there could be one or more tasks involved and hence two or more states included in the blocks. Each of these tasks or states lasts a certain continuous time period and is performed in a certain order. Figure 2.1 shows several typical block designs.
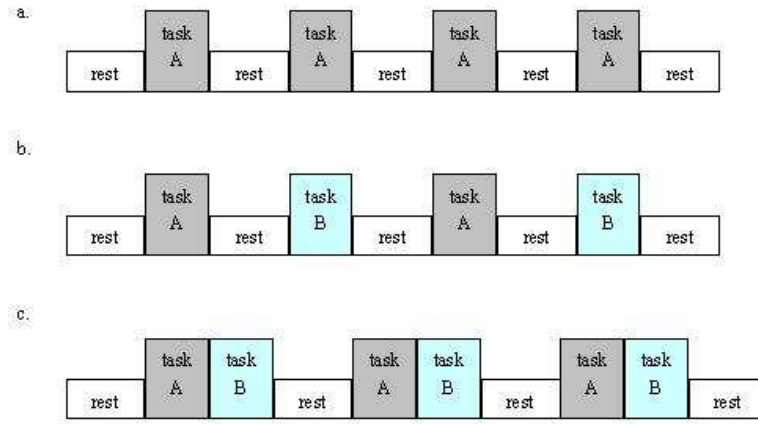
Figure 2.1: *Examples of three different block designs*

Another major method for fMRI experimental design, event-related design, has become more and more popular recently. Different from traditional block designs, in event-

related designs, the stimuli are applied for short bursts in a stochastic manner. Figure 2.2.1 gives a typical example for event-related design.

Figure 2.2: *A typical event-related design. A, B,... stand for different tasks*

When evaluating the strengths and weaknesses of an fMRI experiment, two factors are usually considered. The first is the detection power, that is, determining which brain regions are activated by the experiment stimuli. Another factor is the estimation power, which measures the pattern of blood changes over time within an active brain region (Huettel *et al.*, 2004). Both of the above experimental design strategies have their own advantages and disadvantages as regard to these two factors.

In addition to its simple analysis, block design is very good for detecting significant fMRI activity. But because the experimental conditions are extended in time, block design is relatively insensitive to the shape of the hemodynamic response, and hence poor at estimating the time course of blood changes in activated brain regions. On the contrary, event-related design turns out to be very good at estimating the shape of the hemodynamic response, while poor at the detection power.

Mixed designs that aim to combine both the block and event-related methods are carried out too. They can best combine the detection and estimation power. But the analysis of this kind of design is the most complicated too.

2.2.2 fMRI Data

During the fMRI experiment, a subject will lie in the magnet and perform a predefined task sequence, while a certain number of MR images of the subject's brain are typically recorded (In practice, an MR image can be also called as a volume or a scan). A single MR

image consists of a certain number of slices and each slice is made up of individual cuboid elements called voxels. Hence an fMRI data set can be considered as a three dimensional matrix of voxels that is repeatedly sampled over time. Statistically speaking, an fMRI data set is four dimensional and is usually represented as a spatio-temporal matrix $\mathbf{X}$ of dimension $M \times N$: each column of $\mathbf{X}$ corresponds to an fMRI image with M voxels, and each row of $\mathbf{X}$ is a time series of N time points for one voxel. This 4D fMRI data set can then either be thought of as N volumes, one taken every few seconds, or as M voxels, each with an associated time series of N time points. This is known as the complicated spatio-temporal nature of the fMRI signals. In most fMRI experiments, the number of time points is far less than the number of voxels ($N \ll M$).

Different analysis methods put emphasis on different aspects of fMRI signals. Some methods think of the fMRI data in the spatial representation, while others think of them in the temporal representation.

In Figure 2.3, we give two examples of the time series for two different voxels recorded in an fMRI experiment. On each row, the left image shows a slice of the brain image with an example voxel indicated by the crossings of two lines. The right plot shows the time series related to the highlighted voxel.

2.2.3 Data Preprocessing

The 4D data set acquired by the MR scanner should go through a series of preprocessing steps before it's ready for any statistical analysis. These preprocessing steps take the raw data as input, convert them into images that actually look like brains, then reduce unwanted noise of various types, and precondition the data in order to aid the later statistical analysis.

Usually the preprocessing consists of the following steps:

- Data reconstruction: The raw data are reconstructed into real space so that the image may be viewed and analyzed.
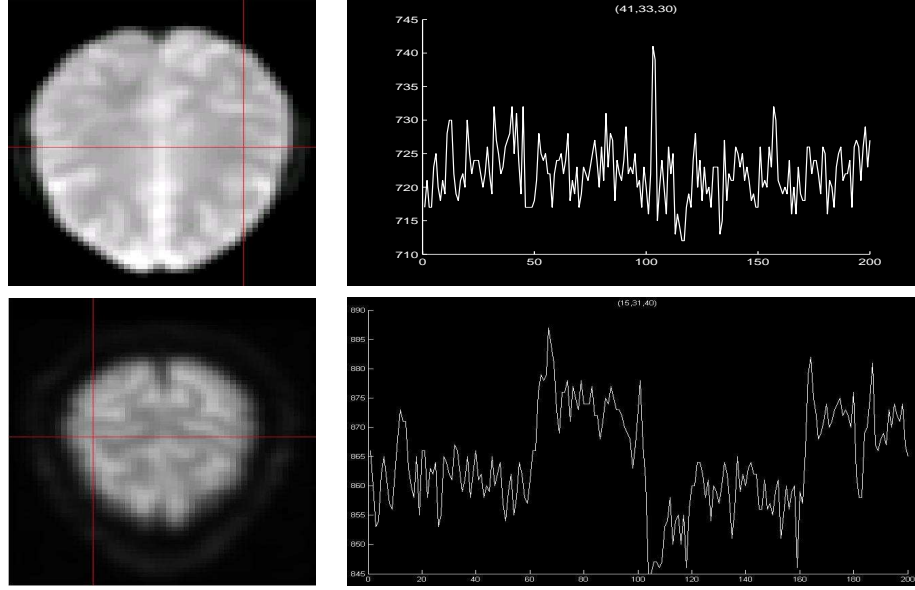
Figure 2.3: *Examples of two time series corresponding to two different voxels recorded in an fMRI experiment. On each row, the left image shows a slice of the brain image with an example voxel indicated by the crossings of two lines. The right plot shows the time series related to the highlighted voxel.*

- Slice-timing correction: Because each slice in each volume is acquired at slightly different times, it's necessary to adjust the data so that it appears that all voxels within one volume had been acquired at exactly the same time.

- Head motion correction: When the head moves during an experiment, some of the images will be obtained with the brain in the wrong location. The head motion correction step is to adjust the time series of images so that the brain is in the same position in every image. To accomplish this, each volume is transformed so that the image of the brain within each volume is aligned with that in every other volume. This step can be also called as a spatial normalization of the data.

- Intensity normalization: Each volume's overall intensity level is adjusted so that all volumes have the same mean intensity. This intensity normalization can help reduce the effect of global changes in intensity over time.

- Spatial filtering: Each volume is spatially blurred to increase signal-to-noise ratio

8

in the data.

- Temporal filtering: Each voxel's time series is filtered by linear or non-linear tools in order to reduce the low and high frequency noise.

The basic goal of these preprocessing steps is to reduce unwanted variability in the experimental data and to improve the validity of later statistical analyses (Huettel *et al.*, 2004). Without the preprocessing procedures, the statistical analysis would be greatly reduced in power and even rendered invalid.

As important the data preprocessing is, it's not the focus of our work. All the data are preprocessed in the software package SPM5 (http://www.fil.ion.ucl.ac.uk/spm/) before any statistical analysis in our work.

2.2.4 Statistical Analysis

After going through the sequence of preprocessing steps, the fMRI data is now ready for the final statistical analysis. One main feature of fMRI data is that the signal changes are small with the presence of lots of noise. This makes statistical analysis, which is concerned with making inference about underlying patterns in data that often contain a large amount of random error, necessary in fMRI analysis.

Over the last decade, a variety of statistical procedures has been proposed for the analysis of BOLD fMRI data. One of the earliest and most direct ways is to simply cross-correlate the voxel time series with a reference time course describing the sequence of stimulant events in the experiment (Bandettini *et al.*, 1993). The general linear model (GLM) is another commonly employed procedure (Friston *et al.*, 1994, 1995b). Lange *et al.* (1999) compared nine analytic methods currently used in BOLD fMRI analysis. Although it has been demonstrated that some statistical procedures outperform others in certain contexts, there is no globally optimal statistical procedure for the analysis of any particular fMRI study. More detailed review of the major statistical analysis methods is given in the following section.

2.3 Common Statistical Analysis Techniques Review

Currently most popular statistical techniques for fMRI analysis can be differentiated into two complementary categories, model-driven methods and data-driven methods. These two kinds of strategies are based on different assumptions of fMRI data and have their own advantages and disadvantages respectively.

2.3.1 Model-Driven Methods

The most widely used model-driven strategy for the analysis of fMRI data is statistical parametric mapping (SPM), which is carried out using a two-stage approach. In the first stage, a general linear model (GLM) with correlated errors is used for each voxel time series. That is,

$$Y_j = \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \qquad \boldsymbol{\epsilon}_j \sim N(0, \sigma^2 \boldsymbol{\Sigma}), \qquad j = 1, 2, \ldots, M, \tag{2.1}$$

where M is the number of voxels of the brain. Y_j is the time series of the jth voxel. The design matrix $\mathbf{X}$ contains terms that model the BOLD response to the stimuli and non-linear trends that are often observed in fMRI voxel time series.

Once this model has been fitted at each voxel, inferences of the model parameters are then made according to the experiment hypothesis. The resulting statistics from all the voxels are assembled spatially into an image, which is the so-called statistical parametric map. The second stage then focuses on the analysis of the statistical map in order to identify those areas of the brain that are activated by the stimuli.

The generality of SPM comes mainly from the flexible forms of the design matrix $\mathbf{X}$ and the variety of the statistics that can be calculated from the model. According to the experimental design, the design matrix can contain both continuous covariates and factorial indicators, which represent different effects of conditions in the experiment. The nature of the BOLD response implies that in areas of activation there is a delayed

and blurred version of the stimulus design. Hence part of the design matrix $\mathbf{X}$, which represents the temporal characteristic of the expected responses, is commonly modeled through the convolution of the stimulus design $s(\cdot)$ with a *hemodynamic response function* (HRF), $h(\cdot)$. That is,

$$\text{BOLD}(t) = \sum_u h(t-u)s(u).$$

Commonly suggested forms for the HRF include discretized Poisson, Gamma and Gaussian density functions. Most model-driven approaches for fMRI analysis can be categorized as SPM, with different assumptions of the HRF, various forms of the general linear model and different approaches for parameter estimation.

Friston *et al.* (1994) gives a typical example of SPM analysis. This work identifies the activated voxels by producing a statistical map with the voxel value being the correlation coefficient between the observed fMRI signal and the input stimulus function. To consider the effects of delay and dispersion of the hemodynamic response, the input stimulus is first convolved with an HRF. The HRF is assumed to have a Poisson distribution, with the mean being estimated using intrinsic autocorrelations in the observed fMRI signals. Based on the estimation of the spectral densities of the observed fMRI signals and the corrected (convolved with HRF) input stimulant signals in the spectral domain at each voxel, this work calculates a statistic (ζ) which has a standard Gaussian distribution under the null hypothesis (no effect in the voxel). Activated regions are then detected by properly thresholding the resulted statistical parametric map with ζ values.

To allow for spatial variation in the HRF at each voxel, a small set of basis functions can be used to model the HRF (Friston *et al.*, 1995a; Josephs *et al.*, 1997). Each basis function is convolved with the design to make one column of the design matrix. Inferences are then made on the linear model coefficients to form the statistical map.

In the paper by Lange and Zeger (1997), a non-linear parametric model for detecting activations in fMRI data is presented. The model at each voxel is a special case of Equation (2.1), with each column of the design matrix $\mathbf{X}$ being the convolution of a

stimulus sequence and an estimated HRF. The coefficient $\boldsymbol{\beta}_j$ shows the magnitude of the linear dependence of the observed fMRI signal at the jth voxel on the corrected stimuli. Here the HRF is modeled as a two-parameter Gamma distribution whose Fourier transform can be evaluated analytically. The two parameters of the HRF, which vary at different voxels, are estimated simultaneously when estimating $\boldsymbol{\beta}_j$. The analysis in this work is carried out in frequency domain too and the "non-linearity" resides in its iterative way for estimating the parameters. Residual images are also obtained in this work. But different from traditional SPM, where random field theory is used for global thresholding to detect the activation regions, this work uses focused tests of activation. Namely, the detection is focused on the potential activation areas given by neuroscientists. In each of such region of interest (ROI), residual spatial autocorrelation functions (ACF) are modeled by exponential and Gaussian forms. This modeling incorporates both the spatial and temporal features of the fMRI data and gives the estimation of the variance-covariance matrix $\boldsymbol{W}$ of the estimated $\hat{\boldsymbol{\beta}}$. In each ROI, under the null hypothesis (no any effect in that region) the statistic $\hat{\boldsymbol{\beta}}'\boldsymbol{W}^{-1}\hat{\boldsymbol{\beta}}$ follows a χ^2-distribution with degrees of freedom being the number of voxels in that ROI.

Bayesian analysis of fMRI data was first fully implemented in Genovese (2000). Instead of modeling the hemodynamic response by convolving the stimulus with a HRF as in most studies, this work models the hemodynamic response directly in a Bayesian framework. Four components are included in the voxelwise model, the baseline signal, drift profile, activation profile and noise. Priors reflecting true prior knowledge are chosen for the parameters involved in these four components. Spatial dependencies are included in the noise model. One advantage of this method is that it produces estimates of meaningful parameters instead of test statistics as in conventional SPM approaches, and hence can answer questions beyond localization problem, e.g. monotonicity is considered as an example in this work. MCMC algorithms are used for inferences, which also causes an intensive computation problem for this method.

Gössl *et al.* (2001b) also uses Bayesian analysis in fMRI and models the hemodynamic response directly as in the above work. However this work adopts the regression model used in Friston *et al.* (1994) instead and defines five different parameters for the signal's baseline, increase, plateau, decrease and undershoot. Prior distributions or numerical values are then given to these parameters. Also based on a Bayesian framework, Gössl *et al.* (2001a) proposed an approach for considering the temporal and spatial dependencies between voxels, which is different from most SPM methods where the spatial correlations between voxels are usually accounted for in the second step. Instead, the spatio-temporal correlations are incorporated in the model formulation through spatial Markov random field priors.

Marchini and Ripley (2000) proposed a nonparametric approach in spectral domain for fMRI analysis. It is not a model-driven approach, but it can be viewed as a special case in the SPM framework. This method is based on the fact that in periodic experimental design, the fundamental frequency of activation contains the majority of information regarding the observed response. Nonparametric techniques are used in all stages of the analysis, including trend removal and correlation structure estimation. At each voxel, the periodogram of the time series for the voxel and a smoothed version of this periodogram are calculated to obtain the so-called ratio statistic. Inferences are then made based on the ratio statistics for all the voxels. A small amount of spatial smoothing to the estimates is applied to consider the spatial correlations. This nonparametric method takes least assumptions about the original data compared with most parametric methods and is more resistant to high frequency artefacts. The idea of this approach is illustrated by periodic experiment examples, while it can be extended for event-related fMRI analysis as well.

Some limitations of the above model-driven methods do exist. First of all, two assumptions are required for GLM: normal distribution of the observed data and the independence of the error terms. Secondly, the validity of modeling the hemodynamic

response effect by the convolution model is yet to be verified. In addition, the forms of the HRF commonly used in the literature are all predefined up to several parameters. These subjective assumptions of the HRF may result in invalid estimates as well.

2.3.2 Data-Driven Methods

On the other hand, for exploratory fMRI analysis, data-driven approaches have been found to be informative. These include clustering analysis (Goutte *et al.*, 1999), principal component analysis (PCA) (Kherif *et al.*, 2002) and independent component analysis (ICA) (McKeown *et al.*, 1998b; Petersen *et al.*, 2000; Calhoun *et al.*, 2003). Among these, ICA is so far the most popular data-driven approach for block-design fMRI studies.

Different from model-driven methods, the data-driven methods are multivariate techniques that account for the dependence among different voxels. In the following chapter, we focus on the study of ICA, with more detail about the theory of ICA and its advantages and disadvantages.

More recently, SPM-ICA is proposed by Hu *et al.* (2005) to unify the model-driven approach SPM with the data-driven approach ICA. In this work, temporal ICA (tICA) was applied to fMRI data sets first to disclose independent components. The resulting components are used to construct the design matrix of a GLM as in Equation (2.1). A conventional SPM analysis is then carried out based on this GLM.

CHAPTER 3

Robust Independent Component Analysis

3.1 Introduction

ICA is a data-driven technique that aims to separate blind sources from their linear mixtures based on the assumption of the statistical independence of the source signals. McKeown *et al.* (1998b) originally introduced ICA to fMRI data analysis and this method has attracted lots of interest in this field ever since. The literature body of the application of ICA in fMRI studies has been growing rapidly.

As an exploratory data analysis tool, ICA is capable of extracting from the recorded fMRI signals the individual signals that correspond to multiple sources such as experiment-stimulus-related components, cardiac and respiratory effects and subject/machine movements. However, most ICA algorithms are sensitive to outliers (McKeown *et al.*, 1998b). One of our contributions is to propose a technique that makes ICA more robust towards outliers.

Outliers usually appear as spikes in fMRI data sets (Kao and MacFall, 2000). The spikes refer to data points with relatively high signal magnitudes, and they are inevitable in fMRI data sets due to radio-frequency problems in MR scanners, static discharge caused by synthetic fibers or even abrupt subject movements. For illustration purposes, Figure 3.1 plots the recorded time series corresponding to three different voxels in the fMRI data set that we analyze later in Section 3.5. The big spikes around time points 110, 100 and 120 in the time series are examples of many other spikes that are commonly
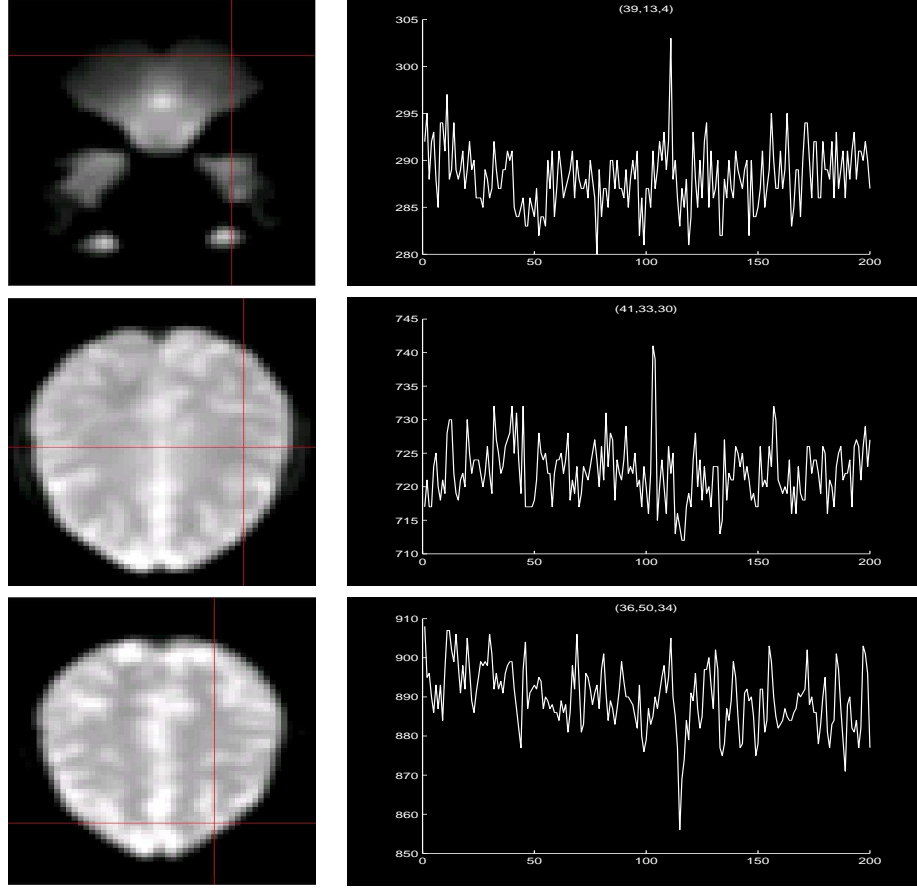
Figure 3.1: *Examples of spikes in our fMRI data set. The left column contains the images for three slices of the brain. The right column plots the time series corresponding to the voxels highlighted by the line crossings on the slices. The big spikes around time points* 110*,* 100 *and* 120 *are examples of many other spikes in the data.*

detected in fMRI data. The spatial locations of the voxels are respectively indicated by the line crossings on the images in the left column. What is more interesting is that the highlighted voxel in the first row even lies outside of the brain tissue; hence, the corresponding spike clearly doesn't represent any brain function. To the best of our knowledge, there has not been much research explicitly addressing statistical issues with the spikes. Luo and Nichols (2003) presented an exploratory diagnostic tool to identify outliers in fMRI data sets.

ICA of fMRI usually involves various preprocessing stages (Calhoun *et al.*, 2003).

A data reduction is usually performed through dimension reduction techniques such as singular value decomposition (SVD) (Petersen *et al.*, 2000). As mentioned earlier, fMRI data sets are usually of high dimension and with far less time points than voxels. Computationally, it is too expensive to apply ICA algorithms directly onto such matrices. To make ICA more efficient, the data reduction aims at reducing the high-dimensional full space to a much smaller feature subspace retained by SVD. Then the ICA decomposition can be focused on the feature subspace. Section 3.2.2 gives more detail about this data reduction step.

As a least squares based method, the SVD used for data reduction, however, is known to be highly susceptible to outliers. Hence the analysis from the follow-up ICA could be contaminated by those spikes. In this work, we propose a supervised SVD (SSVD) procedure that is less sensitive to the spikes in fMRI data sets. Our proposal is motivated by the observation that SVD can be interpreted as a low rank matrix approximation technique. The SVD components can then be obtained from solving a sequence of minimization problems. We introduce some regularization through basis expansion in the corresponding optimization problems to achieve supervised low rank approximations. The basis expansion is constructed using the information of the fMRI experiment designs, particularly the frequencies of the stimulus sequences. Such supervision focuses the SVD on the experimentally interesting directions, which makes the decomposition more efficient and meanwhile less sensitive to spikes as illustrated in Sections 3.4 and 3.5.

The current research is motivated by one of the few fMRI studies focused on elucidating the neurocircuitry involved in Parkinson's disease (PD) and its motor dysfunction. PD patients can have trouble performing simple motor activities, such as finger tapping. One research goal is to identify the brain regions (or Regions of Interest (ROIs)) that are associated with finger tapping in these subjects. To detect all the ROIs associated with performing motor activities robustly, this study employs a block design involving alternating right/left-hand finger tapping (Figure 3.5). Our SSVD-ICA method makes

use of this design information.

The rest of this chapter is organized as follows. Section 3.2 provides an overview of the ICA methodology as well as the conventional data reduction step before ICA. In Section 3.3, we introduce the supervised SVD procedure and its application in ICA for fMRI. Section 3.4 illustrates the performance of our proposed method using a simulation study. In Section 3.5, we compare our method with an existing ICA package using one fMRI data set collected in the aforementioned study. Concluding remarks are given in Section 3.6 to close the chapter.

3.2 Independent Component Analysis

3.2.1 Overview of ICA for fMRI

ICA has been recently used in fMRI studies to extract independent source signals from the recorded fMRI signals (McKeown *et al.*, 1998b; Petersen *et al.*, 2000; Calhoun *et al.*, 2003). The basic idea of ICA can be illustrated using the classic "cocktail party" problem (Hyvärinen *et al.*, 2001; Stone, 2004). Suppose many people talk simultaneously at a party, and several microphones are recording in different locations. The recorded signals are then mixtures of different voices. Using only these recorded mixtures as inputs, ICA aims at identifying the individual voices of different people.

Mathematically, let $\mathbf{x}$ be an M-dimensional vector variable, whose elements are signal mixtures recorded at one time point, and $\mathbf{s} = (s_1, \ldots, s_K)^T$ be a K-dimensional vector variable with each element being a source signal at the time point. The typical ICA model is written as

$$\mathbf{x} = \mathbf{As}, \tag{3.1}$$

where $\mathbf{A}$ is an $M \times K$ mixing matrix, M is the number of signal mixtures and K is the number of source signals.

The independent source components (ICs) $s_1, \ldots, s_K$ are taken as latent variables in this model. Both the unknown mixing matrix $\mathbf{A}$ and the source signals $\mathbf{s}$ need to be

estimated only using the observed signal mixture $\mathbf{x}$. For estimation purpose, the source signals $s_1, \ldots, s_K$ are assumed to be statistically independent and have non-Gaussian distributions. Hyvärinen *et al.* (2001) prove that, without the non-Gaussianity assumption, the mixing matrix $\mathbf{A}$ is not identifiable at all.

According to the Central Limit Theorem, under certain assumptions, the distribution of the sum of several independent random variables is more Gaussian than any of the original random variables. Making use of this fact, ICA recovers the independent components by finding an unmixing matrix $\mathbf{W}$ to maximize the non-Gaussianity of $\mathbf{Wx}$. Then the mixing matrix $\mathbf{A}$ is estimated as $\mathbf{W}^{-1}$, and the source signals are recovered as $\mathbf{s} = \mathbf{Wx}$. For more detail about the theory of ICA, see Hyvärinen *et al.* (2001) and Stone (2004).

As discussed earlier, an fMRI data set is usually represented as a spatio-temporal matrix $\mathbf{X}$ of dimension $M \times N$: each column of $\mathbf{X}$ contains an fMRI image with M voxels recorded at one time point, and each row of $\mathbf{X}$ consists of a time series of N time points for one voxel. Adopting Equation (3.1) into the context of fMRI, we can write the ICA decomposition model for $\mathbf{X}$ as:

$$\mathbf{X} = \mathbf{AS}, \tag{3.2}$$

where each column of the $M \times K$ matrix $\mathbf{A}$ holds a spatial component map, each row of the $K \times N$ matrix $\mathbf{S}$ is the corresponding time series, and K is the number of underlying ICs or source signals.

Due to the spatio-temporal nature of fMRI signals, there are two distinct ICA decomposition options, spatial ICA (sICA) and temporal ICA (tICA). The sICA aims to find independent image components (the columns of $\mathbf{A}$), while tICA looks for independent time courses (the rows of $\mathbf{S}$). In either case, a single ICA component can be interpreted as one spatially distributed set of voxels (one column of $\mathbf{A}$) that is activated by the

corresponding time course in one row of **S** (McKeown *et al.*, 2003).

3.2.2 The Data Reduction Step before ICA

fMRI data are usually of high dimension, especially in the spatial domain. In addition, ICA algorithms are often computationally intensive. Hence before applying ICA, it is common to first perform dimension reduction using SVD. ICA algorithms are then applied in the reduced subspace.

Suppose rank($\mathbf{X}$) $= r \leq \min(M, N)$. The SVD decomposes $\mathbf{X}$ as follows,

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{3.3}$$

where $\mathbf{U}$ is the $M \times r$ matrix of *orthonormal* left singular vectors, $\mathbf{V}$ is the $N \times r$ matrix of *orthonormal* right singular vectors, and $\mathbf{D}$ is the $r \times r$ diagonal matrix of *positive* singular values. Here $\mathbf{U}$ and $\mathbf{V}$ can be viewed as the basis vectors that span the spatial patterns and temporal sequences respectively.

As aforementioned, the sICA of $\mathbf{X}$ looks for independent image components, and tICA for independent time series. Hence, when performing tICA, we can focus on the subspace spanned by $\mathbf{Y} \equiv \mathbf{D}\mathbf{V}^T$, which is of dimension $r \times N$, much smaller than the original $M \times N$. Applying an ICA algorithm on the reduced data $\mathbf{Y}$, we can get $\mathbf{Y} = \tilde{\mathbf{A}}\mathbf{S}$, where $\mathbf{S}$ contains the independent time components. The original spatial maps in the model (3.2) can then be reconstructed as $\mathbf{A} = \mathbf{U}\tilde{\mathbf{A}}$.

Similarly, sICA can be performed by focusing on the subspace retained by $\mathbf{Y}^* \equiv \mathbf{U}\mathbf{D}$. An ICA decomposition on $\mathbf{Y}^*$ results in $\mathbf{Y}^* = \mathbf{A}\tilde{\mathbf{S}}$, where $\mathbf{A}$ consists of the independent spatial maps. The original time courses in the model (3.2) are then recovered as $\mathbf{S} = \tilde{\mathbf{S}}\mathbf{V}^T$.

In summary, the idea of the data reduction step before ICA is to reduce the dimension of the matrices that will be used as inputs to ICA algorithms. Considering the fact that most fMRI experiments have far less time points than voxels, i.e. $r \leq N \ll M$, this preprocessing step can greatly improve the computational efficiency of the whole

procedure.

3.3 Supervised Singular Value Decomposition

Both ICA and SVD are sensitive to spikes that are frequently encountered in fMRI data. To overcome this analysis challenge, we propose to reduce the spike effect in the data-reduction step via a modified SVD technique that is supervised by the experiment design. Consequently, the follow-up ICA is more robust as it focuses on the subspace maintained by the data reduction.

3.3.1 Low Rank Approximation via SVD

To motivate our approach, we note that SVD can be viewed as a matrix low rank approximation method. In the SVD of $\mathbf{X}$ (3.3), let $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_r]$, $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_r]$ and $\mathbf{D} = \mathrm{diag}\{d_1, \ldots, d_r\}$. For an integer $l \leq r$, define

$$\mathbf{X}^{(l)} \equiv \sum_{k=1}^{l} d_k \mathbf{u}_k \mathbf{v}_k^T.$$

Then, $\mathbf{X}^{(l)}$ is the closest rank-l matrix approximation to $\mathbf{X}$ (Harville, 1997). Let $\mathbf{X}^*$ be an arbitrary rank-l matrix, the term "closest" simply means that $\mathbf{X}^{(l)}$ minimizes the squared Frobenius norm between $\mathbf{X}$ and $\mathbf{X}^*$:

$$\|\mathbf{X} - \mathbf{X}^*\|_F^2 = \mathrm{tr}\{(\mathbf{X} - \mathbf{X}^*)(\mathbf{X} - \mathbf{X}^*)^T\}.$$

Suppose, for example, we seek the best rank-one matrix approximation of $\mathbf{X}$. Note that any $M \times N$ rank-one matrix can be written as $\mathbf{u}\mathbf{v}^T$, where $\mathbf{u}$ is a norm-1 M-vector and $\mathbf{v}$ is a N-vector. The problem can be formulated as the following optimization problem,

$$\min_{\mathbf{u},\mathbf{v}} \|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2. \tag{3.4}$$
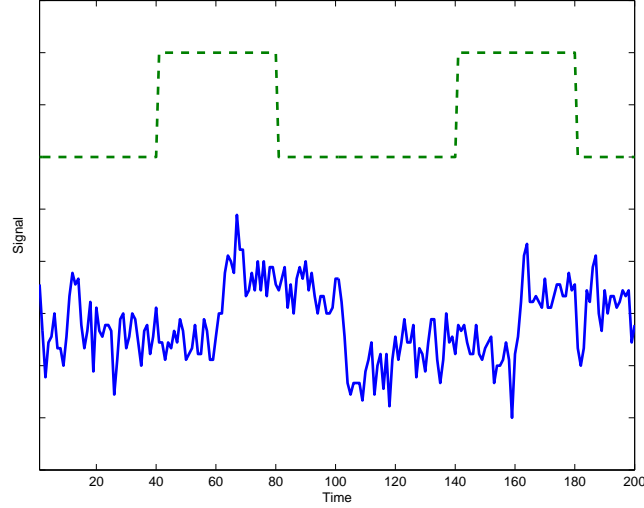
Figure 3.2: *The recorded time course (solid line) of a voxel that is activated by the experiment stimulus sequence (dashed line). The lower and higher levels of the dashed line stand for rest and activation periods of the experiment respectively.*

Then SVD's low rank approximation property implies the following solution

$$\mathbf{u} = \mathbf{u}_1, \quad \mathbf{v} = d_1 \mathbf{v}_1.$$

The subsequent pairs $(\mathbf{u}_k, d_k, \mathbf{v}_k)$, $k > 1$, provide best rank-one approximations of the corresponding residual matrices. For example, $d_2 \mathbf{u}_2 \mathbf{v}_2^T$ is the best rank-one approximation of $\mathbf{X} - d_1 \mathbf{u}_1 \mathbf{v}_1^T$.

3.3.2 Supervised SVD

In block design fMRI studies, experiment tasks or stimuli are typically applied in alternating blocks. In the areas activated by these stimuli, we would observe temporal data that are correlated with the experiment design. Figure 3.2 shows the time series (solid line) of a voxel that is activated by the experiment stimulus (dashed line) in the fMRI study reported later in Section 3.5. We note that the time components can usually be modeled as sinusoidal curves plus some noise in such experiments. This observation motivates us to propose the following supervised SVD (SSVD) procedure.

Suppose the time component of interest, $\mathbf{v} = (v(t_1), \ldots, v(t_N))^T$, can be modeled as

$$v(t_i) = a \sin(2\pi\omega t_i + \phi) = a \cos \phi \sin(2\pi\omega t_i) + a \sin \phi \cos(2\pi\omega t_i),$$

where a is the amplitude, ϕ is the phase, ω is the frequency, and t_i is the ith scanning time. Define

$$\mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2), \quad \boldsymbol{\psi} = (a \cos \phi, a \sin \phi)^T,$$

where

$$\mathbf{b}_1 = (\sin(2\pi\omega t_1), \ldots, \sin(2\pi\omega t_N))^T, \quad \mathbf{b}_2 = (\cos(2\pi\omega t_1), \ldots, \cos(2\pi\omega t_N))^T.$$

Then, one can see that

$$\mathbf{v} = \mathbf{B}\boldsymbol{\psi}, \tag{3.5}$$

which suggests that $\mathbf{v}$ is constrained to be in the linear space spanned by the bases $\mathbf{b}_1$ and $\mathbf{b}_2$.

To make use of the sinusoidal nature of $\mathbf{v}$, we propose to impose the basis-expansion constraint (3.5) on $\mathbf{v}$ in the optimization problem (3.4), and re-formulate the problem as follows,

$$\min_{\mathbf{u}, \boldsymbol{\psi}} \ \|\mathbf{X} - \mathbf{u}\mathbf{v}^T\|_F^2 \quad \text{subject to } \mathbf{v} = \mathbf{B}\boldsymbol{\psi}. \tag{3.6}$$

We name this formulation *Supervised SVD*, because it *supervises* SVD by restricting it to find the best low rank approximation within a certain subspace. On the other hand, the conventional SVD is *unsupervised*.

One nice property of the formulation (3.6) is that it automatically achieves scale invariant as indicated by

$$(c\mathbf{u})(\mathbf{v}/c)^T = \{\mathbf{u}c\}\{\mathbf{B}(\boldsymbol{\psi}/c)\}^T = (\mathbf{u})(\mathbf{B}\boldsymbol{\psi})^T = \mathbf{u}\mathbf{v}^T,$$

23

where c is a nonzero constant. For identifiability purpose, we can standardize $\mathbf{u}$ and $\mathbf{v}$ to unit length and introduce a slope parameter d. The problem can then be rewritten as

$$\min_{d,\mathbf{u},\boldsymbol{\psi}} \quad \|\mathbf{X} - d\mathbf{u}\mathbf{v}^T\|_F^2 \qquad \text{subject to } \mathbf{v} = \mathbf{B}\boldsymbol{\psi}, \mathbf{u}^T\mathbf{u} = 1, \mathbf{v}^T\mathbf{v} = 1. \qquad (3.7)$$

Currently the SSVD is illustrated with only the base Fourier bases in $\mathbf{B}$ for simplicity of the presentation. One can easily extend $\mathbf{B}$ to accommodate higher order Fourier bases, which makes the methodology more flexible to model periodic signals.

3.3.3 Solution and Practical Implementation

The solution of the Supervised SVD (3.7) can be obtained by solving a couple of generalized eigen-problems as stated below.

Theorem 1. *The triplet $\{d, \mathbf{u}, \boldsymbol{\psi}\}$ that minimizes (3.7) satisfies the following equations:*

$$
\begin{cases}
\max_{\mathbf{u}} \quad \mathbf{u}^T\mathbf{X}\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}^T\mathbf{u} \qquad \textit{subject to } \mathbf{u}^T\mathbf{u} = 1, \\[2mm]
\max_{\boldsymbol{\psi}} \quad \boldsymbol{\psi}^T\mathbf{B}^T\mathbf{X}^T\mathbf{X}\mathbf{B}\boldsymbol{\psi} \qquad \textit{subject to } \boldsymbol{\psi}^T\mathbf{B}^T\mathbf{B}\boldsymbol{\psi} = 1, \\[2mm]
d = \boldsymbol{\psi}^T\mathbf{B}^T\mathbf{X}^T\mathbf{u}.
\end{cases}
$$

The proof of the theorem is relegated to Appendix A. The generalized eigen-problems can be solved using standard methods as shown in the Appendix. Below we summarize the computational algorithm to obtain the first supervised time component as well as its corresponding spatial component. See the Appendix for the technical justification. The same algorithm can be applied repeatedly on the residual matrices until the desired number of components is obtained.

Algorithm 1. Supervised SVD (SSVD)

(1) Obtain the frequency of interest ω from the experimenter, or estimate it through spectrum analysis; (See the comments below.)

(2) Form the basis matrix $\mathbf{B}$ and apply Cholesky decomposition on $\mathbf{B}^T\mathbf{B}$ to get $\mathbf{B}^T\mathbf{B} = \mathbf{R}_\mathbf{B}^T\mathbf{R}_\mathbf{B}$, where $\mathbf{R}_\mathbf{B}$ is a 2×2 upper triangular matrix;

(3) Apply SVD on $\mathbf{XBR}_\mathbf{B}^{-1}$ to derive its first left singular vector $\mathbf{u}$ and the first right singular vector $\tilde{\psi}$;

(4) Set $\psi \equiv \mathbf{R}_\mathbf{B}^{-1}\tilde{\psi}$ and $d \equiv \psi^T\mathbf{B}^T\mathbf{X}^T\mathbf{u}$, which leads to $\mathbf{v} = \mathbf{B}\psi$. Hence we obtain the first SSVD triplet $\{d, \mathbf{u}, \mathbf{v}\}$.

Below we want to comment on the choice of the sinusoidal frequency ω. The above algorithm relies on knowing the sinusoidal frequency ω or being able to estimate it from the fMRI data. In most fMRI experiments, the frequency ω of the interesting component is known a priori, for example, the component corresponding to the experimental stimulus. The experimenter might also be interested in some underlying unknown signals, in which case we propose to estimate ω through spectrum analysis on the $\mathbf{v}$ components extracted from the conventional SVD. The effect of spikes is trimmed by the fact that they have a much smaller effect on the peak locations of the spectrums, or the dominating frequencies of the signals.

3.3.4 Application to ICA

Once we obtain the desired number of SSVD components, we propose to apply an ICA algorithm on them. As the SSVD components are insensitive to the spikes, the analysis results from the follow-up ICA procedure should be robust as well (Section 3.4). In addition, because our procedure (SSVD-ICA) makes use of the nature of the fMRI experiment design, it appears to be more powerful in detecting activated brain regions of interest (Section 3.5).

In terms of the number of SSVD components to extract, we propose the following approach. If we know all the signal frequencies of interest, we can just extract the corresponding components. Otherwise, we need to estimate the interesting frequencies. For such, we suggest to extract the first 30 (for example) SVD components, and perform spectrum analysis on them to identify possibly interesting frequencies, before applying our SSVD approach. This approach is consistent with the common practice in ICA for fMRI data, where around 20 or 40 ICA components are usually extracted, and interesting components are then chosen either through visual inspection or correlation analysis.

3.4 A Simulation Study

To illustrate the robustness of the modified ICA using SSVD, we compare it with the conventional ICA using SVD in the following simulation study.

3.4.1 Data Description

According to the ICA decomposition model (3.2), we simulated an $M \times N$ fMRI data matrix $\mathbf{X}$ by first simulating the $M \times K$ spatial component matrix $\mathbf{A}$ and the $K \times N$ time series matrix $\mathbf{S}$ separately. The data matrix $\mathbf{X}$ was then obtained as $\mathbf{X} = \mathbf{A}\mathbf{S}$.

In this study, we set $K = 5$, $M = 30 \times 30 \times 10$ and $N = 240$. The simulated data can be explained as follows: there are 5 underlying independent components; each column of the spatial component matrix $\mathbf{A}$ is a component map that consists of 10 slices and each slice contains 30×30 voxels; while each row of $\mathbf{S}$ is a time series of length 240 that corresponds to the relevant spatial component in $\mathbf{A}$. We simulated the data based on a simple rest-activation block design (the dotted line in the time plot within each panel of Figure 3.3). Each rest or activation period lasted 18 seconds.

Out of the five time components in the matrix $\mathbf{S}$, the first four were simulated based on simple sinusoidal functions plus randomly generated Uniform noise, which are plotted in Figure 3.3 along with their corresponding spectrum plots highlighting the frequencies. The first time component corresponds to the stimulus of some artificial experiment with

a frequency of 0.06Hz. The second and third time components are for the heart beat and breath with frequencies of 1Hz and 0.3Hz respectively. The fourth one is an artifact effect with a frequency of 0.7Hz. The last time component is pure noise. The amplitudes of the first four components are 0.5, 0.45, 0.35 and 0.45 respectively. The noise is sampled from a Uniform distribution between -0.1 and 0.1. Note that the noise distribution has to be non-Gaussian in order for ICA to be well defined as discussed in Section 2.

All the voxels in $\mathbf{A}$ were given a numerical value of either 0 or 1. In each spatial component, the voxels with value 1 correspond to the regions that were activated by the corresponding time stimulus, and they are plotted as dark red areas in Figure 3.3.

To simulate spikes in the data, we randomly selected 10% of the entries in the simulated $\mathbf{X}$ and replaced them with noise randomly generated from Uniform$[-10, -2]$ and Uniform$[2, 10]$. The noise distributions are chosen so that the simulated spikes are indeed outliers, judged by the usual 1.5-Inter-Quartile-Range rule of thumb.

3.4.2 Analysis and Results

Following the standard practice in ICA, we first normalized the contaminated data matrix $\mathbf{X}$ by column centering and row standardization (Hastie and Tibshirani, 2002). Both ICA and SSVD-ICA were applied to the normalized matrix. For computing, we employed the fastICA algorithm (Hyvärinen *et al.*, 2001) in both cases because of its fast computation and popularity .

To effectively display the activated voxels in the extracted spatial maps, the values in each map were standardized to z-scores (McKeown *et al.*, 1998a) by subtracting the component mean and then dividing the component standard deviation. Voxels with $|z| \geq 1$ were then identified as those activated by the corresponding stimulus, and they were given value 1 while the voxels with $|z| < 1$ were assigned as 0 when plotting.

The results from SSVD-ICA are shown in the first column of Figure 3.4. All the four components can be recovered reasonably well, although some noise does exist. Similar to Figure 3.3, dark red areas indicate the activated voxels. The time course plots in each
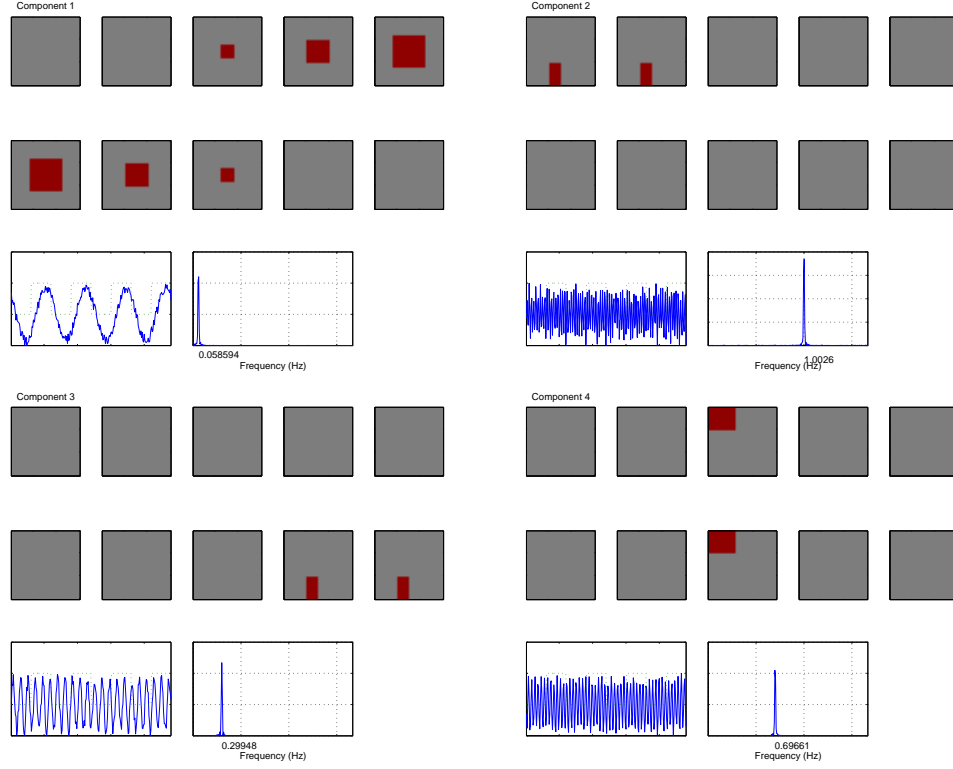
Figure 3.3: *The first four components used in the simulation. In each panel, the first 10 images are the spatial component maps (one column of $\mathbf{A}$), and the dark red areas stand for activated voxels. The solid line in the subsequent plot is the corresponding time series (one row of $\mathbf{S}$). The dotted line stands for the rest-activation block design, 0 for "rest" and 1 for "active". The spectrum plot for each time series is given at the end of each panel highlighting the frequencies. In this simulation study, Component 1 can be viewed as the one related to the experiment stimulus. Components 2 and 3 stand for heart beat and breath respectively. Component 4 could be an artifact effect. Component 5 is not shown here since it is pure noise.*

panel show the corresponding time components, along with their spectrum plots.

The second column of Figure 3.4 shows the results from the conventional ICA. Affected by the spikes, only the first component can be recovered; and the result is more blurred than the corresponding one in the left column. The conventional ICA has trouble identifying the remaining components.

3.5 A Real fMRI Data Analysis

3.5.1 Experiment Paradigm and Data Description

To study brain regions that are related to different finger tapping movements, an fMRI data set was obtained from one human subject performing three different tasks alternately: rest, right-hand movement and left-hand movement. Each rest period lasts 30 seconds and each activation period lasts 120 seconds. The experimental paradigm is shown in Figure 3.5. Note the block design is used because the study is interested in robustly identifying all the regions of interests involving the finger tapping.

During the experiment, two hundred MR scans were acquired on a modified 3T Siemens MAGNETOM Vision system. Each acquisition consisted of 49 contiguous slices. Each slice contained 64×64 voxels. Hence there were $64 \times 64 \times 49$ voxels from each scan. The size of each voxel is 3mm $\times$ 3mm $\times$ 3mm. Each acquisition took 2.9388 seconds, with the scan to scan repetition time (TR) set to be 3 seconds.

3.5.2 Analysis and Results

The data set was preprocessed using SPM5. The preprocessing included realignment, coregistration, segmentation, spatial normalization and smoothing. After the preprocessing, we used the MATLAB function **showsrs** developed by the Duke-UNC Brain Imaging and Analysis Center (BIAC) to visually check the processed data set. Many spikes can be easily seen, three of which are plotted in Figure 3.1.

We then applied both the proposed SSVD-ICA and ICA to the processed data set after column centering and row standardization. When applying ICA, we employed an
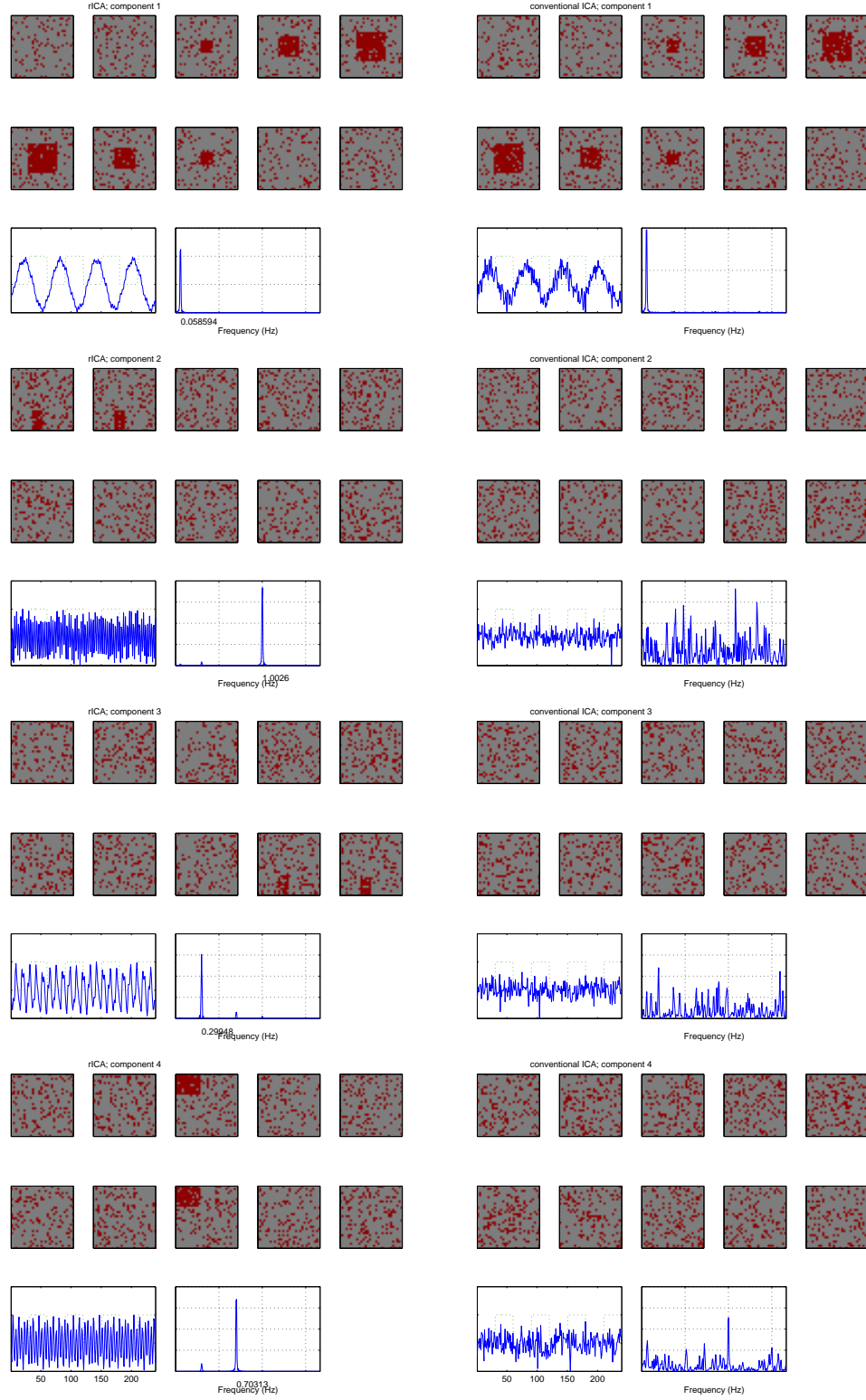
Figure 3.4: *Comparison of the results from the proposed SSVD-ICA (the left column) and the conventional ICA (the right column).*
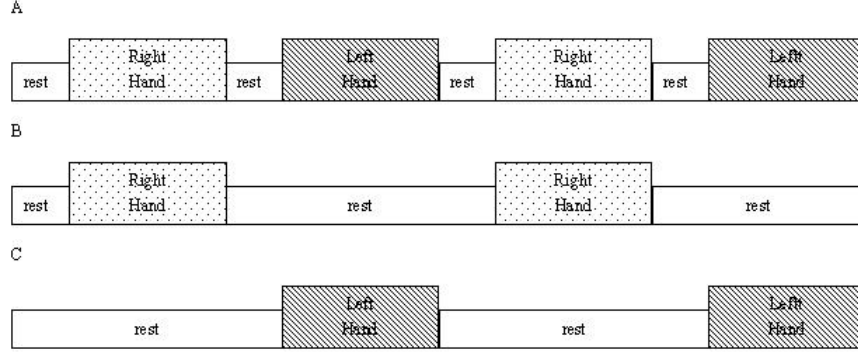
Figure 3.5: *The experimental design used in acquiring the fMRI data. Panel (A) shows the complete design sequence. Panels (B) and (C) show the paradigms for right-hand and left-hand movements separately. Each rest block took* 30 *seconds (*10 *scans when TR = 3 seconds). Each activation block took* 120 *seconds (*40 *scans).*

existing analyzer of fMRI data, *Group ICA of fMRI Toolbox* (GIFT) (Calhoun *et al.*, 2001). With SSVD-ICA, we performed separate analyses using both the true frequencies and the estimated ones. For all the three methods, spatial ICA was carried out using the fastICA algorithm.

The results from the above three methods are displayed in the corresponding panels of Figure 3.6. The goal of this study is to identify the activated brain regions corresponding to right-hand and left-hand movements. We only showed the four related brain slices.

Within each row, the four image slices on the left represent the activated spatial maps. The red and blue areas illustrate activated brain regions. Brighter color indicates higher intensity. All methods demonstrate the classic brain activation patterns during hand movement (Buhmann *et al.*, 2003; Elsinger *et al.*, 2006; Taniwaki *et al.*, 2006), namely

- *contralateral more than ipsilateral in the Primary Motor Cortex* (PMC) (slice 1 and slice 2), *Supplementary Motor Area* (SMA) (the bright area in the middle of slice 2) and *basal ganglia* (slice 3),

- *ipsilateral more than contralateral in the cerebellum* (slice 4).

The plot on the rightmost is the corresponding time component (solid line) with the stimulus sequence (dotted line) overlayed.
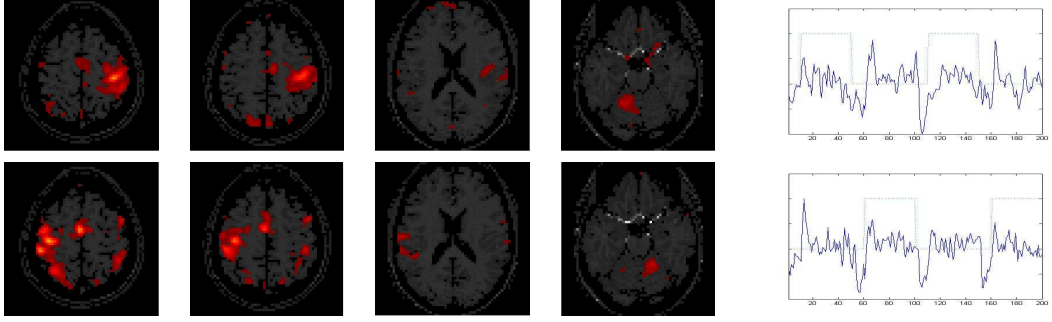
A: GIFT

B: SSVD-ICA using estimated frequencies
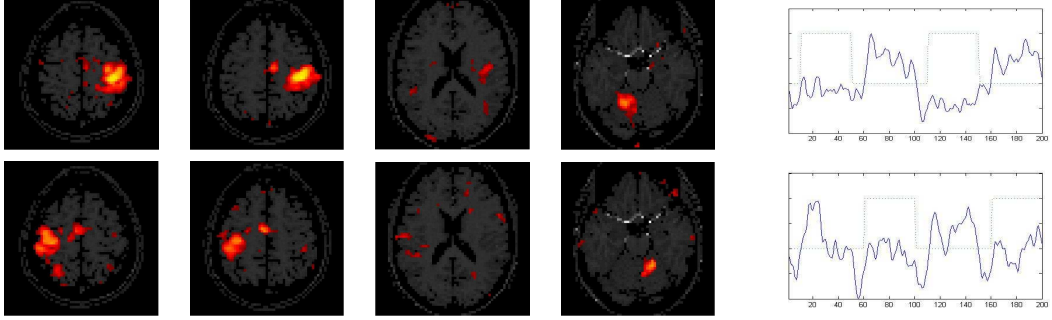
C: SSVD-ICA using specified frequencies
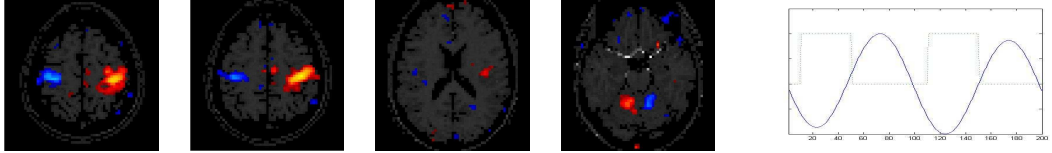
Figure 3.6: *Activated brain regions and their corresponding temporal components detected by three methods, Panel (A): GIFT, right-hand (first row) and left-hand (second row); Panel (B): SSVD-ICA using estimated frequencies, right-hand (first row) and left-hand (second row); Panel (C): SSVD-ICA using specified frequencies, both right-hand and left-hand. Within each row, the first slice shows the primary motor cortex (PMC), the second slice contains both PMC and supplementary motor area (SMA), the third slice shows basal ganglia and the fourth slice shows cerebellum. Red and blue areas illustrate the activated voxels. Brighter color indicates higher intensity. The last plot shows the corresponding time course with the hand-movement stimulus sequence overlayed.*

When applying GIFT, 30 independent components need to be extracted in order to obtain the ones corresponding to right-hand and left-hand movements. The two components are shown in the two rows of Panel (A) respectively. The corresponding components identified by SSVD-ICA using estimated frequencies are presented in Panel (B). It's clear that our method leads to much less noise and higher intensity in all four activated regions. In addition, the time components recovered by our method are better correlated with the experimental stimulus than those obtained by GIFT.

To apply SSVD-ICA using specified frequencies, we specify the true experimental frequency (0.0033). Note that both right-hand and left-hand movements share the same frequency in this study. As a result, the brain regions related to both movements are recovered in a single component as shown in Panel (C), where red areas indicate voxels activated by the right-hand movement and blue areas indicate voxels activated by the left-hand movement. Again, the activated areas are less noisy and more highlighted when compared with GIFT. Another advantage of specifying the frequency is that we only need to extract one component that is relevant to the task paradigm, instead of acquiring 30 components and then identifying the interesting components via visual inspection as done by GIFT.

3.6 Discussion

In this chapter, we introduced an experiment-guided dimension reduction technique into the ICA procedure for fMRI studies, which improves its robustness against spikes that are common in fMRI data sets. We modify the conventional SVD via basis expansion that makes use of the experiment design information. The main benefits are two-fold: first, the extracted components are less sensitive to the spikes; second, the activated brain areas are identified with less noise and higher intensity. Our method is proven to be powerful and advantageous for handling the situation with spikes in a simulation study and a real comparison with GIFT.

Another advantage of our method is to save computing time and the need for visual inspection to pick out the components of interest. One known shortcoming of ICA is that the extracted components are not ordered. Thus the components of interest may not necessarily come out early during the analysis. This explains the common practice that researchers usually acquire around 30 independent components, before searching for the most relevant components by visual inspection. In our method, we can decide on the number of components by specifying the frequencies of interest. This hybrid of data and hypothesis makes the analysis procedure meaningfully more efficient, considering the usually huge size of fMRI data.

In conclusion, our method generates less noisy and faster results with higher intensity when we know the frequencies of interesting components. In addition, by estimating the dominant frequencies instead of specifying them, we can expect to detect components that are not foreseen by experimenters. In either case, our method performs better than the conventional ICA under the presence of spikes.

One further investigation is to refine the procedure for more general experiment designs. As shown in Figure 3.5, neither the paradigm for right-hand movement nor the one for left-hand movement is strictly sinusoidal. Hence the sine curves we used in the SSVD procedure are rather rough approximation to the real sequence. This motivates us to find a more general way to estimate the time components for such situations. Currently, a project that implements a penalized smoothing technique to estimate the time component is under way.

Assessing the Variability of ICA

4.1 Introduction

Independent component analysis (ICA) aims to solve the blind signal separation (BSS) problem by expressing a set of random variables (observations) as linear combinations of statistically independent latent component variables (source signals). It has been applied to fMRI data as an exploratory data analysis technique in order to find independently distributed spatial patterns that depict source processes in the data (Section 3.2.1).

Let $\mathbf{x}$ be an N-dimensional vector variable, whose elements are signal mixtures recorded at one time point, and $\mathbf{s} = (s_1, \ldots, s_K)^T$ be a K-dimensional vector variable with each element being a source signal at the time point. The typical ICA model is written as

$$\mathbf{x} = \mathbf{As}, \tag{4.1}$$

where $\mathbf{A}$ is an $N \times K$ mixing matrix, N is the number of signal mixtures and K is the number of source signals. Both the unknown mixing matrix $\mathbf{A}$ and the source signals $\mathbf{s}$ need to be estimated only using the observed signal mixture $\mathbf{x}$.

As mentioned in Section 3.2.1, to achieve the estimates of $\mathbf{A}$ and $\mathbf{s}$, ICA has two critical assumptions: (1) the source signals are statistically independent and (2) they follow non-Gaussian distributions. According to the Central Limit Theorem, the distribution of a sum of independent random variables goes to a Gaussian distribution under certain assumptions. This means that the distribution of the sum of several independent random

variables is more Gaussian than any of the original random variables. Based on this fact, ICA identifies the independent components by adjusting an unmixing matrix $\mathbf{W}$ to maximize the non-Gaussianity of $\mathbf{Wx}$. Then $\mathbf{A}$ is estimated as $\mathbf{W}^{-1}$ and the source signals are identified as $\hat{\mathbf{s}} = \mathbf{Wx}$. There are several strategies for extracting source signals from the mixtures, each strategy has resulted several methods. One of the early strategies is to seek an unmixing matrix $\mathbf{W}$ which maximizes the joint entropy of the signals $\hat{\mathbf{s}} = \mathbf{Wx}$. The extracted signals will then be independent. To facilitate this, Bell and Sejnowski (1995) used super-gaussian (high-kurtosis) model for the probability density function (pdf) of the source signals, this led to the ICA method known as **infomax**. This method is also closely related to the maximum likelihood approach of ICA (Cardoso, 1997). FastICA is another commonly used ICA algorithm (Hyvärinen and Oja, 2000), in which the source pdf has a parametric functional form. In recent years, there has been a great interest to estimate the source pdf directly, pioneered by Bach and Jordan (2002) and Chen (2005); Chen and Bickel (2005) based on kernel density estimation, Hastie and Tibshirani (2002) using a penalized spline approach, and Kawaguchi and Truong (2007) based on polynomial splines.

The main objective of the current chapter is to evaluate some of these methods by assessing the variability of the estimates of the unmixing matrix $\mathbf{W}$ using a bootstrap methodology. The procedure is to draw bootstrap samples from the estimated source pdf in order to estimate the distribution of $\mathbf{W}$.

The rest of this chapter is structured as follows. Section 4.2 summarizes the methodology of maximum likelihood (ML) ICA algorithms and in particular, introduces two popular ML ICA algorithms in detail. In Section 4.3, we propose a bootstrap procedure to assess the variability of ICA. A simulation study is reported in Section 4.4 to compare two ICA algorithms as regard to the estimate the variability of the procedure. Some concluding remarks will be given in Section 4.5.

4.2 Maximum Likelihood ICA

Early ICA algorithms were developed to minimize the mutual information between the components of the estimate $\mathbf{Wx}$, which is known as infomax (Bell and Sejnowski, 1995). However, the mutual information is difficult to approximate and optimize based on a finite sample. Another common family of algorithms of ICA is to make use of the maximum likelihood (ML) method for estimating the optimal unmixing matrix (Stone, 2004; Hastie and Tibshirani, 2002; Bach and Jordan, 2002; Chen, 2005). It has been proven that the infomax method is essentially equivalent to the ML approaches (Stone, 2004). In the following, we study maximum likelihood ICA in more detail.

In the ICA model (4.1), the independent source components (ICs) $s_1, \ldots, s_K$ are taken as latent variables. To make the problem (4.1) solvable, it's necessary to assume that $N \geq K$. Without loss of generality, we assume that $N = K$. Hence, the mixing matrix $\mathbf{A}$ is of dimension $K \times K$.

The ML approaches include a specification of the probability density function (pdf) of the unknown source signals $\mathbf{s}$. The goal of ML ICA is then to find an unmixing matrix that generates $\mathbf{Wx}$ with a joint pdf as similar as possible to the joint pdf of the unknown source signals.

Suppose that the density function of $\mathbf{x}$ is $f(\cdot)$ and each s_k has a density function $g_k(\cdot)$ for $k = 1, 2, \ldots, K$. Let $\mathbf{W} = \mathbf{A}^{-1}$ be the unmixing matrix. If $s_1, s_2, \ldots, s_K$ are independent with marginal density functions $g_1, g_2, \ldots, g_K$, then $f(\mathbf{x}) = \prod_{k=1}^{K} |\mathbf{W}| g_k(\mathbf{e}'_k \mathbf{Wx})$, where $\mathbf{e}_k$ is the kth column of the $K \times K$ identity matrix so that $s_k = \mathbf{e}'_k \mathbf{Wx}$. Thus the log-likelihood function of $\mathbf{W}$ based on the data is

$$\log f(\mathbf{x}) = \sum_k \log g_k(\mathbf{e}'_k \mathbf{Wx}) |\mathbf{W}|.$$

For fMRI data in which $\mathbf{x} = \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iK})'$ denotes a voxel time series which is a mixture random vector at voxel i with density function $f(\cdot)$, and $\mathbf{s} = \mathbf{s}_i =$

$(s_{i1}, s_{i2}, \ldots, s_{iK})'$ a source vector at voxel i with density function $g(\cdot)$ whose marginal density functions are $g_1, g_2, \ldots, g_K$, to account for all the temporal data, the log-likelihood function of $\mathbf{W}$ is obtained by taking the time average of the above likelihood so that

$$l(\mathbf{W}) = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{K} \log g_k(\mathbf{e}_k' \mathbf{W} \mathbf{x}_i) + \log |W|, \qquad (4.2)$$

where M is the number of voxels in the fMRI data set.

Traditional ICA algorithm FastICA assumes the sources are identically distributed with a common density function g_1 whose functional form is also known (Hyvärinen and Oja, 2000) and it's thus limited to the parametric form of g_1. Recently, some nonparametric methods have been proposed to estimate the unknown distributions of the hidden sources. For example, Hastie and Tibshirani (2002) uses penalized splines for the estimation of $g_1, g_2, \ldots, g_K$, while Bach and Jordan (2002) and Chen (2005) consider kernel estimates (KDICA). More recently, Kawaguchi and Truong (2007) proposes a new ML ICA algorithm that models the distribution of the independent source components using polynomial splines with data-dependent knot locations. This is referred to as SICA. We will provide a more detailed description of these methods in the next two sections.

4.2.1 KDICA

Chen (2005) proposes a fast KDICA algorithm, which considers kernel density estimates of $g_1, g_2, \ldots, g_K$. The goal of the algorithm is to estimate $\mathbf{W}$ and $g_1, g_2, \ldots, g_K$ by maximizing the log-likelihood function (4.2). Since both $\mathbf{W}$ and $g_k, k = 1, \ldots, K$ are unknown, the algorithm starts with an initial $\mathbf{W}$, which can be obtained from FastICA or other ICA algorithms. When $\mathbf{W}$ is known, g_k is identical to the density function of $\mathbf{e}_k' \mathbf{W} \mathbf{x}$. Hence g_k can be estimated by the kernel density estimator $\hat{g}_k(s) = 1/(Kh) \sum_{i=1}^{K} \mathcal{K}((\mathbf{e}_k' \mathbf{W} \mathbf{x}_i - s)/h)$, where the Laplacian kernel is used for the kernel function $\mathcal{K}(\cdot)$ and the bandwidth h is selected as $0.6\hat{\sigma} K^{-1/5}$ with $\hat{\sigma}$ being the sample standard deviation of $\mathbf{e}_k' \mathbf{W} \mathbf{x}_i$. Once the estimates of g_k's are obtained, they can be

plugged back into (4.2) and $\mathbf{W}$ can be updated by maximizing the log-likelihood function. The algorithm iterates in the above way till convergence, using Amari metric (Bach and Jordan, 2002), a measure of the closeness of two matrices defined as

$$d(\mathbf{A}_0, \mathbf{A}) = \frac{1}{2m} \sum_{i=1}^{m} \left(\frac{\sum_{j=1}^{m} |r_{ij}|}{\max_j |r_{ij}|} - 1 \right) + \frac{1}{2m} \sum_{j=1}^{m} \left(\frac{\sum_{i=1}^{m} |r_{ij}|}{\max_i |r_{ij}|} - 1 \right), \qquad (4.3)$$

where $r_{ij} = (\mathbf{A}_0 \mathbf{A}^{-1})_{ij}$ and the dimension of $\mathbf{A}_0$ and $\mathbf{A}$ is $m \times m$.

When estimating the kernel densities of g_k's, Chen (2005) proposes a FastKDE method, which improves the computation performance to a great extent.

4.2.2 SICA

More recently, Kawaguchi and Truong (2007) developed a new approach using polynomial splines to model the logarithmic of $g_1, g_2, \ldots, g_K$ (SICA). In their study, each logarithmic density is modeled using polynomial splines

$$\log g_k(s) = \beta_{k00} + \beta_{k01}s + \sum_{i}^{m_k} \beta_{k1i}(s - r_{ki})_+^3,$$

where $\boldsymbol{\beta}_k = (\beta_{k00}, \beta_{k01}, \beta_{k11}, \ldots, \beta_{k1m_k})$ is a vector of coefficients, r_{ki} are the knots and m_k is the number of knots for the kth source density function. The knot selection in this algorithm starts with an initial knot placement, which is set to be the minimum, median and maximum values of the data. Then the knot selection methodology involves stepwise knot addition, stepwise knot deletion and final model selection based on Bayesian information criterion (AIC), which is defined by

$$\text{BIC}_k = -2l(\hat{\boldsymbol{\beta}}_k) + m_k \log K.$$

The same as KDICA, the algorithm starts with an initial $\mathbf{W}$ and optimize the density functions $g_1, g_2, \ldots, g_K$ and $\mathbf{W}$ in an alternate way until convergence using Amari metric.

4.3 Assessing the Variability of ICA

Based on the above two ICA algorithms, we developed the following bootstrap procedure to assess the variability of the estimated mixing matrix $\mathbf{A}$.

$$\boxed{\textbf{Algorithm 2. Assess the Variability of ICA}}$$

(1) Apply KDICA or SICA on the observed mixtures $\mathbf{x}$ to obtain the estimate of the mixing matrix $\hat{\mathbf{A}}$ and the marginal density function of the hidden sources $\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_K$.

(2) Resample $\mathbf{s}_1^*, \mathbf{s}_2^*, \ldots, \mathbf{s}_B^*$ from $\hat{g}_1, \hat{g}_2, \ldots, \hat{g}_K$, where B is the bootstrap sample size.

(3) Form B new mixture matrices $\mathbf{x}_i^* = \hat{\mathbf{A}}\mathbf{s}_i^*$, $i = 1, 2, \ldots, B$.

(4) Apply KDICA or SICA on $\mathbf{x}_i^*$ to obtain $\mathbf{A}_i^*$, where $\mathbf{x}_i^* = \mathbf{A}_i^*\mathbf{s}_i^*$, $i = 1, 2, \ldots, B$.

(5) Compute the Amari distance between $\hat{\mathbf{A}}$ and each $\mathbf{A}_i^*$, as defined by (4.3), to obtain d_i, $i = 1, 2, \ldots, B$.

(6) Find the 5th and 95th percentiles of $d_1, d_2, \ldots, d_B$. The $\mathbf{A}_i^*$'s that correspond to the two percentiles form the 90% confidence region of $\hat{\mathbf{A}}$.

Remarks For KDICA algorithm, rejection-acceptance resampling method is used (Robert and Casella, 2004). In SICA, bootstrap samples are generated from the logspline density estimate in the manner similar to the r option in R or S.

4.4 A Simulation Study

In this simulation study, we simulated a simple case with two hidden sources. Each source signal is sampled from a Student's t distribution with degree freedom of 15. The 2×2 mixing matrix is randomly generated from a uniform distribution. The mixture matrix is generated by $\mathbf{X} = \mathbf{AS}$, where each row of $\mathbf{S}$ contains a source signal with

$N = 240$
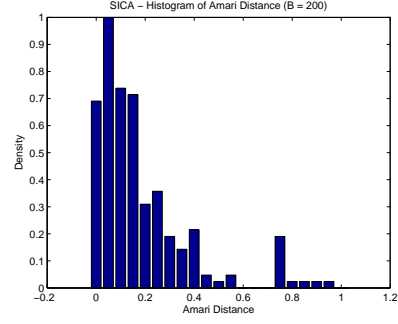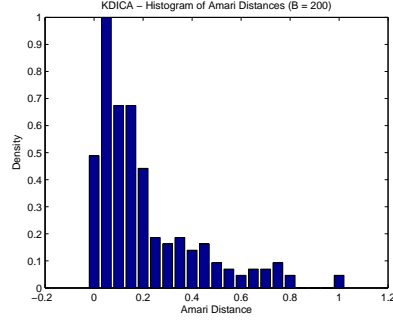
$N = 1000$
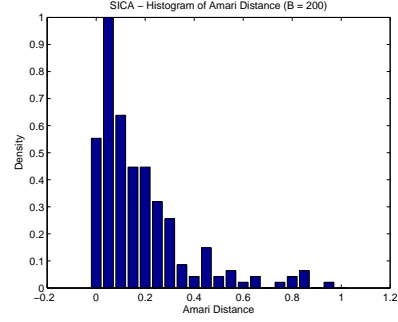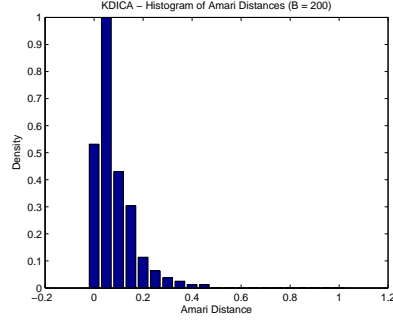
Figure 4.1: *The histograms of the Amari distances obtained by applying Algorithm* 2 *using two ICA methods with sample size being* 200, *when the length of the simulated source signals is* 240 *and* 1000 *respectively. Top Left: KDICA with the length of the source signals* 240; *Top Right: SICA with the length of the source signals* 240; *Bottom Left: KDICA with the length of the source signals* 1000; *Bottom Right: SICA with the length of the source signals* 1000.

length 240 or 1000. Assuming that we do not know the distribution of $\mathbf{S}$, which is true in practice, we use Algorithm 2 to assess the variability of the estimate $\hat{\mathbf{A}}$.

We studied both KDICA and SICA as described in the previous section. For each method, we carried out Algorithm 2 twice with bootstrap sample size being 200, when the length of the source signals is 240 and 1000 respectively. The histograms of the resulting Amari distances are plotted in Figure 4.1. These results are summarized in Table 4.1.

From Figure 4.1 and Table 4.1, we can see that when the length of the independent source signals is relatively small, SICA performs a bit better than KDICA as regard to reducing the estimation variance. While when the length of the independent source signals is large, KDICA performs better than SICA. This means that in different situ-

N	ICA Method	Mean	90% Confidence Interval	
			Lower	Upper
240	KDICA	0.2045	0.0122	0.6771
	SICA	0.1813	0.0077	0.7441
1000	KDICA	0.0835	0.004	0.2428
	SICA	0.181	0.0106	0.5675

Table 4.1: *Comparison of two ICA methods as regard to the variability of the estimate* $\hat{\mathbf{A}}$ *using Algorithm* 2. N *is the length of the simulated source signals.*

ation, different ICA algorithms do have different performance regarding the estimation variance. Using the proposed bootstrap procedure, we can try to find the optimal ICA algorithm as regard to controlling the estimation variance in certain context.

4.5   Discussion

This chapter introduced a nonparametric bootstrap procedure to assess the variability of ICA estimation. This procedure can be summarized as follows. The observed signals are used to provide estimates of the unmixing matrix $\mathbf{W}$ and the source pdf. These are then used to generate bootstrap samples, which in turn yield bootstrap estimates of $\mathbf{W}$. Since these are matrices, and it will be difficult to visualize the distribution unless they are projected to the line. One way to do this is to compute the Amari distances of the bootstrap estimates and the estimate based on the original samples. The distribution (or histogram) of these distances will be constructed to study the variability of the unmixing estimates.

In the parametric ICA procedures such as infomax or FastICA, the bootstrap samples are obtained from the sub-gaussian or high-kurtosis model with parameters estimated by maximum likelihood estimates. In the non-parametric side, the source pdf's are estimated by either kernel or spline based procedures. Bootstrap samples are drawn from these estimates. In our simulated study, random (bootstrap) samples from the kernel estimate are drawn by using the rejection-acceptance method and the performance is associated with the sampling density function. The rejection-acceptance part has been relatively

easy for student's t distributions, but it is not so for other distributions. On the other hand, bootstrap samples are quite easily generated from the spline-based estimates of the source pdf.

In the current study, we have considered the student's t distribution and two non-parametric ICA procedures. Based on these preliminary results, we find that SICA is less variable when the sample signals are short. It will be instructive to plan a more extensive study in which we will compare more ICA algorithms with more variety of density functions such as those studied by Bach and Jordan (2002). More importantly, the number of components should be larger than two for the proposed procedure to be applicable in fMRI studies.

CHAPTER 5

Adaptive SPM

5.1 Introduction

Model-driven strategies have been playing a dominant role for fMRI analysis. Among them, the most widely used approach is statistical parametric mapping (SPM), which includes a two-stage analysis. In the first stage, a general linear model (GLM) with correlated errors is used for each voxel time series (Friston *et al.*, 1995b).

Consider an fMRI matrix $\mathbf{Y} = (y_{ij})_{N \times M}$, where M is the number of voxels on one image, N is the number of time points of the experiment and y_{ij} is the (i, j)th element of $\mathbf{Y}$. The typical GLM of fMRI can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{\Sigma}), \tag{5.1}$$

where each column of $\mathbf{Y}$ contains a time series of N time points for one voxel. Suppose there are K different conditions involved in the experiment. For example, there are four different hand movements included in the experiment of the fMRI study in Section 5.4. Then $\mathbf{X}$ is an $N \times (K + 1)$ design matrix, where the first column is usually constant representing the condition when the subject is at rest with no experimental activities performed, and each of the remaining columns contains the time component for an experiment condition. The $(K + 1) \times M$ matrix $\boldsymbol{\beta}$ contains the related parameters, and due to the special formulation of the design matrix, the second to the last rows naturally give the contrast between each experiment condition and the benchmark when the sub-

ject is at rest. The inferences on each row of $\boldsymbol{\beta}$ will generate a spatial map related to one experiment condition. Random noise $\boldsymbol{\epsilon}$ is assumed to have normal distribution with covariance matrix $\boldsymbol{\Sigma}$. If we assume $\boldsymbol{\epsilon}$ is uncorrelated white noise with variance equal to σ^2 at all time points, then $\boldsymbol{\Sigma}$ becomes identity. Note that in this chapter we are using a different set of notations for the fMRI data matrix from Chapter 3. The reason is to be consistent with conventional notations used in the context of SPM.

At each voxel, suppose $\mathbf{Y}_j$ is a column vector of the time series of the jth voxel, then the GLM for this voxel is

$$\mathbf{Y}_j = \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\epsilon}_j, \qquad j = 1, 2, \ldots, M, \tag{5.2}$$

Once this model has been fitted at each voxel, inferences of the model parameters are then made according to the experiment hypothesis. The resulting statistics from all the voxels are assembled spatially into an image, which is the so-called *statistical parametric map*. The second stage then focuses on the analysis of the statistical map in order to identify those areas of the brain that are activated by the stimuli.

The nature of the BOLD response implies that in areas of activation there is a delayed and blurred version of the stimulus sequence (Figure 3.2). Hence each column of the design matrix $\mathbf{X}$, which represents the temporal characteristic of the expected response, is commonly modeled through the convolution of the stimulus design $s(\cdot)$ with a hemodynamic response function (HRF), $h(\cdot)$. That is,

$$\mathbf{X}_k(t) = \sum_u h(t - u)s_k(u),$$

where $\mathbf{X}_k$ is the kth column of the design matrix. $s_k(\cdot)$ is the stimulus sequence of the kth experiment condition, which usually consists of zeros and ones, where one stands for activation and zero stands for rest.

The modeled $\mathbf{X}_k$ then indicates the temporal characteristic of the brain regions that
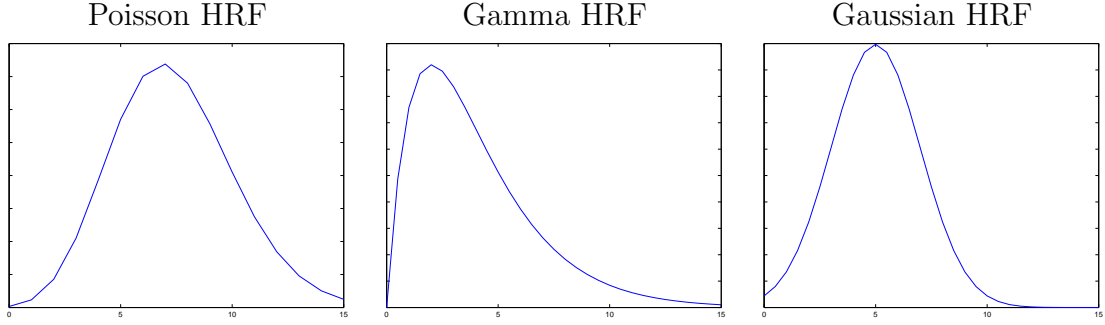
Figure 5.1: *The HRF modeled by Poisson distribution (left), Gamma distribution (middle) and Gaussian distribution (right).*

are activated by the kth experiment condition. In the literature, commonly used forms of HRF include discretized Poisson, Gamma and Gaussian density functions (Figure 5.1). However, these HRFs assume predefined parametric forms, which are rather restrictive. Some work has been proposed to model the HRF using a small set of temporal basis functions to improve the flexibility of its form (Friston *et al.*, 1995b; Josephs *et al.*, 1997). One drawback of the basis approach is that it introduces less sensitivity of the estimation and the results are more difficult to interpret (Kherif *et al.*, 2002). More recently, attempts have been made to combine data-driven and model-driven methods in a complementary way, for example, by Hu *et al.* (2005), Rayens and Andersen (2006) among others. Instead of assuming a certain form for the HRF, the authors model the HRF or the temporal component directly making use of the information extracted by some data-driven methods, such as independent component analysis (ICA) or principal component analysis (PCA).

In this work, we propose a new nonparametric method to model the time component adaptively in the context of SPM. The idea is to start from an initial design matrix $\mathbf{X}$ obtained from ordinary SPM procedure and then apply a penalized smoothing technique to update the shape of the hemodynamic response in an adaptive way. Our approach, namely *Adaptive SPM*, makes use of the timing information of various experimental

stimuli offered by the experiment design, and adjust the shape of the temporal response of the activated voxels so that it's closer to the reality. Because of the improved accuracy of the estimated temporal characteristic, the corresponding detection power of the activated regions is increased as well. The nice performance of the proposed method is illustrated through a simulation study, as well as a comparative study on a real fMRI data analysis.

The rest of this chapter starts with a description of our proposed method in Section 5.2. Section 5.3 reports a simulation study to illustrate its performance. A comparative study on a real fMRI data set is presented in Section 5.4. Some discussion is given in Section 5.5 to close this chapter.

5.2 Adaptive SPM

The GLM model (5.1) can also be expressed as

$$y_{ij} = x_{i0}\beta_{0j} + x_{i1}\beta_{1j} + \cdots + x_{iK}\beta_{Kj} + \epsilon_{ij}, \quad i = 1, \ldots, N, \ j = 1, \ldots, M.$$

Hence, for the image recorded at time i, $i = 1, \ldots, N$, we have

$$
\begin{aligned}
\mathbf{Y}_{(i)} &= (y_{i1}, \ldots, y_{iM})^T \\
&= x_{i0}(\beta_{01}, \ldots, \beta_{0M})^T + x_{i1}(\beta_{11}, \ldots, \beta_{1M})^T + \ldots + x_{iK}(\beta_{K1}, \ldots, \beta_{KM})^T + \boldsymbol{\epsilon}_{(i)} \\
&= x_{i0}\boldsymbol{\beta}_{(0)} + x_{i1}\boldsymbol{\beta}_{(1)} + \cdots + x_{iK}\boldsymbol{\beta}_{(K)} + \boldsymbol{\epsilon}_{(i)},
\end{aligned}
$$

where $\boldsymbol{\epsilon}_{(i)} = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{iM})^T, \boldsymbol{\beta}_{(k)} = (\beta_{k1}, \ldots, \beta_{kM})^T$ for $k = 0, 1, \ldots, K$.

Concatenating all the images recorded at different time points, we have

$$
\begin{pmatrix} \mathbf{Y}_{(1)} \\ \mathbf{Y}_{(2)} \\ \vdots \\ \mathbf{Y}_{(N)} \end{pmatrix} = \sum_{k=0}^{K} \begin{pmatrix} \boldsymbol{\beta}_{(k)} & \tilde{0} & \cdots & \tilde{0} \\ \tilde{0} & \boldsymbol{\beta}_{(k)} & \cdots & \tilde{0} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{0} & \tilde{0} & \cdots & \boldsymbol{\beta}_{(k)} \end{pmatrix} \mathbf{X}_k + \begin{pmatrix} \boldsymbol{\epsilon}_{(1)} \\ \boldsymbol{\epsilon}_{(2)} \\ \vdots \\ \boldsymbol{\epsilon}_{(N)} \end{pmatrix}, \tag{5.3}
$$

47
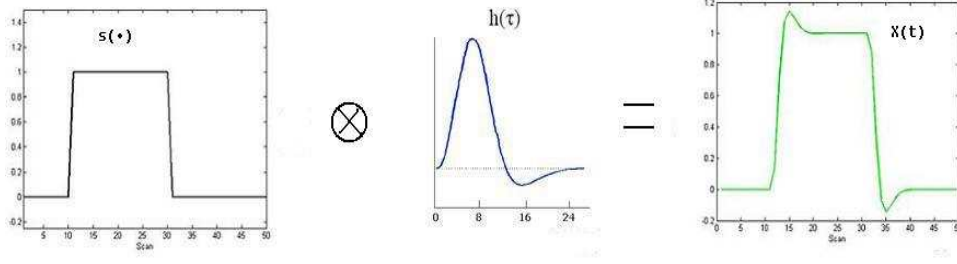
Figure 5.2: *The experiment stimulus sequence (left) is convolved with the canonical HRF (middle) to obtain one column of the design matrix* **X** *(right).*

where $\mathbf{X}_k = (x_{1k}, \ldots, x_{Nk})^T$, for $k = 0, 1, \ldots, K$ and $\tilde{0}$ is an $M \times 1$ vector of zeros.

Classical SPM approaches fit the above GLM by assuming $\mathbf{X}_0 = (1, 1, \ldots, 1)^T$, and obtaining $\mathbf{X}_1, \ldots, \mathbf{X}_K$ by convolving a predefined HRF with the experimental stimulus sequence for each condition respectively. For example, the canonical HRF used in SPM5 is a typical BOLD impulse response characterized by two gamma functions, one modeling the peak and one modeling the undershoot. The canonical HRF is plotted in the middle of Figure 5.2.

Convolve the canonical HRF with the stimulus sequence of each condition in the experiment, we obtain an initial design matrix $\mathbf{X}$. To make the model (5.3) identifiable, we assume $||\mathbf{X}_0|| = ||\mathbf{X}_1|| = \cdots = ||\mathbf{X}_K|| = 1$. In the GLM analysis of fMRI, the stimulus sequence is usually represented by a sequence of zeros and ones, where zero means rest and one means activation by a certain condition. Hence, after convolving with the selected HRF, the modeled time component for a certain condition consists of nonzero values only during the activation period. Figure 5.2 illustrates this process. The plot on the leftmost shows a stimulus sequence with one meaning activation and zero meaning rest. The plot in the middle is the canonical HRF and the time component modeled through the convolution is plotted in the rightmost graph. We can obtain the starting and ending points of the blood flow for each experiment stimulus from the initial design matrix. Our goal is then to estimate the shape of the nonzero part of the blood flow nonparametrically.

We propose to estimate the temporal and spatial components ($\mathbf{X}$ and $\boldsymbol{\beta}$) in an iterative way as described below in Algorithm 3. Each step is then elaborated on in the following sections.

$$\boxed{\textbf{Algorithm 3. Adaptive SPM}}$$

(1) Obtain the initial design matrix $\mathbf{X}$ from SPM5.

(2) Alternate until the convergence of $\mathbf{X}$ and $\boldsymbol{\beta}$.

    (a) Based on the given $\mathbf{X}$, solve a GLM equation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ to obtain $\boldsymbol{\beta}$;

    (b) Once $\boldsymbol{\beta}$ is obtained, use a penalization procedure to derive $X_1, \ldots, X_K$ smoothly.

5.2.1 Obtain the Initial Design Matrix

At the beginning of the algorithm, the initial design matrix $\mathbf{X}$ is obtained from SPM5. The first column of $\mathbf{X}$ is a vector of ones. Each of the remaining columns is obtained by convolving the stimulus sequence of a certain condition with a canonical HRF. Each stimulus sequence consists of ones and zeros, where one means activation and zero means rest. Each column of $\mathbf{X}$ is rescaled to have norm 1, to make the model (5.3) identifiable.

5.2.2 Obtain the Spatial Map

Based on the obtained design matrix $\mathbf{X}$, we can solve the following GLM equation

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 \boldsymbol{\Sigma}),$$

where $\mathbf{X} = (\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_K)$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_{(0)}^T, \boldsymbol{\beta}_{(1)}^T, \ldots, \boldsymbol{\beta}_{(K)}^T)^T$. Assume $\boldsymbol{\Sigma}$ to be identity, the $\boldsymbol{\beta}$ that minimizes $||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2$ is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

5.2.3 Obtain the Smooth Time Components

After $\hat{\boldsymbol{\beta}}$ is obtained, we can then use the following penalization procedure to derive $\mathbf{X}_1, \ldots, \mathbf{X}_K$ smoothly. Note that $\mathbf{X}_0$ remains fixed because it stands for the background rest activity.

Standard SPM assumes the same HRF for all the experimental conditions, hence it's reasonable to assume that their time components share the same nonzero part. However the positions of the nonzero part vary among different time components. Denote the nonzero part of $\mathbf{X}_k$, $k = 1, \ldots, K$, as $\tilde{\mathbf{X}}$ and its length as $\tilde{N}$. Suppose the nonzero part of $\mathbf{X}_k$ starts at index s_k and ends at index e_k.

Then the model (5.3) can be reduced and rewritten as

$$
\begin{pmatrix} \mathbf{Y}_{(1)} \\ \mathbf{Y}_{(2)} \\ \vdots \\ \mathbf{Y}_{(N)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{(0)} & \tilde{0} & \cdots & \tilde{0} \\ \tilde{0} & \boldsymbol{\beta}_{(0)} & \cdots & \tilde{0} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{0} & \tilde{0} & \cdots & \boldsymbol{\beta}_{(0)} \end{pmatrix} \mathbf{X}_0 + \cdots + \sum_{k=1}^{K} \begin{pmatrix} \tilde{0}_{sk} & \cdots & \tilde{0}_{sk} \\ \boldsymbol{\beta}_{(k)} & \cdots & \tilde{0} \\ \vdots & \ddots & \vdots \\ \tilde{0} & \cdots & \boldsymbol{\beta}_{(k)} \\ \tilde{0}_{ek} & \cdots & \tilde{0}_{ek} \end{pmatrix} \tilde{\mathbf{X}} + \boldsymbol{\epsilon},
$$

where $\tilde{0}_{sk}$ is a column vector of zeros with length $(s_k - 1) \times M$ and $\tilde{0}_{ek}$ is a column vector of zeros with length $(N - e_k) \times M$.

Note that in the above model, only $\tilde{\mathbf{X}}$ is unknown and needs to be estimated. Hence it makes sense to re-express the model in the following way,

$$
\tilde{\mathbf{Y}} = \begin{pmatrix} \tilde{\mathbf{Y}}_{(1)} \\ \tilde{\mathbf{Y}}_{(2)} \\ \vdots \\ \tilde{\mathbf{Y}}_{(N)} \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 \\ \tilde{\boldsymbol{\beta}}_2 \\ \vdots \\ \tilde{\boldsymbol{\beta}}_N \end{pmatrix} \tilde{\mathbf{X}} + \boldsymbol{\epsilon}, \tag{5.4}
$$

50

where $\tilde{\mathbf{Y}}_{(i)} = \mathbf{Y}_{(i)} - \boldsymbol{\beta}_{(0)}\mathbf{X}_0$ and $\tilde{\boldsymbol{\beta}}_i$ is a matrix with dimension $M \times \tilde{N}$ that can be obtained from the known $\boldsymbol{\beta}_{(k)}$'s, $i = 1, \ldots, N$, $k = 1, \ldots, K$.

To obtain a smooth $\tilde{\mathbf{X}}$, we consider the following penalized least square (PLS) criterion:

$$
\begin{aligned}
& ||\tilde{\mathbf{Y}} - \tilde{\boldsymbol{\beta}}\tilde{\mathbf{X}}||^2 + \lambda \tilde{\mathbf{X}}^T \boldsymbol{\Omega} \tilde{\mathbf{X}} \\
= & \sum_{i=1}^{N} ||\tilde{\mathbf{Y}}_i - \tilde{\boldsymbol{\beta}}_i\tilde{\mathbf{X}}||^2 + \lambda \tilde{\mathbf{X}}^T \boldsymbol{\Omega} \tilde{\mathbf{X}} \\
= & \sum_{i=1}^{N} (\tilde{\mathbf{Y}}_i^T \tilde{\mathbf{Y}}_i - 2\tilde{\mathbf{Y}}_i^T \tilde{\boldsymbol{\beta}}_i\tilde{\mathbf{X}} + \tilde{\mathbf{X}}^T \tilde{\boldsymbol{\beta}}_i^T \tilde{\boldsymbol{\beta}}_i\tilde{\mathbf{X}}) + \lambda \tilde{\mathbf{X}}^T \boldsymbol{\Omega} \tilde{\mathbf{X}},
\end{aligned}
$$

where $\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \ldots, \tilde{\boldsymbol{\beta}}_N)^T$. The above criterion involves two terms. The first term measures the goodness of fit of the GLM, while the second term penalizes the roughness of $\tilde{\mathbf{X}}$. λ is the smoothing parameter that optimally balances the two terms and $\boldsymbol{\Omega}$ is the $\tilde{N} \times \tilde{N}$ matrix that generates the smoothing penalty.

The minimizer of the above PLS criterion is

$$
\widehat{\tilde{\mathbf{X}}} = \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\beta}}_i^T \tilde{\boldsymbol{\beta}}_i + \lambda \boldsymbol{\Omega} \right)^{-1} \left(\sum_{i=1}^{N} \tilde{\boldsymbol{\beta}}_i^T \tilde{\mathbf{Y}}_i \right).
$$

Once we obtain $\widehat{\tilde{\mathbf{X}}}$, we can normalize it and plug it back to the right positions of each time component to form an updated $\mathbf{X}$.

5.2.4   Selection of the Smoothing Parameter

One thing need to consider in obtaining the smooth time components is the selection of the smoothing parameter λ. We propose to select λ using generalized cross validation (GCV) as following (Hastie *et al.*, 2003).

After obtaining $\widehat{\tilde{\mathbf{X}}}$, we have

$$
\begin{aligned}
\widehat{\tilde{\mathbf{Y}}} &= \tilde{\boldsymbol{\beta}}\widehat{\tilde{\mathbf{X}}} \\
&= \tilde{\boldsymbol{\beta}}\big(\sum_{i=1}^{N}\tilde{\boldsymbol{\beta}}_i^T\tilde{\boldsymbol{\beta}}_i + \lambda\boldsymbol{\Omega}\big)^{-1}\big(\sum_{i=1}^{N}\tilde{\boldsymbol{\beta}}_i\tilde{\mathbf{Y}}_i\big) \\
&= \tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}} + \lambda\boldsymbol{\Omega})^{-1}\tilde{\boldsymbol{\beta}}^T\tilde{\mathbf{Y}} \\
&= \mathbf{S}\tilde{\mathbf{Y}}, &(5.5)
\end{aligned}
$$

where $\mathbf{S} \overset{\text{def}}{=} \tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}} + \lambda\boldsymbol{\Omega})^{-1}\tilde{\boldsymbol{\beta}}^T$.

Define the GCV as:

$$
\begin{aligned}
\text{GCV}(\lambda) &= \frac{\|\widehat{\tilde{\mathbf{Y}}} - \tilde{\mathbf{Y}}\|^2/(MN)}{(1 - \text{tr}(\mathbf{S})/(MN))^2} \\
&= \frac{\sum_{i=1}^{N}\|\tilde{\boldsymbol{\beta}}_i\big(\sum_{i=1}^{N}\tilde{\boldsymbol{\beta}}_i^T\tilde{\boldsymbol{\beta}}_i + \lambda\boldsymbol{\Omega}\big)^{-1}(\sum_{i=1}^{N}\tilde{\boldsymbol{\beta}}_i^T\tilde{\mathbf{Y}}_i)\|^2/(MN)}{(1 - \text{tr}(\mathbf{S})/(MN))^2}.
\end{aligned}
$$

We choose the λ that minimizes the GCV from a set of candidates. We can either choose λ inside the iteration loop or outside the iteration loop.

5.2.5 Remarks and Implementation Details

Remark 1

Our algorithm is proven to be insensitive to the shape of the initial $\tilde{\mathbf{X}}$. All the information we need from the initial design matrix is the starting and ending time points of the blood flow for each component. In practice, if we replace the blood flow $\tilde{\mathbf{X}}$ by a vector of ones with the same length, the algorithm converges to the same result after a few more iterations.

Calculation of tr($\mathbf{S}$)

The matrix $\mathbf{S}$ defined in Equation (5.5) has dimension $MN \times \tilde{N}$. Considering the huge size of typical fMRI data set, the calculation of $\mathbf{S}$ and $\text{tr}(\mathbf{S})$ is rather time and memory consuming. However, in the implementation of Algorithm 3, we can make use

of the following Lemma (Harville, 1997).

Lemma: For any $m \times n$ matrix $\mathbf{A}$ and $n \times m$ matrix $\mathbf{B}$,

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A}).$$

Hence, we have

$$\mathrm{tr}(\mathbf{S}) = \mathrm{tr}\big(\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}} + \lambda\Omega)^{-1}\tilde{\boldsymbol{\beta}}^T\big) = \mathrm{tr}\big(\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}}(\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}} + \lambda\Omega)^{-1}\big).$$

Because both $\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}}$ and $(\tilde{\boldsymbol{\beta}}^T\tilde{\boldsymbol{\beta}} + \lambda\Omega)^{-1})$ have been calculated in previous steps, we can make use of the stored results to obtain the trace of $\mathbf{S}$ instead of calculating $\mathbf{S}$ directly. This way we can improve the performance of the algorithm considering the large size of fMRI data.

Calculation of GCV

Note that in the model (5.4), only the nonzero $\tilde{\boldsymbol{\beta}}_i$'s, that is, the i's $(i = 1, 2, \ldots, N)$ that belong to at least one experiment task period, contribute to the model. Hence we can ignore those times where there are no activities and reindex the nonzero $\tilde{\boldsymbol{\beta}}_i$'s as $\tilde{\boldsymbol{\beta}}_1, \tilde{\boldsymbol{\beta}}_2, \ldots, \tilde{\boldsymbol{\beta}}_{N^*}$, where N^* is the number of time points that belong to at least one activation period. The model (5.4) then becomes

$$\tilde{\mathbf{Y}} = \begin{pmatrix} \tilde{\mathbf{Y}}_{(1)} \\ \tilde{\mathbf{Y}}_{(2)} \\ \vdots \\ \tilde{\mathbf{Y}}_{(N^*)} \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 \\ \tilde{\boldsymbol{\beta}}_2 \\ \vdots \\ \tilde{\boldsymbol{\beta}}_{N^*} \end{pmatrix} \tilde{\mathbf{X}} + \boldsymbol{\epsilon}$$

The remaining computations can be carried out based on this reduced model then. This is another way to make the algorithm more efficient in the implementation.

5.3 A Simulation Study

5.3.1 Data Description

According to the GLM model (5.1), we simulated an $N \times M$ fMRI data matrix $\mathbf{Y}$ by first simulating the $N \times (K+1)$ design matrix $\mathbf{X}$ and the $(K+1) \times M$ parameter matrix $\boldsymbol{\beta}$ separately. The data matrix $\mathbf{Y}$ was then obtained as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is a randomly generated noise matrix from a normal distribution with mean 0 and variance 0.4788. The variance is estimated from the real fMRI study in Section 5.4.

In this study, we set $K = 4$, $M = 30 \times 30 \times 10$ and $N = 120$. The simulated data can be explained as follows: there are 4 different tasks involved in the experiment and we want to detect which brain regions are related to the four tasks respectively. Each column of the design matrix $\mathbf{X}$ represents the blood changes triggered by a particular task. Each row of $\boldsymbol{\beta}$ consists of the parameters related to one experiment task, which form a spatial map that indicates the brain areas activated by the certain task. The brain image recorded at one time point consists of 10 slices and there are 30×30 voxels on each slice. The brain image is recorded 120 times, and hence each column of $\mathbf{X}$ is a time series of length 120. We "simulated" the time components using a recorded blood flow from a real fMRI study (Section 5.4). The length of the blood flow triggered by the task spans 30 time points ($\tilde{N} = 30$). All the four experimental components share the same shape of the hemodynamic response. However, the nonzero response for each component starts at different times since, in our experiment, the subject performs the four tasks sequentially with some rest period in between.

All the voxels in $\boldsymbol{\beta}$ are given a numerical value of either 0 or 1. In each spatial component, the voxels with value 1 correspond to the regions that are activated by the corresponding experiment task, and they are plotted as dark red areas in Figure 5.3. The last time component is not fully observed in this experiment because the fMRI scan stopped right after the subject finished the last task. Due to the delay of the blood flow
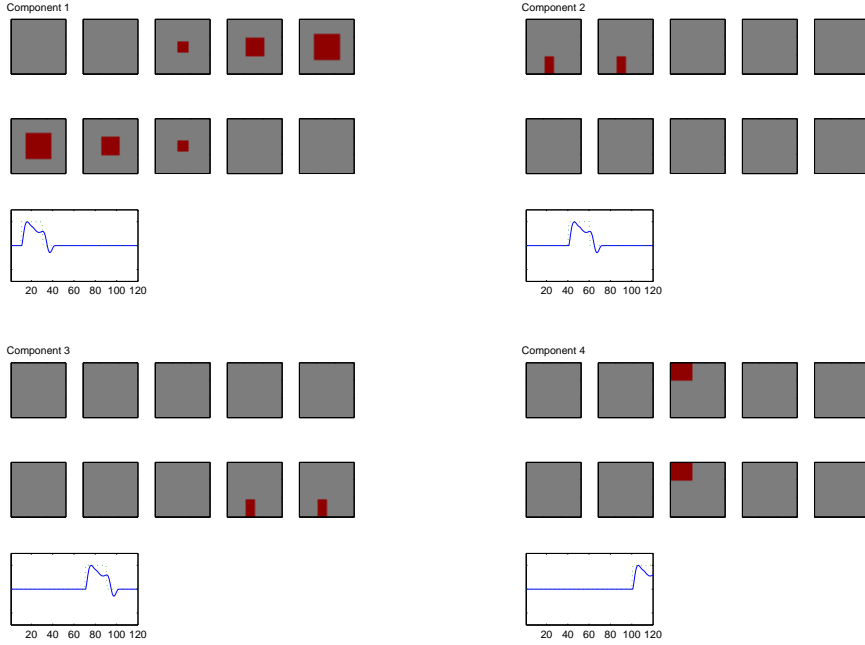
Figure 5.3: *The four components generated in the simulation. In each panel, the first 10 images are the spatial maps (one row of $\boldsymbol{\beta}$), and the dark red areas stand for activated voxels. The solid line in the subsequent plot shows the temporal characteristic of the activated voxels (one column of $\mathbf{X}$). The dotted line stands for the experiment stimulus, 0 meaning "rest" and 1 meaning "active".*

caused by hemodynamic response, we can only observe part of the blood flow in this case.

5.3.2   Analysis and Results

We then applied both our proposed approach and standard SPM on the simulated data. As mentioned in Section 5.3.1, there are four different conditions involved in the experiment ($K = 4$). Hence there are five columns in the design matrix $\mathbf{X}$, where the first column is a vector of 1. The length of $\tilde{\mathbf{X}}$, that is, the length of the blood flow triggered by the tasks, is set to be 30 ($\tilde{N} = 30$). The initial design matrix was obtained from SPM5. Namely, each column of $\mathbf{X}$, except for the first one, was obtained by convolving the corresponding stimulus sequence with the canonical HRF (Figure 5.2). Starting from this initial design matrix, our iterative algorithm converged after 10 iterations. The GCV

criterion selected λ to be 10^3 from a set of candidate points of $\{0, 10^1, 10^2, \ldots, 10^{10}\}$. We also applied the algorithm using a different initial design matrix, where $\tilde{\mathbf{X}}$ is replaced by a vector of 1. The algorithm converged to the same results after 15 iterations. The results are shown in the first column of Figure 5.4.

All the four components were recovered reasonably well, although some noise does exist due to the random noise added to the GLM model. Similar to Figure 5.3, dark red areas indicate the activated voxels. The time course plot in each panel shows the corresponding time component (solid lines). The estimated time components are very close to the real ones as in Figure 5.3.

The second column of Figure 5.4 shows the results from standard SPM, using the design matrix obtained in SPM5. That is, we obtained the spatial maps by $\boldsymbol{\beta} = (\breve{\mathbf{X}}^T \breve{\mathbf{X}})^{-1} \breve{\mathbf{X}}^T \mathbf{Y}$, where $\breve{\mathbf{X}}$ is the design matrix used by SPM5. The time component modeled by the canonical HRF is shown in the last plot (solid line) in each panel. The shape of the time component is far away from the real ones. It's also obvious that the activation areas shown by the spatial maps are much more noisy than the activation indicated by our proposed method.

5.4 A Real fMRI Data Analysis

5.4.1 Experiment Paradigm and Data Description

To study brain regions that are related to different finger tapping movements, an fMRI data set was obtained from one human subject performing five different tasks alternately: rest, externally-guided (EG) right-hand movement, externally-guided left-hand movement, internally-guided (IG) right-hand movement and internally-guided left-hand movement. Each rest period lasts 30 seconds and each activation period lasts 60 seconds. Here EG movement means the subject did the finger tapping following video instructions, while IG movement means the subject needed to finish the task according to their memory. The experimental paradigm is shown in Figure 5.5. When acquiring
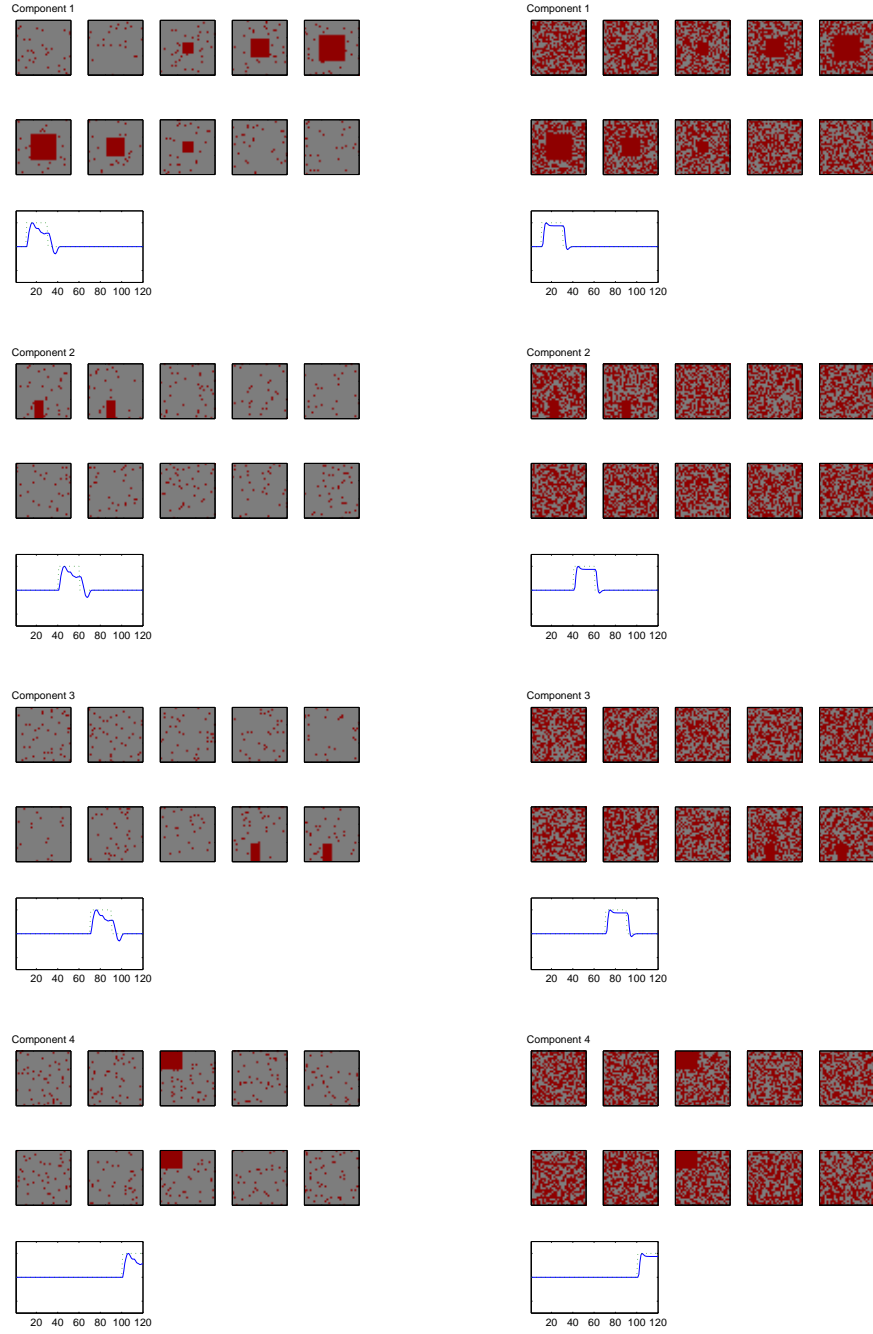
Figure 5.4: *Comparison of the results from the proposed adaptive SPM approach (the left column) and standard SPM (the right column).*
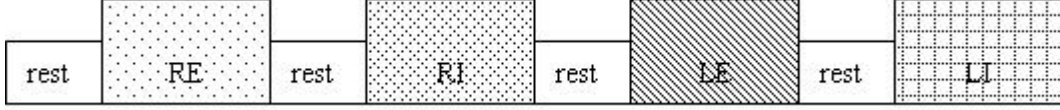
Figure 5.5: *The experimental design used in acquiring the fMRI data. RE: EG Right-hand; RI: IG Right-hand; LE: EG Left-hand; LI: IG Left-hand. Each rest block took* 30 *seconds (*10 *scans when TR* = 3 *seconds). Each activation block took* 60 *seconds (*20 *scans).*

the data, the subject repeated this paradigm sequence twice.

During the experiment, 240 MR scans were acquired on a modified 3T Siemens MAG-NETOM Vision system. Each acquisition consisted of 46 contiguous slices. Each slice contained 53×63 voxels. Hence there were $53 \times 63 \times 46$ voxels from each scan. The size of each voxel is 3mm $\times$ 3mm $\times$ 3mm. Each acquisition took 2.9388 seconds, with the scan to scan repetition time (TR) set to be 3 seconds.

5.4.2 Analysis and Results

The goal of this fMRI study is to detect which brain regions are responsible for the four different finger tapping movements, EG right-hand (RE), IG right-hand (RI), EG left-hand (LE) and IG left-hand (LI).

The data set was preprocessed using SPM5. The preprocessing included realignment, coregistration, segmentation, spatial normalization and smoothing.

We than analyzed the preprocessed data set using both SPM5 with the canonical HRF and our proposed method with the initial design matrix obtained from SPM5.

The results from the two methods are displayed in the corresponding panels of Figure 5.6. As mentioned earlier, the goal of this study is to identify the brain regions activated by the four different finger movements. We only showed the four related brain slices for each movement according to the classic pattern mentioned in Section 3.5.

The upper panel of Figure 5.6 shows the four components detected by adaptive SPM. The lower panel presents the four components recovered by general SPM. Within each row, the four image slices represent the activated spatial maps. The red areas illus-

trate activated brain regions. Brighter color indicates higher intensity. Both methods demonstrate the classic brain activation patterns during hand movement as mentioned in Section 3.5. But our method shows dominantly higher intensity and less noise for each component. In addition, the subject is supposed to use more SMA and less cerebellum when doing IG movement. This hypothesis is proven to be true using our method.

Figure 5.7 presents the time component estimated by both methods. The solid line in each plot is the estimated time component and the dotted line is the stimulus sequence, where zero means "rest" and one means "activation". Because all the four components are assumed to share the same shape of blood flow, we only plot one time component here for each method. The time component estimated by our method has a shape that is much closer to reality.

5.5   Discussion

In this chapter, we introduced an adaptive SPM method to estimate the time components related to the fMRI experiment stimulus nonparametrically, which leads to better detection of brain areas activated by the stimulus. The motivation is the fact that the detection power of standard SPM approaches is constrained by usually predefined HRF. The idea of our proposed method is to start from a design matrix $\mathbf{X}$ modeled by the canonical HRF and then apply a penalized smoothing technique to refine the time component ($\mathbf{X}$) and spatial component in an iterative way. We make use of the timing information of the blood flow offered by the initial design matrix and adjust the shape of the blood flow of the activated voxels to be closer to the reality. We illustrated the performance of our method through a comparative study on both simulated data and a real fMRI data set. Our method gives better results in both cases. In addition, our method is easy to implement and the computation is fast.

I: Adaptive SPM

RE

RI

LE

LI

II: General SPM

RE

RI

LE

LI
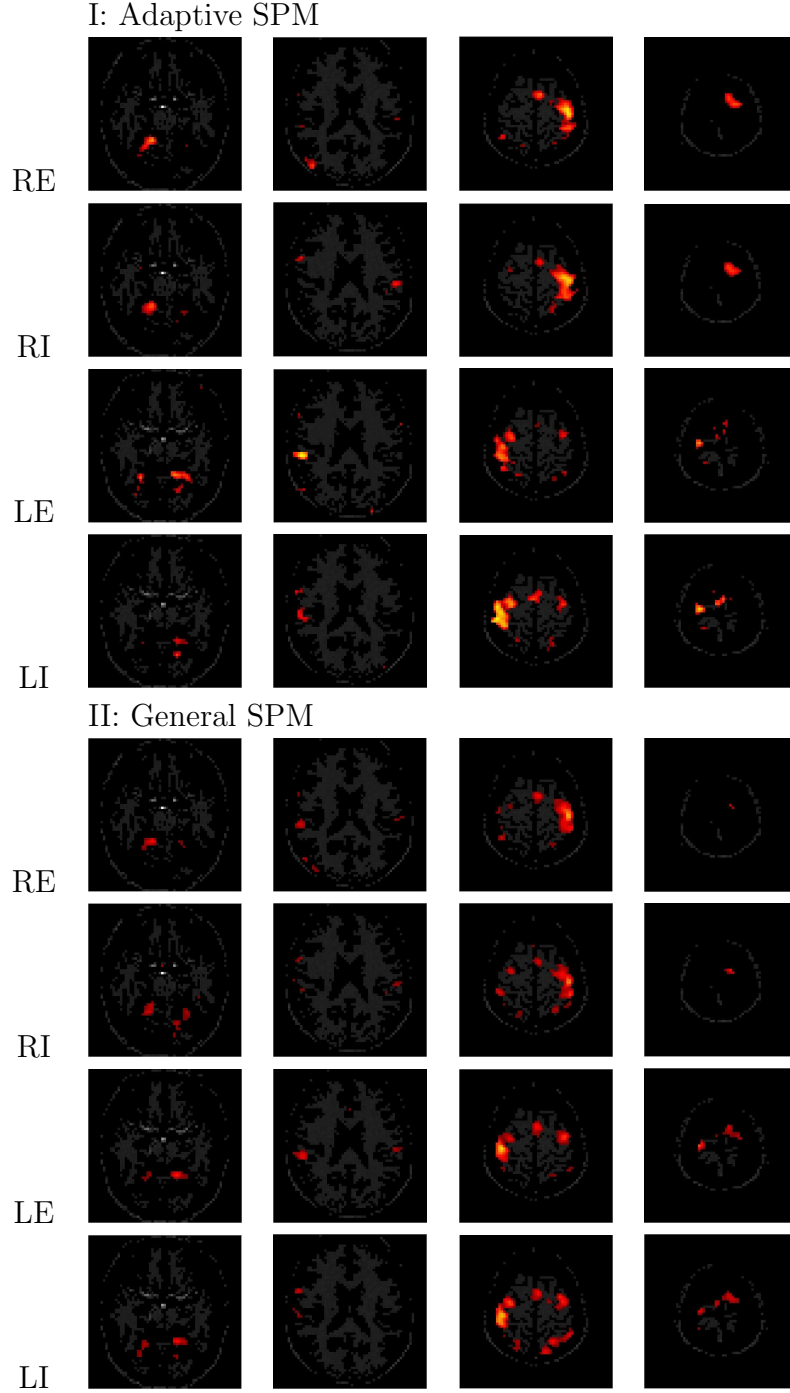
Figure 5.6: *Brain regions activated by the four finger movements detected by two methods, Panel (I): Adaptive SPM, EG right-hand (first row), IG right-hand (second row), EG left-hand (third row) and IG left-hand (fourth row); Panel (II): General SPM, EG right-hand (first row), IG right-hand (second row), EG left-hand (third row) and IG left-hand (fourth row). Within each row, the first slice shows the primary motor cortex (PMC), the second slice contains both PMC and supplementary motor area (SMA), the third slice shows basal ganglia and the fourth slice shows cerebellum. Red areas illustrate the activated voxels. Brighter color indicates higher intensity.*
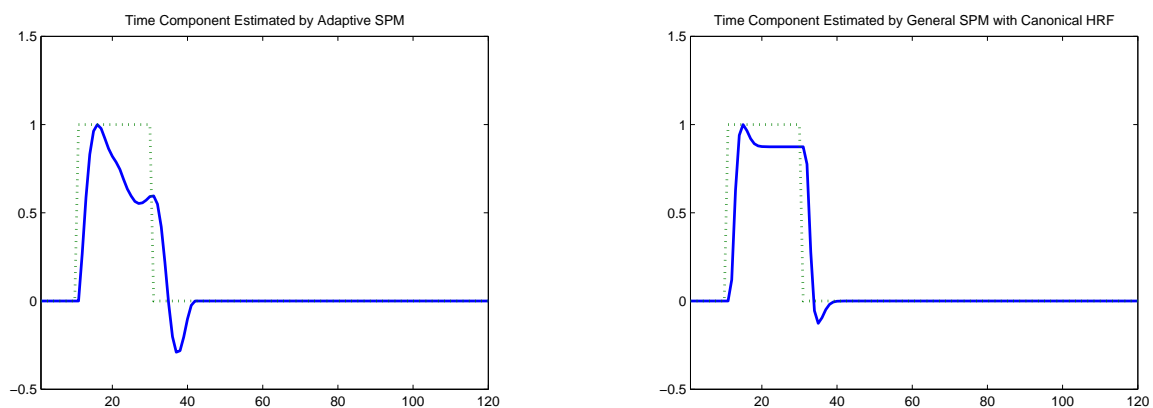
Figure 5.7: *The time components (solid lines) estimated by two methods with the stimulus sequence (dotted lines) overlayed. Left: Adaptive SPM; Right: General SPM with canonical HRF.*

CHAPTER 6

A Novel Method for Event-Related fMRI Analysis

6.1   Introduction

Because of the good estimation power and experimental flexibility they provide, event-related experiment designs have become more and more popular in fMRI studies. In event-related fMRI (ER-fMRI), stimuli (or events) are applied for short bursts in a stochastic manner. The recorded BOLD fMRI signals measure transient changes in brain activity associated with discrete events. This feature makes ER-fMRI a powerful tool to estimate the change in the MR signal triggered by neuronal activity, which is known as the hemodynamic response.

The basis for ER-fMRI is that the changes in hemodynamics are rapid and occur within seconds after a neuronal event. It's shown that even a stimulus duration of as little as 2 seconds could produce detectable signal changes (Rosen *et al.*, 1998). The fact that fMRI is sensitive to transient signal change to brief neuronal events makes it possible to model the timing of hemodynamic response by single-trial events, which is not achievable in traditional block-design fMRI.

It is hemodynamic response to the underlying neuronal activity that makes the fMRI signal in areas of activation a blurred and delayed version of the stimulus sequence. Figure 6.1 shows recorded BOLD signals (solid line) triggered by a single event (dashed line on the left panel) and a typical block-design sequence (dashed line on the right panel) respectively. Both of them show the blur-and-delay effect caused by the hemodynamic
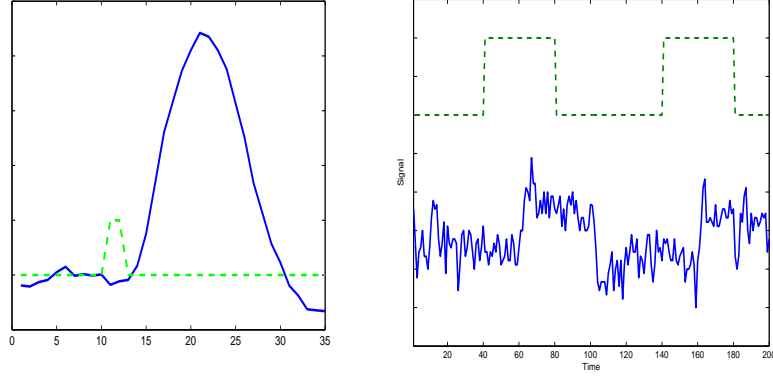
Figure 6.1: *Left Panel: The recorded BOLD signal (solid line) triggered by a single event (dashed line). Right Panel: The recorded BOLD signal (solid line) triggered by a typical block-design sequence (dashed line).*

response.

In common practice, the blur-and-delay effect is modeled through the convolution of the stimulus sequence $(X(\cdot))$ and a hemodynamic response function (HRF, $h(\cdot)$) as

$$\text{BOLD}(t) = h \otimes X(t) = \int h(t-u)X(u)du. \tag{6.1}$$

The basis of the above model is the so-called linearity of BOLD fMRI responses when multiple stimuli are presented in succession. This linearity property was firstly studied by Boynton *et al.* (1996). The authors studied the linear transform model for BOLD fMRI, which indicates that fMRI response is a linear transform of the neural activity. A basic framework of this linear transform model is shown in Figure 6.2.

According to this framework, an experiment stimulus induces the neural activity in a specific region of the brain. The neural activity then brings blood flow changes in that region of the brain, while BOLD fMRI responses are measured from these blood flow changes. In addition to giving this clear picture of how BOLD fMRI works, the linear transform model in Figure 6.2 is important in two respects. Firstly, the linearity of the fMRI response and neural activity makes it possible to determine changes in
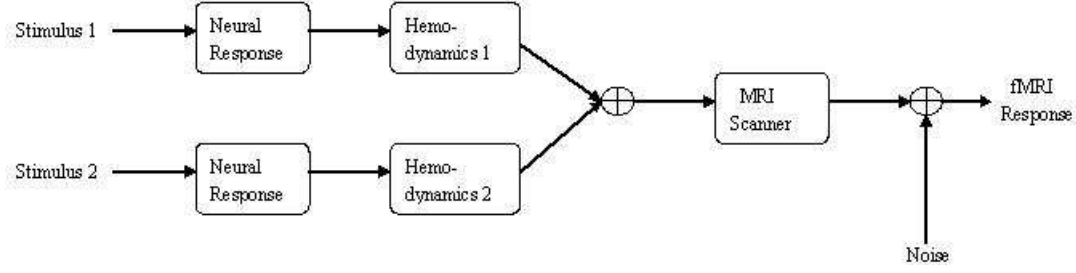
Figure 6.2: *The basic framework of the linear transform model. fMRI response is a linear transform of the neural activity. Adapted from Boynton* et al. *(1996).*

neural activity by the amplitude changes in hemodynamic response. Secondly, this linear transform model also shows that when multiple stimuli are presented in succession, the hemodynamic response would be the summation of the individual responses generated by the single stimulus respectively.

Dale and Buckner (1997) reconfirmed this linearity property. In their study, the authors investigated clusters of one, two or three stimuli at interstimulus intervals of either 2 or 5 seconds. By subtracting the single-stimulus response from that to a pair of stimuli, the resulted response is supposed to be the same as the single-stimulus response. Similarly, the response to the third one can be obtained by subtracting the response to the pair of stimuli from that to the triple of stimuli. The detected responses to the one, two and three stimuli are respectively shown on the left panel of Figure 6.3. The single-stimulus responses from subtraction are on the right panel. These pictures give evidence of the linearity.

This linear transform model of fMRI response is the basis for most model-driven fMRI studies in the literature (Friston *et al.*, 1994; Lange and Zeger, 1997; Friston *et al.*, 1995b; Josephs *et al.*, 1997; Woolrich *et al.*, 2004; Lindquist and Wager, 2007). Based on this linearity property, BOLD signals are commonly modeled as the convolution of the stimulus sequence and a certain form of HRF (Equation (6.1)). The form of the HRF is either specified or estimated through a set of parameters. There has been a big body of literature on the HRF modeling. Most of them are parametric methods. The
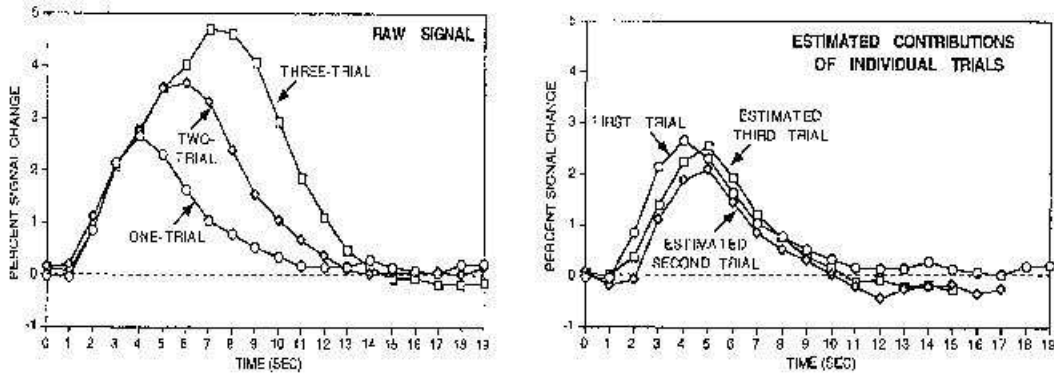
Figure 6.3: *Linear addition of hemodynamic responses to individual stimulus events. Adapted from Dale and Buckner (1997).*

assumptions made about the shape of the HRF vary among different methods. Some of them are very stringent, while some of them are relatively more flexible.

As an early study of ER-fMRI, Dale and Buckner (1997) gives a completely fixed form of HRF to every voxel. Friston *et al.* (1994) also assumes the HRF as a Poisson function with fixed parameter. Lange and Zeger (1997) uses a two-parameter gamma function to model the HRF. The two parameters vary at different voxels and hence the estimated HRF varies from voxel to voxel. These methods all give very limited, if any, flexibility of the HRFs and hence cause biases of estimation. To allow the variation of HRFs at different voxels of the brain, Genovese (2000) and Gössl *et al.* (2001b) propose to model the HRF and BOLD response voxel-wise in a Bayesian framework. In their work, the HRF is defined by certain number of parameters and prior distributions are given to each parameter. Inferences of the parameters are then made at each voxel using Markov Chain Monte Carlo (MCMC) technique. The disadvantage of these methods is the slow performance of general MCMC techniques for the inferences.

Friston *et al.* (1995a) is one of the first ER-fMRI studies in the GLM framework, and the form of the HRF is estimated by two given temporal basis functions. Josephs *et al.* (1997) extended this idea and more basis functions are included to give better flexibility. These are very important work in the sense that the basis sets allow one

to estimate an HRF of arbitrary shape for different events at different voxels of the brain, and at the same time the inferences can be easily made. Many studies on the HRF modeling are focused on the refinement and improvement of the idea of using basis sets ever since. For example, Woolrich *et al.* (2004) introduced a technique to apply some constraints to avoid nonsensical HRF, which is a big problem when using simple basis functions. More recently, Lindquist and Wager (2007) proposed another method, using three superimposed inverse logistic functions, to model the HRF. This work also summarizes some currently most popular HRF modeling techniques, such as smooth finite impulse response (FIR) filter (Goutte *et al.*, 2000), canonical HRF with time and dispersion derivatives (Caulhoun *et al.*, 2004) and the canonical SPM HRF (Friston *et al.*, 1998).

The limitations of these methods reside in two aspects. First of all, the linear transform model of fMRI response is only tested empirically and some recent studies even show some non-linearities in fMRI responses. Secondly, all the above methods are in a parametric framework, where the choice of parametric forms naturally introduces biases of the estimation. When the selected model contains less parameters, it cannot capture the variation of the HRFs across the brain. On the other hand, when more parameters are included in the model, they tend to introduce more estimation error and less estimation power.

To overcome the analysis challenges caused by these limitations, we propose a novel regression approach to estimate the HRF directly. The approach is based on point processes modeling to account for the event-related designs.

A point process is a special kind of stochastic process. It measures random collections of point occurrences in a certain time interval. Point processes can be used for modeling identification problems emerged in, e.g., neurophysiology as in (Brillinger, 1975). In the literature, point processes are modeled to identify the relationship between the so-called input process X and the output process Y. This relationship indicates how the input

influences the output by estimating related parameters (Brillinger, 1974).

In ER-fMRI, experiment stimuli are applied to subjects for short burst in a random manner. By analyzing the recorded fMRI signals, we want to estimate the shape of the HRFs associated with different stimuli. Taking the stimulus sequence as the input process and the recorded fMRI signal as the output, we can use the point process theory to identify the relationship between the stimuli and fMRI signals.

Compared to the existing methods, the proposed procedure yields simultaneously the nonparametric estimate of the HRF and a test for the linearity assumption. To illustrate its usefulness and the scientific implications, we applied this procedure to study the spatial variation of the HRF, and the extent to which the linear relationship holds in various regions of interest for Parkinson's disease patients.

The rest of this chapter is organized as follows. Section 6.2 introduces the point process theory. In Section 6.3, we apply the point process theory to ER-fMRI study. Section 6.4 illustrates the performance of our proposed method through a simple real data analysis. For the purpose of comparison, a more complicated ER-fMRI data set is analyzed using both the proposed method as well as an existing popular ER-fMRI analyzer in Section 6.5. Concluding remarks are given in Section 6.6 to close the chapter.

6.2 Point Processes

A point process X is a special kind of stochastic process. It measures random collections of point occurrences. Given a time interval A in practice, $X(A)$ measures the number of occurrences in A of some event of interest. Typical examples of point processes include Poisson processes and renewal processes (Cox and Isham, 1980).

The same as ordinary stochastic processes, a point process is said to be stationary when the probability properties are invariant under translations of the arbitrary interval A. Orderliness is a special property for point processes. A point process is said to be orderly if $P\{X(t, t+\delta] > 1\} = o(\delta), \forall t \in \mathcal{R}$, as $\delta \to 0+$, that is, the process X doesn't

have any multiple simultaneous occurrences. In the following of this chapter, we focus on the analysis of stationary orderly point processes.

6.2.1 Point Process Parameters and Spectral Properties

Let X and Y be two stationary point processes with orderliness. We define $dX(t) = X(t, t + dt]$ and similarly $dY(t) = Y(t, t + dt]$. Then the orderliness implies that $P\{dX(t) > 1\} = o(dt), t \in \mathcal{R}$ and $P\{dY(t) > 1\} = o(dt), t \in \mathcal{R}$. We now define some parameters that describe point processes.

The rate of the process X is defined as

$$p_X = \lim_{dt \to 0+} \frac{E(dX(t))}{dt} = \lim_{dt \to 0+} \frac{E(X(t, t + dt])}{dt}. \tag{6.2}$$

Due to the assumed orderliness, (6.2) can be interpreted as

$$P\{X \text{ points in } (t, t + dt]\} = p_X dt.$$

The rate of the process Y is defined in the same way. In the second-order case, one defines the second-order product density as

$$p_{XY}(\tau) = \lim_{dt, d\tau \to 0} \frac{E(dY(t + \tau)dX(t))}{dt d\tau}, \quad \tau \neq 0, \tag{6.3}$$

with the interpretation that

$$P\{X \text{ points in } (t, t + dt] \wedge Y \text{ points in } (t + \tau, t + \tau + d\tau]\} = p_{XY}(\tau)dt d\tau, \quad \tau \neq 0.$$

p_{XX} and p_{YY} can be defined similarly.

Hence, the covariance density of the processes can be defined by

$$q_{XY}(\tau) = p_{XY}(\tau) - p_X p_Y, \quad \tau \neq 0, \tag{6.4}$$

68

with the interpretation given by

$$dC_{XY}(\tau)dt = \text{Cov}\{dY(t+\tau), dX(t)\} = q_{XY}dtd\tau.$$

But in the case of dC_{XX} and dC_{YY}, we have to note that $E\{[dX(t)]^2\} = E\{dX(t)\}$, hence the covariances are

$$
\begin{aligned}
dC_{XX}(\tau)dt &= \text{Cov}\{dX(t+\tau), dX(t)\} = (q_{XX}(\tau) + \delta(\tau)p_X)dtd\tau, \\
dC_{YY}(\tau)dt &= \text{Cov}\{dY(t+\tau), dY(t)\} = (q_{YY}(\tau) + \delta(\tau)p_Y)dtd\tau,
\end{aligned}
$$

where $\delta(\cdot)$ is the Dirac delta function.

Based on the above parameters, the conditional intensity of the processes is then defined by

$$E\{dY(t)|X(t) = 1\} = p_{XY}(t)dt/p_X, \tag{6.5}$$

which can be interpreted as

$$P\{Y \text{ points in } (t, t+dt] \mid X \text{ events at } t\} = p_{XY}(t)dt/p_X.$$

According to Bartlett (1963), the cross-spectrum of the two processes at frequency λ, $f_{XY}(\lambda)$, is defined by

$$f_{XY}(\lambda) = \frac{1}{2\pi} \int \exp(-iu\lambda)dC_{XY}(u) = \frac{1}{2\pi} \int \exp(-iu\lambda)q_{XY}(u)du. \tag{6.6}$$

The spectrum of the process X at frequency λ is

$$f_{XX}(\lambda) = \frac{1}{2\pi} \int \exp(-iu\lambda)dC_{XX}(u) = \frac{1}{2\pi}p_X + \frac{1}{2\pi} \int \exp(-iu\lambda)q_{XX}(u)du, \tag{6.7}$$

with a similar definition for Y.

6.2.2 Linear Systems

To study the relationship between two processes, we consider the following linear regression model (Brillinger, 1975):

$$E(dY(t) \mid X) = \left(s_0 + \int s_1(t-u)dX(u) \right) dt; \tag{6.8}$$

that is,

$$P\{Y \text{ events in } (t, t+dt] \mid X\} = \left(s_0 + \int s_1(t-u)dX(u) \right) dt, \tag{6.9}$$

where $s_1(t)$ is called the *average impulse response*. It represents the effect on the intensity of Y with X eventing at time 0.

Equation (6.8) then leads to the following results:

Result 1: *Let $S_1(\cdot)$ be the Fourier transform of $s_1(\cdot)$, then*

$$
\begin{aligned}
p_Y &= s_0 + p_X \int s_1(u)du, \tag{6.10} \\
p_{YX}(t) &= s_0 p_X + s_1(t)p_X + \int s_1(t-u)p_{XX}(u)du, \tag{6.11} \\
q_{YX}(t) &= s_1(t)p_X + \int s_1(t-u)q_{XX}(u)du, \tag{6.12} \\
f_{YX}(\lambda) &= S_1(\lambda)f_{XX}(\lambda). \tag{6.13}
\end{aligned}
$$

Proof of Result 1 is given in Appendix B.

An important statistic for studying the linear system is the coherency defined by

$$W_{XY}(\lambda) = \left| \frac{f_{XY}(\lambda)}{\sqrt{f_{XX}(\lambda)f_{YY}(\lambda)}} \right|. \tag{6.14}$$

Regressing one process Y on the other X, the squared coherency at frequency λ is the so-called multiple correlation coefficient. When the processes are linearly related as in (6.8), $W_{XY}(\lambda) = 1$ for $0 < \lambda < 1$.

6.2.3 Parameter Estimation and Inference

Result 1 suggests the following estimates of $S_1(\cdot)$ and $W_{XY}(\cdot)$:

$$\hat{S}_1 = \hat{f}_{YX}(\hat{f}_{XX})^{-1} \tag{6.15}$$

and

$$\hat{W}_{XY}(\lambda) = \left| \frac{\hat{f}_{XY}(\lambda)}{\sqrt{\hat{f}_{XX}(\lambda)\hat{f}_{YY}(\lambda)}} \right|, \tag{6.16}$$

where $\hat{f}_{XY}(\cdot)$, $\hat{f}_{XX}(\cdot)$ and $\hat{f}_{YY}(\cdot)$ are the smooth estimates of f_{XY}, f_{XX} and f_{YY} respectively. These estimates are described in the following as given in Brillinger (1974).

Let $d_X(\cdot)$ be the discrete Fourier transformation of X:

$$d_X(\lambda) = \sum_{t=1}^{T} \exp(-i\lambda t)X(t),$$

where T is the total number of time points in the process. The discrete Fourier transformation of Y $d_Y(\cdot)$ is defined similarly. And the periodogram of X and Y are then given as

$$I_{XY}(\lambda) = (2\pi T)^{-1}d_X(\lambda)d_Y(\lambda)^*,$$

where a^* is the conjugation of a. $I_{YX}(\cdot)$, $I_{XX}(\cdot)$ and $I_{YY}(\cdot)$ are defined similarly.

Suppose $K(\cdot)$ is a weight function, then the estimates of f_{XY}, f_{XX} and f_{YY} are given by the following respectively:

$$
\begin{aligned}
\hat{f}_{XY}(\lambda) &= 2\pi(bT)^{-1}\sum_{s\neq0} K(b^{-1}(\lambda - 2\pi s/T))I_{XY}(2\pi s/T), \\
\hat{f}_{XX}(\lambda) &= 2\pi(bT)^{-1}\sum_{s\neq0} K(b^{-1}(\lambda - 2\pi s/T))I_{XX}(2\pi s/T), \\
\hat{f}_{YY}(\lambda) &= 2\pi(bT)^{-1}\sum_{s\neq0} K(b^{-1}(\lambda - 2\pi s/T))I_{YY}(2\pi s/T),
\end{aligned}
$$

where b is the smoothing parameter.

Now if we set

$$f_{\epsilon\epsilon}(\lambda) = f_{XX}(\lambda) - \frac{|f_{XY}(\lambda)|^2}{f_{YY}(\lambda)} = f_{XX}(\lambda)\left(1 - \frac{|f_{XY}(\lambda)|^2}{f_{XX}(\lambda)f_{YY}(\lambda)}\right) = f_{XX}(\lambda)(1 - W_{XY}^2(\lambda)),$$

it can be shown that under certain conditions, the estimate $\hat{S}_1(\lambda)$ is asymptotically complex normal with mean $S_1(\lambda)$ and variance $2\pi(bT)^{-1}\int K(\alpha)^2 d\alpha f_{\epsilon\epsilon}(\lambda)f_{XX}(\lambda)^{-1}$. And $\hat{S}_1(\lambda_1)$, $\hat{S}_1(\lambda_2)$, ..., $\hat{S}_1(\lambda_J)$ are asymptotically independent normal for distinct λ_1, λ_2, ..., λ_J (Brillinger, 1974).

The estimate of the impulse response function $s_1(\cdot)$ is then given by

$$\hat{s}_1(u) = \left[\sum_{q=-Q_T}^{Q_T} \hat{S}_1(C_T q)\exp(iuC_T q)\right]\left[(1 - \cos C_T u)/(\pi C_T u^2)\right],$$

for some small C_T and large Q_T. Under certain conditions, $(\hat{s}_1(u_1), \hat{s}_1(u_2), \ldots, \hat{s}_1(u_J))$ is asymptotically normal with mean $(s_1(u_1), s_1(u_2), \ldots, s_1(u_J))$ and covariance matrix

$$C_T(bT)^{-1}\int K(\alpha)^2 d\alpha \int \exp\{i(u_j - u_k)\alpha\}f_{\epsilon\epsilon}(\alpha)f_{YY}(\alpha)^{-1}d\alpha(2\pi)^{-1}, \quad j,k = 1,2,\ldots J.$$

On the other hand, the estimate of the coherence $W_{XY}(\cdot)$ is given by

$$\hat{W}_{XY}(\lambda) = \frac{\hat{f}_{XY}(\lambda)}{\sqrt{\hat{f}_{XX}(\lambda)\hat{f}_{YY}(\lambda)}}.$$

The coherence lies between 0 and 1, the closer to 1 the stronger the linear relationship between the processes.

Under certain conditions, $\hat{W}_{XY}(\lambda)$ is asymptotically normal with mean $W_{XY}(\lambda)$ and variance proportional to constant $\frac{1-W_{XY}^2(\lambda)}{Tb}$. Moreover, if $W_{XY}(\lambda) = 0$, then

$$F(\lambda) = \frac{c|\hat{W}_{XY}(\lambda)|^2}{1 - |\hat{W}_{XY}(\lambda)|^2} \sim F_{2,2c}, \quad \text{where} \quad c = (bT/\gamma) - 1 \text{ and } \gamma = \sum \lambda. \qquad (6.17)$$

This result can be used to test for a response to the stimulus by computing a test statistic for significant activation $F(\lambda_\alpha)$ at the fundamental frequency of activation λ_α. Under the null hypothesis of no activation, the F-statistic at the fundamental frequency of activation, $F(\lambda_\alpha)$, has a F distribution with 2 and $2c$ degrees of freedom. Large values of $F(\lambda_\alpha)$ indicate a large effect at the fundamental frequency. The theoretical justification of the above results is relegated to Appendix C.

6.3　A Novel Method for Event-Related fMRI Analysis

As mentioned before, in ER-fMRI, experiment stimuli are applied to subjects for short bursts in a random manner. By analyzing the recorded fMRI signals, we want to estimate the shape of the HRF associated with different stimuli. The linear property of fMRI system mentioned earlier motivated us to apply the point process technique in ER-fMRI analysis.

Let X be the stimulus sequence, Y the recorded fMRI signals at a single voxel and $h(\cdot)$ the corresponding HRF. Then Y contains the BOLD signal of the voxel plus random noise. According to the linear transform model for BOLD fMRI (Equation (6.8)), we have

$$E(Y(t)) = \text{BOLD}(t) = h \otimes X(t) = \int h(t-u)X(u)du. \qquad (6.18)$$

Here $h(\cdot)$ represents the influence of the stimulus on the BOLD signal, which has the same meaning as the $s_1(\cdot)$ in (6.8) and (6.9).

Our goal in this study is then to estimate the form of $h(\cdot)$ and test the linear relationship (6.18). Suppose there is only one event in the stimulus sequence and this event happens in a random manner. Then the stimulus function $X(\cdot)$ has only two values, 1 when there is an event and 0 when no event. If $h(\cdot)$ is the HRF, then we have

$$\text{BOLD}(t) = h \otimes X(t) = \int h(t-u)X(u)du. \qquad (6.19)$$

As in Section 6.2.2, by estimating the Fourier transform of $h(\cdot)$, $H(\alpha) = \int h(u)\exp(-i\alpha u)du$, we can get the estimation of desired HRF.

In the most general case, there would be random number of events in the stimulus sequence. Let X and Y be stationary point processes, we propose to consider the linear system for studying the input-response relationship:

$$E\big(dY(t) \mid X\big) = \big(\mu + \int s(t-u)dX(u)\big)dt. \tag{6.20}$$

As a specific application, let $X(t) = \{\tau_0, \tau_1, \ldots\}$ denote the times of event (stimulus) and $Y(t) = \{\tau_0 + \gamma_0, \tau_1 + \gamma_1, \ldots\}$ be the times leaving the peak of the response, with γ_j being the time to reach the peak following the stimulus. We may consider

$$dY(t) = \sum_j \delta(t - \tau_j - \gamma_j)dt,$$

where $\delta(\cdot)$ is the Dirac delta function. In the case that times γ_j reaching the peaks are independent of the stimulus $X(\cdot)$ and have the same marginal density function $g(\cdot)$, then

$$
\begin{aligned}
E(dY(t)|X) &= \int \sum_j \delta(t - \tau_j - \gamma)dt \cdot g(\gamma)d\gamma \\
&= \sum_j \int \delta(t - \tau_j - u)g(u)dudt \\
&= \sum_j g(t - \tau_j)dt \\
&= \big(\int g(t-u)dX(u)\big)dt.
\end{aligned}
$$

This is useful and important for studying the variation of the peak times because an estimate of $g(\cdot)$ would provide information about their distribution. Similarly we can use this approach to examine the variability of other important response times following the event of interest. For example, Figure 6.4 gives an empirical shape of the hemodynamic response to a single event stimulus. We can approximately divide it into four stages,
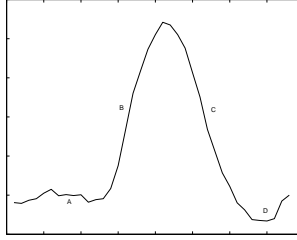
74

Figure 6.4: *An empirical shape of the hemodynamic response in fMRI to a single event stimulus. The four stages of the hemodynamic response are: A: lag-on; B: rise; C: decay; D: dip.*

which are lag-on, rise, decay and dip as shown in the graph. By defining and estimating the time limits for each of the stages, we can estimate the shape of the HRF generally.

6.4 A Simple Real Data Analysis

6.4.1 Data Description

To illustrate the performance of our new method , we applied our method on a simple data set taken from Huettel *et al.* (2004). This data set contains averaged epochs of 35 time points collected using fMRI at 4.0T. The time courses are averaged from regions of interest in visual cortex within a set of 11 subjects. The experimental design presented a single static checkerboard for 1500ms. The inter-scan time (TR) was 500ms. The epoch consists of time points from 5s before stimulus onset through 12s after stimulus onset. The values are expressed in proportional signal change, i.e., 0.01 corresponds to a 1% change. Figure 6.5 plots the 35 observations of this data set.

6.4.2 Analysis and Results

Before the analysis, we repeated the 35 observations twice, that is, "repeated" the experiment twice. Hence we obtained a bigger data set with 70 time points, as shown in the top panel in Figure 6.6.

In addition to applying our proposed method to the revised data set, we also implemented the basis function modeling of Friston *et al.* (1995b). In their method, the HRF
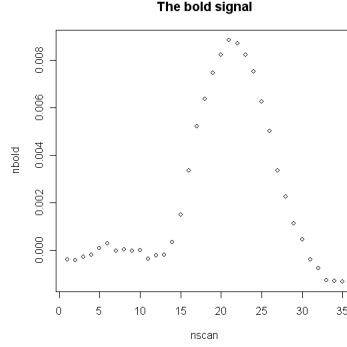
Figure 6.5: *An example dataset from a simple event-related fMRI experiment. Within this time course: From time points $1 - 10$, the subject was watching a dark display. At time points 11, a checkerboard pattern was presented $1500ms$. After the offset of the checkerboard until time point 35, the subject was watching a dark display.*

is modeled using two basis functions

$$f(t) = \sin(\pi t/(n+1)) \cdot \exp(-t/(4n)),$$

$$f(t) = \sin(\pi t/(n+1)) \cdot \exp(t/n),$$

where n is the number of scans in the experiment, which is 70 for our data.

Applying our method to this revised data set, we obtained the estimated HRF which is plotted in the middle panel in Figure 6.6. By convolving the estimated HRF with the stimulus we obtained the predicted BOLD signal, which is the solid line in the bottom panel in Figure 6.6. The perfect agreement between the predicted BOLD and the observed signals is shown in the bottom panel in Figure 6.6.

In addition, Figures 6.7 and 6.8 show the comparison between our method and the basis function modeling method from Friston *et al.* (1995b). The estimated HRF using our method appears to have the typical shape of a HRF more than the one using basis function modeling does. Moreover, the predicted BOLD signal by convolving the stimulus sequence with the estimated HRF using our method apparently agrees with the recorded signals much more than the one with basis function modeling.
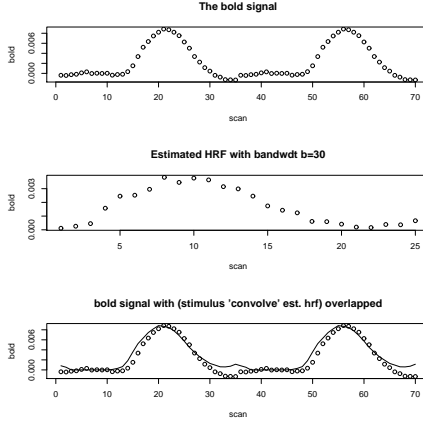
Figure 6.6: *Top: the recorded BOLD signal; Middle: the estimated HRF by applying the new method; Bottom: dotted line – the recorded BOLD signal, solid line – the predicted BOLD signal by convolving the stimulus with the estimated HRF.*

On the other hand, we also estimated the coherence statistic defined by (6.14). At the fundamental frequency, $\hat{W}_{XY} = 0.99556$ with a F-statistic equal to 90.953. These values support the linear relationship between the stimulus sequence and the fMRI signals.

6.5 A Second Real Data Analysis

6.5.1 Experiment Paradigm and Data Description

In this study, an fMRI data set was obtained from one human subject performing a predefined event sequence as visually instructed. The stimulus sequence includes two different events: right-hand and left-hand finger tapping. Each finger tapping movement last around 1 second. The order of the sequence was predefined in a random way. To avoid the overlapping of consecutive events, the time interval between two successive events was randomly selected from Uniform$[18, 22]$. The experiment paradigm is shown in Figure 6.9.

During the experiment, 47 MR scans were acquired on a modified 3T Siemens MAG-NETOM Vision system. Each acquisition consisted of 49 contiguous slices. Each slice contained 64×64 voxels. Hence there were $64 \times 64 \times 49$ voxels from each scan. The size of each voxel is 3mm $\times$ 3mm $\times$ 3mm. Each acquisition took 2.9388 seconds, with the

Estimated HRF. Proposed method

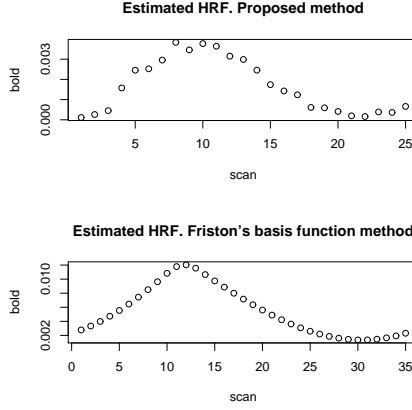**Estimated HRF. Friston's basis function method**

Figure 6.7: *Top: the estimated HRF using our proposed method; Bottom: the estimated HRF using the basis function modeling in Friston* et al. *(1995b).*

scan to scan repetition time (TR) set to be 3 seconds.

6.5.2 Analysis and Results

The data set was preprocessed using SPM5. The preprocessing included realignment, slice timing correction, coregistration and spatial smoothing.

We then analyzed the processed data set using both our proposed method and SPM5. When using SPM5, we used a canonical HRF with time and dispersion derivatives to model the hemodynamic response (Friston *et al.*, 1998) and it's shown in Figure 6.10. A t-statistic map was generated to show the activations triggered by the stimuli and part of them is shown on the first row of Figure 6.11.

When using the proposed method, to detect which regions of the brain were activated by the finger tapping movements, we generated a spatial map of the p-value for each voxel. The p-values were calculated based on the test defined by Equation (6.17). The p-map generated by the proposed method is shown on the second row in Figure 6.11.

The four image slices represent the spatial maps of the right-hand activation. The red areas illustrate activated brain regions. Brighter color indicates higher intensity. The p-maps generated by the proposed method demonstrate the classic brain activation patterns during hand movement as described in Section 3.5. However, the t-maps of the

78
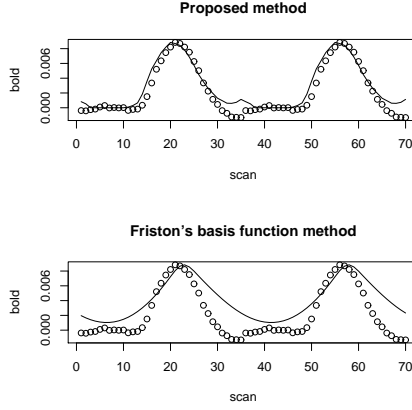
Figure 6.8: *Top: dotted line – the recorded BOLD signal, solid line – the predicted BOLD signal by convolving the stimulus with the estimated HRF using our proposed method; Bottom: dotted line – the recorded BOLD signal, solid line – the predicted BOLD signal by convolving the stimulus with the estimated HRF using the basis function modeling in Friston* et al. *(1995b).*
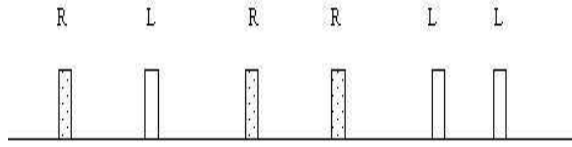
Figure 6.9: *The experiment paradigm. R: right-hand finger tapping; L: left-hand finger tapping.*

same four slices generated using SPM5 do not show any activation, as seen from the first row of Figure 6.11.

Using our proposed method, we selected the voxels which are shown to be activated according to Figure 6.11 and plotted the estimated HRFs at those voxels. Figure 6.12 displays the HRFs for five voxels selected from PMC. Figure 6.13 displays the HRFs for three voxels from SMA and Figure 6.14 shows the HRFs for three voxels from cerebellum.

All of them have the form that agrees with empirical experience. But it is clear that the HRFs in PMC, SMA and cerebellum all have different shapes from each other, which indicates the variation of the hemodynamic responses among different brain regions.

79
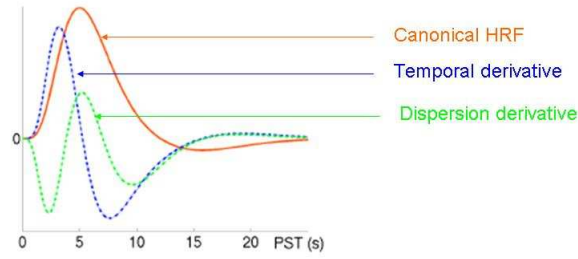
Figure 6.10: *The HRF modeled by SPM5 using a canonical HRF with time and dispersion derivatives.*

SPM t-map
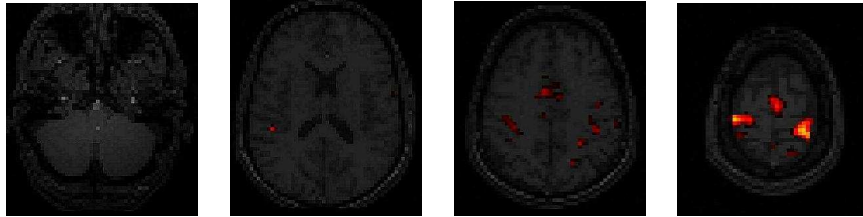
Proposed method p-map
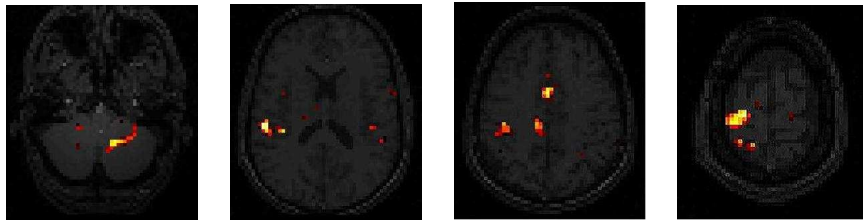
Figure 6.11: *The four related slices that contain the areas activated by right-hand finger tapping. The first row consists of the t-maps generated by SPM5 and they don't show any activation. The second row contains the p-maps generated by the proposed method. The first slice indicates the activated areas in cerebellum. The second slice contains basal ganglia. The third slice contains SMA and the fourth slice shows PMC.*
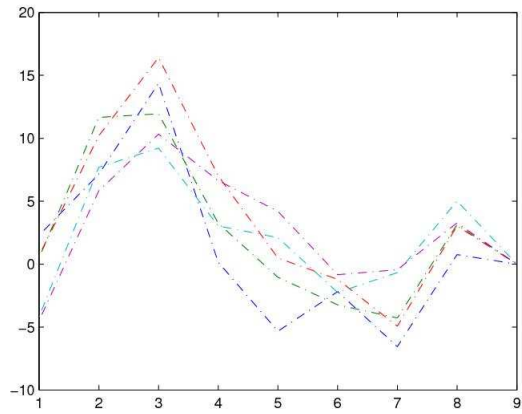
Figure 6.12: *The estimated HRFs for five voxels from PMC.*
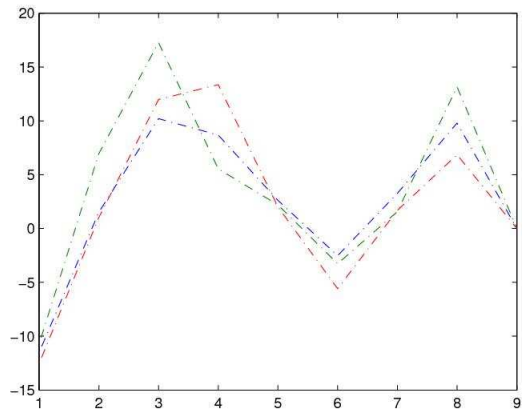
Figure 6.13: *The estimated HRFs for three voxels from SMA.*
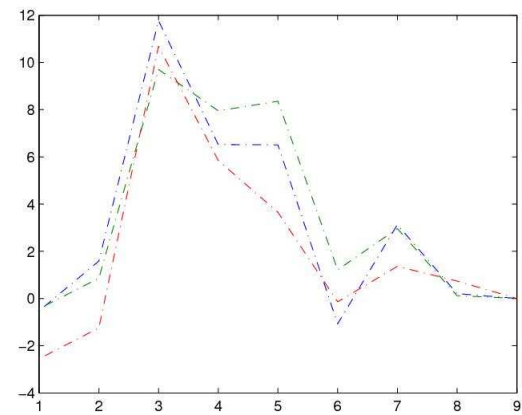
Figure 6.14: *The estimated HRFs for three voxels from cerebellum*

6.6 Discussion

In this chapter, we introduced a new method for ER-fMRI analysis. At each voxel, the HRF is estimated nonparametrically without specifying a functional form a priori. Hence the variation of the HRF across different brain regions can be studied systematically, and our method can be an useful diagnostic tool for other approaches that may be biased because of misspecification of HRF. In addition, the linear relationship specified through the BOLD signal can be examined statistically by carrying out a formal test of hypothesis. The strength of the linear relationship and hence the degree of association between the stimulus and the response can be evaluated.

On the other hand, the p-map that is generated based on the F-test (6.17) is more informative than the SPM t-map which is based on the t-statistic of the parameter estimates where no test for linearity is given. Further investigations will be done for an extensive comparative study on these maps. Another important thing for the estimate in point processes is the selection of weight function $K(\cdot)$ and bandwidth b (Section 6.2.3). Newton (1988) recommended eight commonly used weight functions, while the problem of bandwidth selection is more flexible and important. This project is currently underway.

CHAPTER 7

Further Thoughts

7.1 Connectivities and Networks

After detecting the activated voxels, a question that is attracting more and more interest in fMRI analysis is: is there any connectivity among the activated voxels? that is, does the activation of one voxel or one network of voxels drive the other voxel or network? If it does, we are interested in finding the voxel and how it drives the others. The problem here is called the connectivities of voxels.

One way to address this problem is to study the HRFs from different regions of the brain. Because HRFs can indicate the timing of the blood flow, we can decide the connectivity among different voxels by comparing the time of the initial rise of the HRFs.

7.2 Group Analyses in fMRI

Most traditional fMRI analyses are based on observations from single subject or an average of multiple subjects. However, it's been noticed that intra-subject findings cannot be easily generalized to a population. This makes it a necessary trend to develop certain robust method for group analysis in fMRI.

A classical approach for doing group analysis is offered in SPM, which calculates a t-map for each of the subjects and then applies certain statistical test based on all the t-maps. This method is restricted by most of its assumptions about the data such as normality of each subject's estimated effect and normality of the original data itself.

As implemented in the aforementioned package GIFT, Calhoun *et al.* (2001) proposed

a group analysis using ICA. The idea is to apply two-level SVD on the group data as a preprocessing step. Firstly, a SVD is applied on each subject data and dominant components are saved respectively. Concatenate all these components, a second SVD is then applied on the resulted matrix for the group. ICA is carried out in the subspace spanned by the second SVD to recover interesting spatial and temporal components.

Lukic *et al.* (2002) proposed a robust ICA algorithms for multiple fMRI datasets, which is extended from Molgedey-Schuster ICA method and is based on non-vanishing source autocorrelation functions. How to develop more interesting algorithms based on this group ICA method is another direction for future work.

Proof of Theorem 1 and Justification for Algorithm 1

Below we first provide the proof of Theorem 1 using matrix algebra.

Note that (3.7) is equivalent to minimizing

$$\text{tr}(\mathbf{X}\mathbf{X}^T) - 2d\boldsymbol{\psi}^T\mathbf{B}^T\mathbf{X}^T\mathbf{u} + d^2(\boldsymbol{\psi}^T\mathbf{B}^T\mathbf{B}\boldsymbol{\psi})(\mathbf{u}^T\mathbf{u}), \tag{A.1}$$

subject to $\mathbf{u}^T\mathbf{u} = 1$ and $\mathbf{v}^T\mathbf{v} = 1$. For fixed d and $\mathbf{u}$, the minimizer is

$$\boldsymbol{\psi} = (d\mathbf{u}^T\mathbf{u}\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}^T\mathbf{u} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}^T\mathbf{u}/d,$$

which we can plug back into (A.1) to obtain

$$\text{tr}(\mathbf{X}\mathbf{X}^T) - \mathbf{u}^T\mathbf{X}\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}^T\mathbf{u}.$$

Hence the minimization problem (3.7) is reduced to

$$\max_{\mathbf{u}} \ \mathbf{u}^T\mathbf{X}\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{X}^T\mathbf{u} \qquad \text{subject to } \mathbf{u}^T\mathbf{u} = 1, \tag{A.2}$$

which is a generalized eigen-problem. Similarly, the $\boldsymbol{\psi}$ that solves (3.7) is the solution to the following generalized eigen-problem,

$$\max_{\boldsymbol{\psi}} \ \boldsymbol{\psi}^T\mathbf{B}^T\mathbf{X}^T\mathbf{X}\mathbf{B}\boldsymbol{\psi} \qquad \text{subject to } \boldsymbol{\psi}^T\mathbf{B}^T\mathbf{B}\boldsymbol{\psi} = 1. \tag{A.3}$$

According to (A.1), the d that solves (3.7) is given by $d = \boldsymbol{\psi}^T\mathbf{B}^T\mathbf{X}^T\mathbf{u}$.

Given the above proof of Theorem 1, the numerical implementation (Algorithm 1)

can be justified as follows.

To solve (A.2), consider the Cholesky decomposition $\mathbf{B}^T\mathbf{B} = \mathbf{R}_\mathbf{B}^T\mathbf{R}_\mathbf{B}$ where $\mathbf{R}_\mathbf{B}$ is a 2×2 upper triangular matrix. The maximization problem (A.2) is then equivalent to

$$\max_{\mathbf{u}} \quad \mathbf{u}^T\mathbf{X}\mathbf{B}\mathbf{R}_\mathbf{B}^{-1}(\mathbf{R}_\mathbf{B}^{-1})^T\mathbf{B}^T\mathbf{X}^T\mathbf{u} \qquad \text{subject to } \mathbf{u}^T\mathbf{u} = 1.$$

The solution to this problem, denoted as $\mathbf{u}^*$, is actually the first left eigenvector of the matrix $\mathbf{X}\mathbf{B}\mathbf{R}_\mathbf{B}^{-1}$ (Harville, 1997).

Problem (A.3) can be solved similarly. Let $\tilde{\boldsymbol{\psi}} = \mathbf{R}_\mathbf{B}\boldsymbol{\psi}$, then (A.3) is equivalent to

$$\max_{\tilde{\boldsymbol{\psi}}} \quad \tilde{\boldsymbol{\psi}}^T(\mathbf{R}_\mathbf{B}^{-1})^T\mathbf{B}^T\mathbf{X}^T\mathbf{X}\mathbf{B}\mathbf{R}_\mathbf{B}^{-1}\tilde{\boldsymbol{\psi}} \qquad \text{subject to } \tilde{\boldsymbol{\psi}}^T\tilde{\boldsymbol{\psi}} = 1.$$

The maximizer $\tilde{\boldsymbol{\psi}}^*$ is the first right eigenvector of $\mathbf{X}\mathbf{B}\mathbf{R}_\mathbf{B}^{-1}$. Consequently, the maximizer of the original problem (A.3) is $\boldsymbol{\psi}^* = \mathbf{R}_\mathbf{B}^{-1}\tilde{\boldsymbol{\psi}}^*$.

The above derivation suggests that one only needs to perform a single SVD of $\mathbf{X}\mathbf{B}\mathbf{R}_\mathbf{B}^{-1}$ to obtain both $\boldsymbol{\psi}^*$ and $\mathbf{u}^*$. Then the scale parameter d can be estimated as $d^* = \boldsymbol{\psi}^{*T}\mathbf{B}^T\mathbf{X}^T\mathbf{u}^*$.

Proof of Result 1

(6.10): According to Equation (6.8),

$$
\begin{aligned}
E\{dY(t)\} &= E\{E[dY(t)|X]\} \\
&= E\{(s_0 + \int s_1(t-u)dX(u))dt\}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
p_Y &= \frac{E\{dY(t)\}}{dt} \\
&= E\{s_0 + \int s_1(t-u)dX(u)\} \\
&= s_0 + \int s_1(t-u)E(dX(u)) \\
&= s_0 + \int s_1(t-u)p_X du \\
&= s_0 + p_X \int s_1(u)du.
\end{aligned}
$$

(6.11): Again, based on Equation (6.8), we have:

$$
\begin{aligned}
E\{dY(t+s)dX(s)\} &= E\{E[dY(t+s)dX(s)|X]\} \\
&= E\{E[dY(t+s)|X] \cdot dX(s)\} \\
&\quad (\text{since } dX(s) = 0 \text{ or } 1) \\
&= E\{(s_0 + \int s_1(t+s-u)dX(u))dt \cdot dX(s)\} \\
&= E\{s_0 dX(s) + \int s_1(t+s-u)dX(u)dX(s)\}dt \\
&= \{s_0 E dX(s) + \int s_1(t+s-u)E(dX(s)dX(u))\}dt
\end{aligned}
$$

$$= s_0 p_X ds dt + \left[\int s_1(t+s-u)E\big(dX(s)dX(u)\big)\right]dt$$

$$= s_0 p_X ds dt + \left[s_1(t)p_X ds + \right.$$

$$\left.\int_{u \neq s} s_1(t-v)E\big(dX(s)dX(s+v)\big)\right]dt$$

$$(\text{Let} \quad v = u - s)$$

$$= s_0 p_X ds dt + s_1(t)p_X ds dt + \left[\int s_1(t-u)p_{XX}(u)du ds\right]dt$$

Hence,

$$p_{YX}(t) = \frac{E\{dY(t+s)dX(s)\}}{dt ds}$$

$$= s_0 p_X + s_1(t)p_X + \int s_1(t-u)p_{XX}(u)du.$$

(6.12): Equations (6.10) and (6.11) give us:

$$q_{XY}(t) = p_{XY}(t) - p_X p_Y$$

$$= s_0 p_X + s_1(t)p_X + \int s_1(t-u)p_{XX}(u)du - p_X\left(s_0 + p_X \int s_1(u)du\right)$$

$$= s_1(t) + \int s_1(t-u)(p_{XX}(u) - p_X p_X)du$$

$$= s_1(t)p_X + \int s_1(t-u)q_{XX}(u)du.$$

(6.13): The cross-spectrum of the two processes at frequency λ $f_{YX}(\lambda)$ and the spectrum of the process X at frequency λ $f_{XX}(\lambda)$ are defined by Equations (6.6) and (6.7). Plugging in the result of Equation (6.12), we have:

$$f_{YX}(\lambda) = \frac{1}{2\pi}\int \exp(-iu\lambda)q_{YX}(t)dt$$

$$= \frac{1}{2\pi}\int \exp(-it\lambda)\left[s_1(t)p_X + \int s_1(t-u)q_{(}XX)(u)du\right]du$$

$$= \frac{1}{2\pi}\int \exp(-it\lambda)s_1(t)p_X dt + \frac{1}{2\pi}\int \exp(-it\lambda)\left(\int s_1(t-u)q_{XX}(u)du\right)dt$$

$$= \frac{1}{2\pi} p_X \int \exp(-it\lambda) s_1(t) dt +$$

$$\frac{1}{2\pi} \int \exp(-is\lambda) s_1(s) ds \cdot \int \exp(-iu\lambda) q_{XX}(u) du$$

$$(\text{Let} \quad s = t - u)$$

$$= \int \exp(-it\lambda) s_1(t) dt \cdot \left(\frac{1}{2\pi} p_X + \frac{1}{2\pi} \int \exp(-iu\lambda) q_{XX}(u) du \right)$$

$$= S_1(\lambda) f_{XX}(\lambda),$$

where $S_1(\lambda) = \int \exp(-it\lambda) s_1(t) dt$ is the Fourier transform of $s_1(t)$.

Asymptotic Properties of the Point Process Parameter Estimates

C.1 Preliminaries

Let $X_1(t)$ and $X_2(t)$, $t \in \mathbb{R}$, be two stationary orderly point processes. Suppose $X_1(t)$ is the stimulus sequence and $X_2(t)$ is the BOLD signal. Define

$$P\{dX_a(t) = 1\} = C_a \, dt, \qquad a = 1, 2,$$

and

$$\text{cov}\{dX_a(t+u), dX_b(t)\} = C_{ab}(du) \, dt, \qquad a, b = 1, 2.$$

The power spectrum at frequency ω is defined by

$$f_{ab}(\lambda) = \frac{1}{2\pi} \int e^{-i\lambda u} C_{ab}(du), \qquad \lambda \in \mathbb{R}, \quad a, b = 1, 2.$$

C.2 Discrete Fourier Transforms

Let $\text{tap}(\cdot) : \mathbb{R} \to \mathbb{R}$ denote a **tapering function**. Suppose there are T time points in each series. The discrete Fourier transform (DFT) for the univariate series X_a is defined by

$$d_a^T(\lambda) \equiv d^T(\lambda; X_a) = \sum_t \text{tap}_a(t/T) X_a(t) \exp(-i\lambda t), \qquad \lambda \in \mathbb{R}, \quad a = 1, 2.$$

For vector-valued series $\mathbf{X}$, it is given by

$$\mathbf{d}^T(\lambda) \equiv \mathbf{d}^T(\lambda; \mathbf{X}) = \sum_t [\text{tap}_a(t/T) X_a(t)] \exp(-i\lambda t), \qquad \lambda \in \mathbb{R}.$$

We introduced the tapering function here to simplify the argument in the following proof. Chapter 6 presents the case with $\text{tap} \equiv 1$. In the implementation of this procedure, this is an option that can be set by the user.

Condition 1. *The tapering function* $\text{tap}(\cdot) : \mathbb{R} \to \mathbb{R}$ *has a compact support with bounded first derivative. Furthermore,*

$$\int \text{tap}(u) \, du = 1 \qquad and \qquad \int |\text{tap}(u)| \, du < \infty.$$

Set $\text{tap}_a^T(t) = \text{tap}_a(t/T)$ and

$$\text{TAP}_{a_1,\ldots,a_k}^T(\lambda) = \sum_t \left(\prod_{j=1}^k \text{tap}_{a_j}^T(t) \right) \exp(-i\lambda t), \qquad \lambda \in \mathbb{R}, a_j \in \{1,2\}, j = 1, \ldots, k.$$

Condition 2. *The covariance function satisfies*

$$\sum_u C_{a_1 a_2}(u) < \infty, \quad and \quad \sum_{u_1,\ldots,u_{k-1}} C_{a_1 \ldots a_k}(u_1, \ldots, u_{k-1}) < \infty, \qquad a_1, \ldots, a_k = 1, 2.$$

The second part of the above condition is necessary for establishing the asymptotic properties of the estimates to be considered in this section.

Lemma 1. *Suppose Conditions 1 and 2 hold. Then*

$$\sup_{\lambda_1,\ldots,\lambda_k} \; \left| \text{cum}(d_{a_1}^T(\lambda_1), \ldots, d_{a_k}^T(\lambda_k)) - \right.$$
$$\left. (2\pi)^{k-1} \text{TAP}_{a_1,\ldots,a_k}^T (\lambda_1 + \cdots + \lambda_k) \, f_{a_1,\ldots,a_k}(\lambda_1, \ldots, \lambda_k) \right| = o(T),$$

where $\text{cum}(d_{a_1}^T(\lambda_1), \ldots, d_{a_k}^T(\lambda_k))$ *denotes the joint cumulant of* $d_{a_1}^T(\lambda_1), \ldots, d_{a_k}^T(\lambda_k)$.

Condition 3. *The covariance function satisfies*

$$\sum_u |u| C_{ab}(u) < \infty, \quad and \quad \sum_{u_1,\dots,u_{k-1}} |u_j| C_{a_1\dots a_k}(u_1,\dots,u_{k-1}) < \infty, \quad a_1,\dots,a_k = 1,2.$$

Lemma 2. *Under Conditions 1 and 3,*

$$\sup_{\lambda_1,\dots,\lambda_k} \ \left| \mathrm{cum}(d_{a_1}^T(\lambda_1),\dots,d_{a_k}^T(\lambda_k)) - \right.$$

$$\left. (2\pi)^{k-1} \mathrm{TAP}_{a_1,\dots,a_k}^T(\lambda_1 + \cdots + \lambda_k) f_{a_1,\dots,a_k}(\lambda_1,\dots,\lambda_k) \right| = O(1),$$

where $\mathrm{cum}(d_{a_1}^T(\lambda_1),\dots,d_{a_k}^T(\lambda_k))$ *denotes the joint cumulant of* $d_{a_1}^T(\lambda_1),\dots,d_{a_k}^T(\lambda_k)$.

The DFT is asymptotically normal with mean specified according to the frequency λ as described below.

Theorem 2. *Under Conditions 1 and 2,* $d_a^T(\lambda)$ *is asymptotically*

1. $N_1^c(0, 2\pi T f_{aa}(\lambda) \mathrm{TAP}_{aa}(0))$ *if* $\lambda \neq 0 \mod \pi$,

2. $N_1(T c_a \mathrm{TAP}_{aa}(0), 2\pi T f_{aa}(\lambda) \mathrm{TAP}_{aa}(0))$ *if* $\lambda = 0, \pm 2\pi, \dots$,

3. $N_1(0, 2\pi T f_{aa}(\lambda) \mathrm{TAP}_{aa}(0))$ *if* $\lambda = \pm\pi, \dots$.

Note that $\mathrm{TAP}_{aa}(0) = \int \mathrm{tap}^2$. The above result implies that the real and the imaginary part of $d_a^T(\lambda)$ are approximately independent. Each is approximately normal with mean and variance $\pi T f_{aa}(\lambda) \int \mathrm{tap}^2$.

C.3 Periodogram

The distributions of the DFT suggests the following statistic:

$$I_{aa}^T(\lambda) = |d_a^T(\lambda)|^2 / \left(2\pi \sum_t [\mathrm{tap}(t/T)]^2 \right), \qquad \lambda \in \mathbb{R}. \tag{C.1}$$

This is called **periodogram** and is an estimate of the spectral density function f_{aa}. For more historical remarks, see Brillinger (1981).

Note that if there is no tapering function, the periodogram is given by

$$I_{aa}^T(\lambda) = (2\pi T)^{-1}|d_a^T(\lambda)|^2, \qquad \lambda \in \mathbb{R}.$$

Let $\lambda_j = 2\pi j/T$, $j = 0, \pm 1, \pm 2, \ldots$ denote the Fourier frequencies. The result below describes the asymptotic distribution of the periodograms.

Theorem 3. *Under Conditions 1–3, $I_{aa}^T(\lambda_j)$, $j = 1, \ldots, J$, are asymptotically independent $f_{aa}(\lambda_j)\chi_2^2/2$. Also $I_{aa}^T(\lambda)$ is asymptotically $f_{aa}(\lambda)\chi_1^2$ for $\lambda = \pm\pi, \pm 3\pi, \ldots$, independent of the $I_{aa}^T(\lambda_j)$, $j = 1, \ldots, J$. J is the number of frequencies.*

The above result shows that the asymptotic variance of the periodogram is approximately $f_{aa}(\lambda)^2$, which is usually positive. Thus the periodogram is not a consistent estimate of the spectral density function. The following section will present a class of consistent estimates obtained by smoothing the periodograms.

C.4 Window Estimates — The Smoothed Periodograms

A class of consistent estimates can be obtained by using a running mean or local average of the periodograms. Specifically, set

$$\hat{f}(\lambda_k) = (2m+1)^{-1} \sum_{j=-m}^{m} I_{aa}^T\left(\frac{2\pi(k+j)}{T}\right).$$

It follows from the asymptotic distributional properties of the periodograms (Theorem 3) that $\hat{f}(\lambda_j)$, $j = 1, \ldots, J$, are asymptotically independent with $\hat{f}(\lambda) \sim f(\lambda)\chi_{4m+2}^2/(4m+2)$ if $\lambda \neq 0$, and $\hat{f}(0) \sim f(0)\chi_{2m}^2/(2m)$. An important implication of the above result is that consistency can be achieved by letting $m \to \infty$ and $m/T \to 0$ as $T \to \infty$.

The local average estimate can be expressed in the form of

$$\hat{f}(\lambda_k) = \begin{cases} \sum_{j=-m}^{m} W_j I_{aa}^T \left(\frac{2\pi(k+j)}{T} \right) & k \neq 0, \\ \left(\sum_{j=1}^{m} W_j \right)^{-1} \sum_{j=1}^{m} W_j I_{aa}^T \left(\frac{2\pi j}{T} \right) & k = 0, \end{cases}$$

where the weights W_j, $j = 0, \pm 1, \pm 2, \ldots, \pm m$, satisfy $\sum_j W_j = 1$.

More generally, let $W(\cdot)$ denote a weight function and set

$$\hat{f}(\lambda) = \sum_{s \neq 0} B_T^{-1} W \left(B_T^{-1} \left(\lambda - \frac{2\pi s}{T} \right) \right) I_{aa}^T \left(\frac{2\pi s}{T} \right), \tag{C.2}$$

where B_T is referred to as the **bandwidth** or **window width** that will be specified more clearly later. Certain properties of the weight function $W(\cdot)$ will be required in order to assure that the above estimate is consistent.

Condition 4. *The weight function $W(\cdot) : \mathbb{R} \to \mathbb{R}$ has a compact support with bounded first derivative. Furthermore,*

$$\int W(\lambda) \, d\lambda = 1 \qquad and \qquad \int |W(\lambda)| \, d\lambda < \infty.$$

Under this condition, the bias of the window estimate is given by

$$E(\hat{f}_{aa}(\lambda)) = \int W(\beta) f_{aa}(\lambda - B_T \beta) \, d\beta + O(T^{-1} B_T^{-1}).$$

In fact, more properties can be obtained and are stated in the following result.

Theorem 4. *Under Conditions 1–3 and suppose that the spectral density function f_{aa} does not vanish. Let $B_T \to 0$ and $B_T T \to \infty$ as $T \to \infty$. Then, $\hat{f}(\lambda_j)$, $j = 1, \ldots, J$, are*

asymptotically normal with mean zero and covariance structure given by

$$\lim_{T\to\infty} B_T T \operatorname{cov}(\hat{f}(\lambda), \hat{f}(\mu)) = \begin{cases} 0 & \text{if } \lambda \neq \mu, \\ 2\pi f(\lambda)^2 \int W^2 & \text{otherwise.} \end{cases} \tag{C.3}$$

C.5 Transfer Function

In this section, we will consider the transfer model given by

$$E(X_2(t)|X_1) = \mu + \sum_u s(t-u)X_1(u). \tag{C.4}$$

Alternatively, this is expressed as

$$X_2(t) = \mu + \sum_u s(t-u)X_1(u) + \epsilon(t). \tag{C.5}$$

Suppose $S(\cdot)$ is the Fourier transform of the *impulse response function $s(\cdot)$*, according to Cramér Representation, there are random measures Z_1, Z_2 and Z_ϵ such that (by omitting μ for simplicity)

$$\int \exp(i\lambda t)\, dZ_2(\lambda) = \sum_u s(t-u) \int \exp(i\lambda u)\, dZ_1(\lambda) + \int \exp(i\lambda t)\, dZ_\epsilon(\lambda)$$

$$= \int \sum_u s(t-u)\exp(-i\lambda(t-u))\exp(i\lambda t)\, dZ_1(\lambda) + \int \exp(i\lambda t)\, dZ_\epsilon(\lambda)$$

$$= \int S(\lambda)\exp(i\lambda t)\, dZ_1(\lambda) + \int \exp(i\lambda t)\, dZ_\epsilon(\lambda).$$

Therefore,

$$E|dZ_2(\lambda)|^2 = |S(\lambda)|^2 E|dZ_1(\lambda)|^2 + E|dZ_\epsilon(\lambda)|^2,$$

or, using $E|dZ_2(\lambda)|^2 = f_{22}(\lambda)$, $E|dZ_1(\lambda)|^2 = f_{11}(\lambda)$ and $E|dZ_\epsilon(\lambda)|^2 = f_{\epsilon\epsilon}(\lambda)$,

$$f_{22}(\lambda) = |S(\lambda)|^2 f_{11}(\lambda) + f_{\epsilon\epsilon}(\lambda). \tag{C.6}$$

Moreover, it follows from (C.5) that

$$f_{21}(\lambda) = S(\lambda)f_{11}(\lambda). \tag{C.7}$$

To estimate these quantities, we now set

$$I_{ab}^T(\lambda) = (2\pi T)^{-1}d_a^T(\lambda)\overline{d_b^T(\lambda)}, \tag{C.8}$$

$$\hat{f}_{ab}(\lambda_k) = (2m+1)^{-1}\sum_{j=-m}^{m} I_{ab}^T\left(\frac{2\pi(k+j)}{T}\right), \qquad a,b=1,2, \tag{C.9}$$

$$\hat{S}(\lambda) = \hat{f}_{21}(\lambda)/\hat{f}_{11}(\lambda), \qquad \lambda \in \mathbb{R}. \tag{C.10}$$

Then under Condition 4,

$$E(\hat{f}_{ab}(\lambda)) = \int W(\nu)f_{ab}(\lambda - B_T\nu)\,d\nu + O(T^{-1}B_T^{-1}), \qquad a,b=1,2.$$

Moreover, we have the following asymptotic result:

Theorem 5. *Under Conditions 1–3 and suppose that $f_{ab} \neq 0$ and $f_{ab}'' < \infty$. Let $B_T T \to \infty$, $B_T^5 T \to 0$ as $T \to \infty$. Then $\hat{S}(\lambda)$ is approximately normal with mean $S(\lambda)$ and variance*

$$2\pi f_{\epsilon\epsilon}(\lambda)\int W^2/(B_T T f_{11}(\lambda)),$$

where

$$f_{\epsilon\epsilon}(\lambda) = f_{22}(\lambda) - f_{21}(\lambda)f_{11}(\lambda)^{-1}f_{12}(\lambda). \tag{C.11}$$

We estimate $f_{\epsilon\epsilon}(\lambda)$ by

$$\hat{f}_{\epsilon\epsilon}(\lambda) = \frac{2m+1}{2m+1-r}[\hat{f}_{22}(\lambda) - \hat{f}_{21}(\lambda)\hat{f}_{11}(\lambda)^{-1}\hat{f}_{12}(\lambda)]. \tag{C.12}$$

96

C.6 Proofs

C.6.1 Proofs of Lemmas 1 and 2

If follows from

$$|w_a(t+u)w_b(t+v) - w_a(t)w_b(t)|$$

$$\leq |w_a(t+u)w_b(t+v) - w_a(t+u)w_b(t)| + |w_a(t+u)w_b(t) - w_a(t)w_b(t)|$$

and Condition 1 that there is a constant K_1 such that

$$|\sum_t w_{a_1}^T(t+u_1)\cdots w_{a_{k-1}}^T(t+u_{k-1})w_{a_k}^T(t)\exp(-i\lambda t) - W_{a_1\ldots a_k}^T(\lambda)|$$

$$\leq K_1(|u_1| + \cdots + |u_{k-1}|).$$

and

$$\operatorname{cum}(d_{a_1}^T(\lambda_1),\ldots,d_{a_k}^T(\lambda_k))$$

$$= \sum_{t_1}\cdots\sum_{t_k} w_{a_1}^T(t_1)\cdots w_{a_k}^T(t_k)\exp\left(-i\sum_{j=1}^k \lambda_j t_j\right) C_{a_1,\ldots,a_k}(t_1 - t_k,\ldots,t_{k-1} - t_k)$$

$$= \sum_{u_1=-2(T-1)}^{2(T-1)}\cdots\sum_{u_{k-1}=-2(T-1)}^{2(T-1)} \exp\left(-i\sum_{j=1}^{k-1}\lambda_j t_j\right) C_{a_1,\ldots,a_k}(u_1,\ldots,u_{k-1}) \times$$

$$\sum_t w_{a_1}^T(t+u_1)\cdots w_{a_{k-1}}^T(t+u_{k-1})w_{a_k}^T(t)\exp\left(-i\sum_{j=1}^k \lambda_j t\right)$$

$$= \sum_{u_1=-2(T-1)}^{2(T-1)}\cdots\sum_{u_{k-1}=-2(T-1)}^{2(T-1)}$$

$$\exp\left(-i\sum_{j=1}^{k-1}\lambda_j t_j\right) C_{a_1,\ldots,a_k}(u_1,\ldots,u_{k-1})W_{a_1\ldots a_k}^T(\lambda_1 + \cdots + \lambda_k) + \epsilon_T,$$

where

$$|\epsilon_T| \leq K_2 \sum_{u_1=-2(T-1)}^{2(T-1)} \cdots \sum_{u_{k-1}=-2(T-1)}^{2(T-1)} (|u_1| + \cdots + |u_{k-1}|) C_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1}).$$

It now follows from Condition 3,

$$T^{-1}|\epsilon_T| \leq K_2 \sum_{u_1=-2(T-1)}^{2(T-1)} \cdots \sum_{u_{k-1}=-2(T-1)}^{2(T-1)} T^{-1}(|u_1| + \cdots + |u_{k-1}|) C_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1}),$$

$T^{-1}(|u_1| + \cdots + |u_{k-1}|) \to 0$ and the dominated convergence theorem that

$$|\epsilon_T| = o(T). \tag{C.13}$$

Lemmas 1 and 2 follow from this and

$$f_{a_1,\ldots,a_k}(\lambda_1, \ldots, \lambda_{k-1}) = (2\pi)^{k-1} \sum \cdots \sum \exp\left(-i \sum_1^{k-1} \lambda_j u_j\right) C_{a_1,\ldots,a_k}(u_1, \ldots, u_{k-1}) + o(1).$$

C.6.2 Proof of Theorem 2

From Condition 1, $\text{TAP}_{a_1,\ldots,a_k}^T(\lambda) = O(T)$. Recall that the Gaussian distribution has cumulants of order greater than 2 vanishes. The desired result now follows from Lemmas 1, 2 and the fact that

$$\begin{aligned}
&T^{-k/2} \text{cum}(d_{a_1}^T(\lambda_1), \ldots, d_{a_k}^T(\lambda_k)) \\
&= T^{-k/2}(2\pi)^{k-1} \text{TAP}_{a_1,\ldots,a_k}^T (\lambda_1 + \cdots + \lambda_k) f_{a_1,\ldots,a_k}(\lambda_1, \ldots, \lambda_k) + o(T^{1-k/2}) \\
&\to 0 \qquad \text{for } k > 2 \text{ as } T \to 0.
\end{aligned}$$

C.6.3 Proof of Theorem 3

This follows from Theorem 2 and the definition of the chi-square distribution.

C.6.4 Proof of Theorem 4

Direct computation shows that

$$\text{cov}(I(\lambda), I(\mu)) = f(\lambda) \left\{ \left(\frac{\sin T(\lambda + \mu)/2}{T \sin (\lambda + \mu)/2} \right)^2 + \left(\frac{\sin T(\lambda - \mu)/2}{T \sin (\lambda - \mu)/2} \right)^2 \right\} + O(1/T).$$

Moreover,

$$
\begin{aligned}
\text{cov}(\hat{f}(\lambda), \hat{f}(\mu)) &= 2\pi T^{-1} \int W^T(\lambda - \alpha) W^T(\mu - \alpha) f(\alpha)^2 \, d\alpha + \\
&\quad 2\pi T^{-1} \int W^T(\lambda - \alpha) W^T(\mu + \alpha) f(\alpha)^2 \, d\alpha + \\
&\quad O(B_T^{-2} T^{-2}) + O(T^{-1}).
\end{aligned}
$$

The indicated covariance structure C.3 is an easy consequence of these results.

To obtain the asymptotic normality, we need to show that all cumulants of order higher than 2 tend to zero as $T \to \infty$. This is carried out by directly computing the cumulants of the window estimates in a manner similar to the proof of Lemma 1.

C.6.5 Proof of Theorem 5

We begin with two lemmas.

Lemma 3. *Let $(\mathbf{V}_n)$ denote a sequence of random vectors converging in distribution to $\mathbf{V}$. Then there exists a probability space such that $\mathbf{V}_n$ converges to $\mathbf{V}$ almost surely.*

The proof can be found in Billingsley (1995).

Lemma 4. *Let $(\mathbf{V}_n)$ denote a sequence of random vectors in $\mathbb{R}^p$ converging in distribution to $N_p^c(\mathbf{0}, \mathbf{I}_p)$ and $(\mathbf{U}_n)$ a sequence of $p \times p$ unitary matrices. Then $\mathbf{U}_n \mathbf{V}_n$ converges to $N_p^c(\mathbf{0}, \mathbf{I}_p)$ as $n \to \infty$.*

The proof follows from Lemma 3.

Now we proceed to the proof of the main result. Let d_j^T be the Fourier transform of X_j, $j = 1, 2$. Let $2\pi k/T$ denote the Fourier frequency that is nearest to λ. Then

$$d_2^T(2\pi(k + q)/T)$$
$$= S(2\pi(k + q)/T)d_1^T(2\pi(k + q)/T) + d_\epsilon^T(2\pi(k + q)/T) + O(1)$$
$$= S(\lambda)d_1^T(2\pi(k + q)/T) + d_\epsilon^T(2\pi(k + q)/T) + O(1), \quad q = 0, \pm 1, \ldots, \pm m,$$

where $O(1)$ is uniformly in q. Now let $\mathbf{D}_2$ denote the $1 \times (2m + 1)$ matrix given by

$$\mathbf{D}_2 = (2\pi T)^{-1/2} \left[\begin{array}{cccc} d_2^T(2\pi(k - m)/T) & \cdots & d_2^T(2\pi k/T) & \cdots & d_2^T(2\pi(k + m)/T) \end{array} \right].$$

Define $\mathbf{D}_1$ and $\mathbf{D}_\epsilon$ similarly. Then

$$\mathbf{D}_2 = S(\lambda)\mathbf{D}_1 + \mathbf{D}_\epsilon + O(T^{-1/2}).$$

Let $\mathbf{U} \equiv \mathbf{U}^T = [\mathbf{U}_1, \mathbf{U}_2]$ be a $(2m + 1) \times (2m + 1)$ unitary matrix whose first column is $\mathbf{U}_1 = \mathbf{D}_1^H(\mathbf{D}_1\mathbf{D}_1^H)^{-1/2}$, where $\mathbf{D}^H = \overline{\mathbf{D}}^\top$ is the conjugate transpose of $\mathbf{D}$. Then

$$\mathbf{D}_2\mathbf{U} = S(\lambda)\mathbf{D}_1\mathbf{U} + \mathbf{D}_\epsilon\mathbf{U} + O(T^{-1/2}).$$

The first and the remaining columns of these matrices yield

$$[\hat{S}(\lambda) - S(\lambda)]\hat{f}_1(\lambda)^{1/2}(2m + 1)^{1/2} = \mathbf{D}_\epsilon\mathbf{U}_1 + O(T^{-1/2}), \tag{C.14}$$

$$\mathbf{D}_2\mathbf{U}_2 = \mathbf{D}_\epsilon\mathbf{U}_2 + O(T^{-1/2}). \tag{C.15}$$

By the property of the unitary matrix,

$$(2m + 1)\hat{f}_2 = \mathbf{D}_2\mathbf{D}_2^H$$
$$= \mathbf{D}_2\mathbf{U}_1\mathbf{U}_1^H\mathbf{D}_2^H + \mathbf{D}_2\mathbf{U}_2\mathbf{U}_2^H\mathbf{D}_2^H$$

$$= \mathbf{D}_2\mathbf{D}_1^H(\mathbf{D}_1\mathbf{D}_1^H)^{-1}\mathbf{D}_1\mathbf{D}_2^H + \mathbf{D}_2\mathbf{U}_2\mathbf{U}_2^H\mathbf{D}_2^H.$$

Thus

$$\hat{f}_{\epsilon\epsilon} = \mathbf{D}_2\mathbf{U}_2\mathbf{U}_2^H\mathbf{D}_2^H = \mathbf{D}_\epsilon\mathbf{U}_2\mathbf{U}_2^H\mathbf{D}_\epsilon^H + O_p(T^{-1/2}). \qquad (C.16)$$

Now, according to Theorem 2, $\mathbf{D}_\epsilon \to_d N_{2m+1}^c N(\mathbf{0}, f_\epsilon(\lambda)\mathbf{I})$ and therefore $f_\epsilon(\lambda)^{-1/2}\mathbf{D}_\epsilon \to_d$ $N_{2m+1}^c N(\mathbf{0}, \mathbf{I})$. By Lemma 4, $f_\epsilon(\lambda)^{-1/2}\mathbf{D}_\epsilon\mathbf{U} \to_d N_{2m+1}^c N(\mathbf{0}, \mathbf{I})$, or $\mathbf{D}_\epsilon\mathbf{U} \to_d N_{2m+1}^c N(\mathbf{0}, f_\epsilon(\lambda)\mathbf{I})$. This, together with (C.14) and (C.16) yield the desired result. This completes the proof of the theorem.

BIBLIOGRAPHY

Bach F.R. and Jordan M.I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research* **3**, 1–48.

Bandettini P.A., Jesmanowicz A., Wong E.C. and Hyde J.S. (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine* **30**, 161–173.

Bell A. and Sejnowski T.J. (1995). An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation* **7**, 1129–1159.

Billingsley P. (1995). *Probability and Measures*. Wiley, New York, 3 edition.

Boynton G.M., Engel S.A., Glover G.H. and Heeger D.J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *The Journal of Neuroscience* **16**, 4207–4221.

Brillinger D.R. (1974). Cross-Spectral Analysis of Processes with Stationary Invrements Including the Stationary G/G/∞ Queue. *The Annals of Probability* **2**, 815–827.

Brillinger D.R. (1975). The identification of point process systems. *The Annals of Probability* **3**, 909–929.

Brillinger D.R. (1981). *Time Series: Data Analysis and Theory*. Holden-Day, San Francisco, CA.

Buhmann C., Glauche V., Sturenburg H. *et al.* (2003). Pharmacologically modulated fMRI–cortical responsiveness to levodopa in drug-naive hemiparkinsonian patients. *Brain* **126**, 451–461.

Calhoun V.D., Adali T., Hansen L.K., Larsen J. and Pekar J.J. (2003). ICA of functional MRI data: an overview. In: *4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pp. 281–288.

Calhoun V.D., Adali T., Pearlson G.D. and Pekar J.J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping* **14**, 140–151.

Cardoso J.F. (1997). Infomax and maximum likelihood for blind source separation. *IEEE signal processing letters* **4**, 112–114.

Caulhoun V.D., Stevens M.C., Pearlson G.D. and Kiehl K.A. (2004). fMRI analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. *NeuroImage* **22**, 252–257.

Chen A. (2005). Fast kernel density independent component analysis. In: *the 6th International Conference on Independent Component Analysis and Blind Signal Separation*, volume 3889, pp. 24–31. Charleston, SC, USA.

Chen A. and Bickel P.J. (2005). Consistent independent component analysis and prewhitening. *IEEE Transactions on Signal Processing* **53**, 3625–3632.

Cox D.R. and Isham V. (1980). *Point Processes*. Chapman and Hall.

Dale A.M. and Buckner R.L. (1997). Selective averaging of rapidly presented individual trials using fmri. *Human Brain Mapping* **5**, 329–340.

Elsinger C., Harrington D. and Rao S. (2006). From preparation to online control: reappraisal of neural circuitry mediating internally generated and externally guided actions. *NeuroImage* **31**, 1177–1187.

Friston K.J., Frith C.D., Turner R. and Frackowiak R.S.J. (1995a). Characterizing evoked hemodynamics with fmri. *NeuroImage* **2**, 157–165.

Friston K.J., Holmes A.P., Worsley K.J., Poline J.P., Frith C.D. and Frackowiak R.S.J. (1995b). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping* **2**, 189–210.

Friston K.J., Jezzard P. and Turner R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping* **1**, 153–171.

Friston K.J., Josephs O., Rees G. and Turner R. (1998). Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine* **39**, 41–52.

Genovese C.R. (2000). A Bayesian Time-Course Model for Functional Magnetic Resonance Imaging Data (with discussion). *Journal of the American Statistical Association* **95**, 691–703.

Gössl C., Auer D.P. and Fahrmeir L. (2001a). Bayesian spatiotemporal inference in functional magnetic resonance imaging. *NeuroImage* **57**, 554–562.

Gössl C., Fahrmeir L. and Auer D.P. (2001b). Bayesian modeling of the hemodynamic response function in BOLD fMRI. *NeuroImage* **14**, 140–148.

Goutte C., Nielsen F.A. and Hansen L.K. (2000). Modeling the haemodynamic response in fMRI using smooth FIR filters. *IEEE Transactions on Medical Imaging* **19**, 1188–1201.

Goutte C., Toft P., Rostrup E., Nielsen F.A. and Hansen L.K. (1999). On clustering fmri time series. *NeuroImage* **9**, 298–310.

Harville D.A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer.

Hastie T. and Tibshirani R. (2002). Independent component analysis through product density estimation. Technical report, Department of Statistics at Stanford University.

Hastie T., Tibshirani R. and Friedman J.H. (2003). *The Elements of Statistical Learning*. Springer.

Hu D., Yan L.R., Liu Y.D., Zhou Z.T., Friston K.J., Tan C.L. and Wu D.X. (2005). Unified SPM-ICA for fMRI analysis. *NeuroImage* **25**, 746–755.

Huettel S.A., Song A.W. and McCarthy G. (2004). *Functional Magnetic Resonance Imaging*. Sinauer Associates, Inc.

Hyvärinen A., Karhunen J. and Oja E. (2001). *Independent Component Analysis*. John Wiley & Sons.

Hyvärinen A. and Oja E. (2000). Independent component analysis: algorithms and applications. *Neural Networks* **13**, 411–430.

Jezzard P., Matthews P.M. and Smith S.M. (editors) (2001). *Functional MRI: an introduction to methods*. Oxford University Press.

Josephs O., Turner R. and Friston K. (1997). Event-related fmri. *Human Brain Mapping* **5**, 243–248.

Kao Y. and MacFall J.R. (2000). Correction of MR κ-space data corrupted by spike noise. *IEEE Transactions on Medical Imaging* **19**, 671–680.

Kawaguchi A. and Truong Y.K. (2007). Spline independent component analysis. Manuscript.

Kherif F., Poline J.B., Flandin G., Benali H., Simon O., Dehaene S. and Worsley K.J. (2002). Multivariate model specification for fmri data. *NeuroImage* **16**, 1068–1083.

Lange N. (1996). Statistical Approaches to Human Brain Mapping by Functional Magnetic Resonance Imaging. *Statitics in Medicine* **15**, 389–428.

Lange N., Strother S.C., Anderson J.R., Nielsen F.A., Holmes A.P., Kolenda T., Savory R. and Hansen L.K. (1999). Pluratlity and Resemblance in fMRI Data Analysis. *NeuroImage* **10**, 282–303.

Lange N. and Zeger S.L. (1997). Non-linear Fourier Time Series Analysis for Human Brain Mapping by Functional Magnetic Resonance Imaging (with discussion). *Journal of the Royal Statistical Society. Series B* **46**, 1–29.

Lindquist M. and Wager T. (2007). Validity and power in hemodynamic response modeling: A comparison study and a new approach. *Human Brain Mapping* Available online.

Lukic A.S., Wernick M.N., Hansen L.K., Anderson J. and Strother S.C. (2002). A spatially robust ICA algorithm for multiple fMRI data sets. In: *IEEE International Symposium on Biomedical Imaging, Proceedings*, pp. 839–842.

Luo W. and Nichols T.E. (2003). Diagnosis and exploration of massively univariate neuroimaging models. *NeuroImage* **19**, 1014–1032.

Marchini J.L. and Ripley B.D. (2000). A new statistical approach to detecting significant activation in functional MRI. *NeuroImage* **12**, 366–380.

McKeown M.J., Hansen L.K. and Sejnowski T.J. (2003). Independent component analysis of functional MRI: what is signal and what is noise? *Current Opinion in Neurobiology* **13**, 620–629.

McKeown M.J., Jung T.P., Makeig S. *et al.* (1998a). Spatially independent activity patterns in functional MRI data during the stroop color-naming task. *PNAS* **95**, 803–810.

McKeown M.J., Makeig S., Brown G.G., Jung T., Kindermann S.S., Bell A.J. and Sejnowski T.J. (1998b). Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping* **6**, 160–188.

Newton H.J. (1988). *Timeslab: A Time Series Analysis Laboratory*. Wadsworth & Brooks/Cole.

Petersen K.S., Hansen L.K., Kolenda T., Rostrup E. and Strother S.C. (2000). On the independent components of functional neuroimages. In: *Proceedings of the Third International Conference on Independent Component Analysis and Blind Source Separation (ICA2000)*, pp. 615–620.

Rayens W.S. and Andersen A.H. (2006). Multivariate analysis of fMRI data by oriented partial least squares. *Magnetic Resonance Imaging* **24**, 953–958.

Robert C.P. and Casella G. (2004). *Monte Carlo Statistical Methods*. Springer.

Rosen B.R., Buckner R.L. and Dale A.M. (1998). Event-related functional MRI: past, present, and future. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 773–780.

Stone J.V. (2004). *Independent Component Analysis: a tutorial introduction*. The MIT Press.

Taniwaki T., Okayama A., Yoshiura T. *et al.* (2006). Functional network of the basal ganglia and cerebellum motor loops in vivo: different activation patterns between self-initiated and externally triggered movements. *NeuroImage* **31**, 745–753.

Woolrich M.W., Behrens T.E. and Smith M. S (2004). Constrained linear basis sets for hrf modelling using variational bayes. *NeuroImage* **21**, 1748–1761.