TRUE SPATIO-TEMPORAL DETECTION AND ESTIMATION FOR FUNCTIONAL MAGNETIC RESONANCE IMAGING

by

Joonki Noh

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2007

Doctoral Committee:

Professor Jeffrey A. Fessler, Co-Chair
Professor Victor Solo, Co-Chair
Professor Douglas C. Noll
Assistant Professor Clayton D. Scott

This dissertation is dedicated to my parents and Hyun Kyung.

ACKNOWLEDGEMENTS

This dissertation owes the existence to many persons. First and foremost, I am immensely grateful to my advisor, Professor Victor Solo, for his guidance, advices, encouragement and patience. I admire his enthusiasm, insight, and attitude to science. His unique view and vast knowledge of many scientific fields showed me directions to go as a researcher and laid the cornerstone of my research. From him, I earned one invaluable maxim, ”a good researcher never gives up”, which I always keep in my mind when I struggle. Without doubt, these influences will continue on my future works and life in general.

I would like to express my special thanks to my co-advisor, Professor Jeffrey A. Fessler. From the beginning of my graduate study in the University of Michigan, he has provided me invaluable advices, not only for medical imaging but also for academic life. Especially, in the last year of my graduate study, I deeply appreciate the opportunity to participate in his dual-energy computed tomography project and financial supports. It has been great pleasure, honor and very nice experiences to work and interact with a scholar and a mentor of his caliber.

I would also like to thank the members of my dissertation committee, Professor Douglas C. Noll and Professor Clayton D. Scott for their careful reading and priceless comments on this dissertation.

I am sincerely grateful to my wife, Hyun Kyung, for her sacrifices and supports with endless prayers during my graduate study. She is the greatest gift that God gives me. Without her, I would not be the person who I am in all aspects. I would also like to thank my parents, parents-in-law and family in South Korea for giving me unconditional love and for always believing me.

I would like to extend my thanks to my colleague, Magnus O. Ulfarsson. Not only technical discussions but also personal interactions with him were invaluable to me throughout the years. I also thank many Korean friends for being with me and helping me for the past 4 years. In particular, my thanks should go to Seyoung Chun in Image Computing Laboratory and Yoon Chung Kim in FMRI Laboratory.

Finally, I sincerely thank God for allowing me the opportunity to work with many wonderful persons in the University of Michigan and for leading me this far.

The research presented in this dissertation was partially funded by Riethmiller Fellowship of the Department of Electrical Engineering and Computer Science, the University of Michigan, and a research fund from the University of Michigan.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

CHAPTER 1

Introduction to FMRI Data Analysis

1.1 Overview

In the human body, the brain is probably the most complicated and the least understood organ in spite of the developments in science. Due to recent technological advances made in the last decade of the 20th century, we have dramatically broadened and deepened our understandings of the human brain. Especially, the advent of brain imaging techniques allows researchers to study a healthy and living human subject while the subject's brain is functioning and performing cognitive tasks without the need for surgery. Among proposed brain imaging techniques, positron emission tomography (PET) and functional magnetic resonance imaging (FMRI) have been found useful and widely used for the investigation of the brain. To investigate functions of the brain using PET, a radioactive tracer, e.g., $[^{15}O]H_2O$ is injected into subject's blood vessel and the changes of regional cerebral blood flow (rCBF) in the brain induced by given cognitive tasks are directly measured. Although PET is non-invasive, it involves exposure to ionizing radiation and has poor spatial resolution.

A dominant contrast mechanism used in FMRI is the blood oxygenation level dependent (BOLD) contrast, which is known as the result of complex interactions between localized changes in cerebral blood flow, cerebral blood volume, metabolism, and blood oxygenation level caused by neuronal activities. The difference of magnetic susceptibilities of hemoglobins in different oxygenation states causes regional distortion of an MR decaying parameter T_2^*, leading to the changes in the intensity of an MR image. The localized signal changes observed through this BOLD contrast occurs in the range of several seconds and the rapid acquisition of images allows visualization of interesting brain dynamics. Another contrast mechanism used in FMRI is the perfusion contrast by arterial spin labeling (ASL), using magnetically labeled arterial blood water to measure rCBF. The BOLD-based FMRI measures the changes of T_2^* indirectly reflecting the changes in rCBF, whereas the ASL-based FMRI measures the changes of T_1 reflecting the changes in rCBF. For readable introductory materials to brain mapping with FMRI, the reader is referred to [10] and [12].

A typical blocked experiment in FMRI consists of two conditions, a functional condition inducing neuronal activities on certain regions in the human brain and a rest condition. In a simple experiment, when a visual stimulus is presented to a subject in an MRI scanner, the functional condition involves visual fixation on a flickering checkerboard image and the rest condition involves fixation on a non-flickering image, e.g., an image with a cross sign at the center. Typically, these two conditions are repeated during the experiment. This repetition, namely a pre-specified temporal stimulus, is usually described as a periodic pulse, a sequence of an "on-off" pattern, in which "on" indicates the functional condition and "off" represents the rest condition. As an alternative to the simple blocked design, an event-related FMRI experiment in which multiple simple stimuli are given to a subject during an experiment has been developed [14, 43].

While the subject in the MRI scanner is responding to the given stimuli, a sequence of images reflecting functional activities of the subject's brain is taken in rapid succession, usually every second. The sequence of images obtained by the experiment has four dimensions (4D), which consist of three spatial dimensions (3D) and one temporal dimension (1D). The basic element of these 3D images is called a voxel and a two-dimensional (2D) cross section of a 3D image is called a slice. A 2D slice is typically shown on x-y plane and different slices are denoted by different z coordinates. After the image reconstruction is done by inverse Fourier transform or its equivalent iterative methods, several operations in the pre-processing step are performed before FMRI data analysis. For example, since each slice in a 3D image is acquired at a different time, slice timing correction is necessary, usually done by temporal interpolation. The image registration to match the collected images containing subject's functional activities to a high resolution structural image of the subject, typically a T_1 weighted image, and the correction for motion artifacts are required as well before performing data analysis. The image normalization to transform the collected functional images into the standard Talairach space is also necessary in the case of multi-subject experiments.

It is known that the signal changes on MR images observed through the BOLD contrast between the functional state and the rest state is typically very small, from 2-5% at 1.5T, a moderate magnetic field strength, to approximately 15% at 4T, a strong magnetic field strength [12]. Therefore to obtain statistically valid results from FMRI observations such as an activation map to show localized responses of subject's brain to presented tasks, sophisticated and carefully designed statistical procedures are required. In FMRI data analysis, many statistical methods have been proposed until now, which can be categorized into two types, a model-driven approach and a data-driven approach. In the model-driven method, a model is assumed and FMRI measurements are fitted to that model. Using statistical hypothesis testing based on the fitted model, conclusions can be drawn. Since techniques in time series analysis have been applied to voxel time courses, the model-driven method is also known as a univoxel approach. In the data-driven method, no underlying model is assumed and some meaningful spatial or temporal components in the measurements are searched. These found components are expected to associate with some physiological processes of interest in the brain. The data-driven approach is useful when an appropriate model generating data is not available. However, computational demand is usually high and interpretations are not easily made from obtained results by the data-driven method. Among several data-driven methods, principal component analysis (PCA) and independent component analysis (ICA) serve as important roles in FMRI data analysis.

With rare exceptions, most of proposed methods, including model-driven and data-driven approaches, have been derived from a spatial independence assumption with implicit space and time separability. In this thesis, our attention focuses on the model-driven approach for a single subject. We now introduce important and challenging tasks in FMRI data analysis with the contributions of this thesis.

1.2 Challenging Tasks

1.2.1 Spatio-Temporal Detection Statistics

One of the most significant tasks in FMRI data analysis is probably creating an activation map to show localized brain responses to pre-specified temporal stimuli. This task is typically called "the activation study". Activation map is just a spatial plot of a detection statistic built up from a statistical hypothesis testing that can be conceptually described as follows. Suppose that we have voxels in a region of interest (ROI) and are interested in whether a particular voxel in the ROI is activated while the subject is performing given tasks. Then, we need to test if the collected data provide enough evidences to support the null hypothesis H_0 or the alternative hypothesis H_1, where

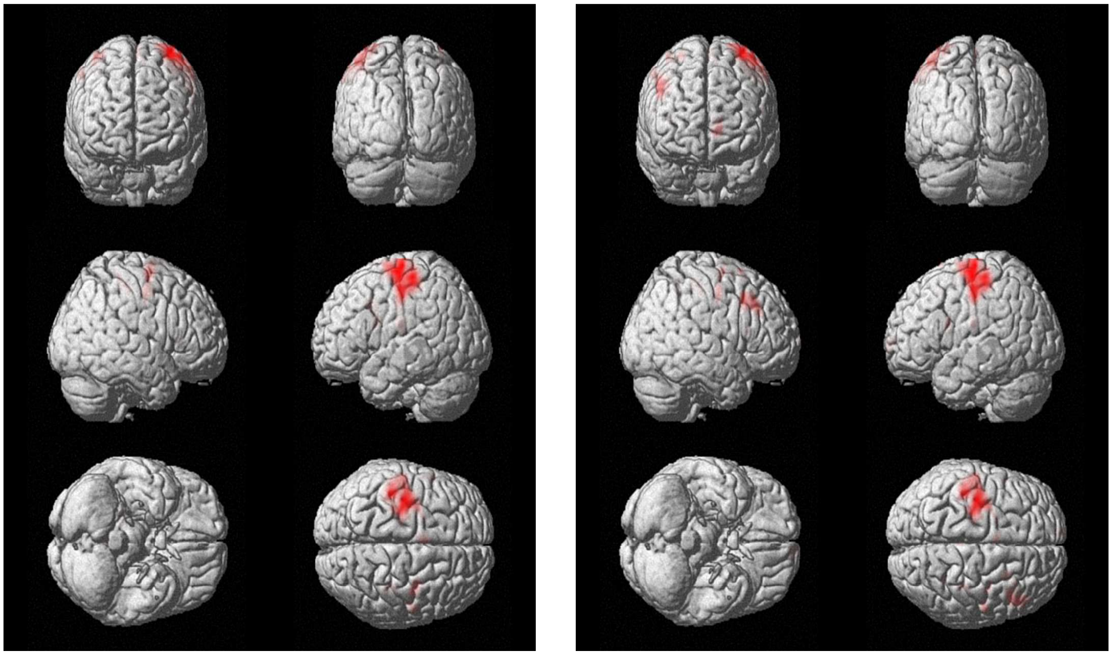
$$\begin{aligned} &H_0 : \text{the particular voxel is not activated,} \\ &H_1 : \text{the particular voxel is activated,} \end{aligned} \tag{1.1}$$

by given temporal stimuli. Now, we have an important question: how do we construct detection statistics for activation ?

To make a decision about the activation hypotheses, a thresholding rule is usually applied to obtained detection statistics. If we have an observation y_v at a voxel location v, called voxel time course, then the decision rule is

$$\Phi(y_v) = \begin{cases} 1 \quad , & \text{if} \quad T_v > \gamma(\alpha) \\ 0 \quad , & \text{otherwise} \end{cases} , \tag{1.2}$$

where T_v is a detection statistic based on the hypotheses in (1.1) and $\gamma(\alpha)$ is a threshold for a given significance level α, i.e., Type I error. By computing detection statistics over all voxels in a ROI and thresholding them with $\gamma(\alpha)$, we create a thresholded activation map, also known as a binary activation map. In the thresholded activation map, 1 is assigned to activated voxels and 0 is assigned to non-activated ones. Fig. 1.1 shows examples of two thresholded activation maps from different BOLD response modelings, parametric approach and FIR approach, rendered onto different anatomical views of the human brain. On Fig 1.1, red spots indicates that those voxels are activated, thus 1s are assigned to them. Notice that two methods show slightly different thresholded activation maps, indicating sophisticated statistical modeling is required in the activation study.



(a) Parametric approach (b) FIR approach

Figure 1.1: Examples of thresholded activation maps using two different BOLD response modelings : (a) parametric approach with a canonical hemodynamic response function and (b) FIR approach. Activation maps are generated from statistical parametric mapping (SPM) by simple regression models. For details of the BOLD response modeling, refer to section 1.3.2.

In Chapter 2, we review a standard approach to build up a detection statistic which only focuses on temporal aspects of voxel time courses under the implicit assumption of space-time separability. As a matter of fact, with rare exceptions, most previously proposed approaches have two main assumptions: **spatial independence** and **space-time separability**. Then, we introduce a new method to construct a detection statistic, jointly considering spatial and temporal correlations without space-time separability in Chapter 3. The removal of two assumptions which have been dominantly used is one of the main contributions of this thesis. Under the space and time separability, it will be shown that the new detection statistic has a simplified form and allows more intuitive interpretations, naturally leading to the conclusion that a spatial whitening operation is needed instead of spatial smoothing by a Gaussian kernel (SSK).

1.2.2 Spatio-Temporal Signal and Noise Modeling

Since measurements available for FMRI data analysis are spatiotemporal, a sequence of images, to build up a proper activation statistic, spatiotemporal signal and noise modeling is required. However, in multi dimensions, e.g., three dimensions or four dimensions in FMRI, since the fundamental theorem of algebra does not hold as in one-dimensional temporal case, noise modeling is a even more challenging problem than signal modeling. Thus, now we have an important question: how do we perform signal and noise modeling, in other words, how do we implement the constructed detection statistic for activation ?

To model spatial and temporal correlations, one possible approach is parametric spectral estimation, thus we need to estimate spatiotemporal power spectral density (PSD), i.e., spatiotemporal autocorrelation function (ACF). For this purpose, one natural approach would be using a spatiotemporal autoregressive (AR) based model which requires very complicated computations due to the non-linearity of asymptotic likelihood function. In multi dimensions, there does not exist an efficient algorithm like Levinson-Durbin algorithm in time series.

A simplest way to avoid this demanding task is assuming spatial independence as in conventional noise modeling approaches. However, the assumption of spatial independence is not supported by knowledge in neuroscience and physiology. Instead of AR-based model or ARMA model, in this thesis, the parametric cepstrum will be used to model spatial and temporal correlation, which allows several advantages, e.g., linear model fitting and linear description of space-time separability, over the conventional AR-based modeling. Details will be given in Chapter 4.

1.2.3 Family-Wise Error Rate Control

Since there are typically a large amount of voxels in a ROI, e.g., more than 4000 in a 64 × 64 image, controlling overall error rate is a multiple comparison problem (MCP). A widely used controlling measure is family-wise error (FWE) rate. The definition and an equivalent expression of FWE rate are given by, under H_0 (no activation),

$$\mathrm{FWE} \triangleq \Pr\left(\bigcup_{v=0}^{M-1} \{T_v > \gamma(\alpha)\} \,\middle|\, H_0 \right) = \Pr\left(\max_v T_v > \gamma(\alpha) \,\middle|\, H_0 \right), \tag{1.3}$$

where T_v is a detection statistic at a voxel $v(= 0, \ldots, M-1)$, M denotes the number of voxels in a ROI, and $\gamma(\alpha)$ is a threshold determined by a given significance level α, typically set to 0.05 in FMRI. Note that spatial correlation in an observed dataset induces spatial correlation of the detection statistic T_v. Since there is no closed form solution for the maximum distribution of T_v in the presence of spatial correlation, the statistical dependence between T_vs at different spatial locations introduces the main obstacle in MCP. Now, we have an important question: how do we compute a threshold $\gamma(\alpha)$ for a given significance level α ?

To control FWE in the presence of spatial correlation, a number of approaches have been proposed in FMRI so far. The most traditional method is Bonferroni correction which usually gives too conservative results. This conservativeness caused by Bonferroni correction becomes more severe when the spatial correlation of T_v becomes stronger. An alternative method suggested by [65] is random field theory (RFT), approximating the maximum distribution of T_v for a high value of $\gamma(\alpha)$. However, RFT requires assumptions on measured data, e.g., spatial smoothness of the observed data, which is against evidences in neuroscience. Details about RFT and spatial smoothing by a Gaussian kernel (SSK) to satisfy the requirements of RFT will be discussed in Chapter 2. As an alternative measure to FWE, false discovery rate (FDR) has been used with permutation test (PT) [38].

One advantage of the detection statistic proposed in this thesis is the simplification of the step for FWE rate control. Since the proposed detection statistic involves spatiotemporal whitening operation, detection statistics at different locations are asymptotically independent, thus not requiring complicated methods such as RFT. For details, refer to Chapter 3.

1.2.4 Space-Time Separability

Since space-time separability assumption allows a considerable amount of simplifications in analyzing and modeling of FMRI measurements, it has been assumed and applied to most of developed techniques for data analysis without the proper justifications. This assumption can be found in other literatures. For example in electromagnetism, to solve Maxwell's equations, the space-time separability has been usually assumed and given physically satisfactory solutions.

Conceptually, space-time separability implies that pure spatial operations and temporal operations can be separately applied to FMRI measurements to detect activations properly. For example, in SPM that is a widely used software package for FMRI data analysis, spatial smoothing by Gaussian kernel (SSK), a pure spatial operation, is first applied to the collected dataset. Then, general linear model (GLM), leading to purely temporal filtering, is sequentially performed to build up an activation map through t-test or F-test. For given data, however, the validity of the separability assumption is unknown without a proper test for it. To the best of author's knowledge, any testing for the space-time separability has so far not been treated properly in FMRI.

In the temporal frequency and spatial wave-number domain, **space-time separability** is defined in the following multiplicative form:

$$F_{k,l} = F_k G_l, \tag{1.4}$$

where $F_{k,l}$ denotes a spatiotemporal PSD, F_k is a pure temporal PSD, and G_l is a pure spatial PSD. A non-linearity of the space-time separability causes difficulty to develop a testing procedure for it. Now, we have an important question: how do we develop a test procedure for testing the space-time separability ?

We tackle this problem with the parametric cepstrum, since the space and time separability can be linearly described in the cepstral domain. One of the main contributions of this thesis is the development of space-time separability test and the analysis of the asymptotic power of the proposed separability test procedure for the first time in FMRI. In simulation and application to real data, it is shown that our separability test works properly. In addition, it turns out that the derived asymptotic power function involves cepstral coefficients only in the non-separable region, that are parameters of interest, and is independent of nuisance parameters such as drift and activation amplitudes. We discuss the details of the proposed separability test in Chapter 5.

1.2.5 Comparison of Competing Approaches

Another important task in FMRI data analysis is how to construct appropriate methods to compare different approaches to build up activation maps. Although many statistical methods for activation detection have been suggested in FMRI, the construction of an appropriate comparison criterion to rank different approaches have not attracted much attention. One example of ad hoc and naive ideas is to compare two approaches by looking at two generated activation maps, and then to claim that a method to have more smooth activation shapes (before thresholding) or to show more activation spots (after thresholding) is better. However, both claims can be significantly misleading. In this thesis, we suggest two methods for the comparison of different approaches to make activation statistics. The first is based on a model comparison technique, Akaike information criterion. The second is a method to compare the efficiencies of competing detection statistics for a given significance level and a detection power.

Model Comparison

We construct a properly formulated model comparison criterion using Akaike information criterion (AIC), which consists of a term measuring model fit and a term measuring model complexity. The AIC measures average discrepancy, an estimate of expected Kullback-Leibler (KL) distance, between fitted model and underlying unknown truth [57, 32]. Therefore, if AIC value from one model are substantially lower than that from another model, then one is on average much closer to the underlying truth than the other. Model selection can be dealt with as a special case of model comparison. To show localized discrepancy between fitted models and underlying noise-free truth, AIC maps, which are newly introduced in this thesis, are defined and compared. Details will be given in Chapter 4.

Performance Comparison of Detection Statistics

A well known method to compare two competing detection statistics is receiver operating characteristic (ROC) curve. In ROC curves, we can compare powers of detection statistics for a given false alarm rate. The comparison of an existing detection statistic involving spatial smoothing by Gaussian kernel and the proposed detection statistic is performed with ROC curves in Chapter 3.

Another possibility is to measure asymptotic relative efficiency of competing detection statistics. The philosophy behind asymptotic relative efficiency (ARE) is to measure relative sample sizes of two detection statistics to achieve a given power for a fixed false alarm rate. In FMRI data analysis, a more efficient test procedure in the sense of ARE reduces experiment time to achieve the same detection power than a less efficient test procedure, thus allowing several advantages, e.g., subject's reduced exposure time to strong magnetic fields. We derive ARE of an existing detection statistic involving spatial smoothing and the proposed detection statistic involving spatial whitening in Chapter 6. This asymptotic comparison is also new in FMRI data analysis.

1.2.6 Rician Modeling and Activation Detection

Most statistical methods for FMRI data analysis are based on magnitude voxel time courses and their approximation by a Gaussian distribution. As a matter of fact, since the magnitude images are originally produced from complex valued data, they obey a Rician distribution. A Rician probability density function (PDF) can be approximated as a Gaussian PDF under the assumption of high signal to noise ratio (SNR). Therefore, statistical methods based on the approximated Gaussian modeling may perform poorly when the SNR is low. It is known that high SNR assumption typically works well for FMRI data analysis with the BOLD response until now.

However, because of two reasons, we need Rician modeling which is SNR robust and Rician activation detection. Firstly, it is known that there is a fundamental tradeoff between SNR and spatial resolution. There have been many researches in FMRI whose main goal is to improve spatial resolution for obtaining more exact information about the brain, decreasing SNRs of voxels in a ROI. Secondly, even in images with a moderate spatial resolution, it is known that there may exist some regions in the brain which have significant signal dropouts, inducing that SNRs on those regions decrease. In addition, it is known that SNRs are not sufficiently high to apply a Gaussian approximation to a Rician distribution in approaches using non-BOLD contrast mechanisms, e.g., ASL-based FMRI. However, since Rician PDF involves the zeroth order modified Bessel function, parameter estimation by directly maximizing Rician log-likelihood function has been known to be non-trivial.

One of the main contributions of this thesis is developing a method to estimate parameters from a Rician distributed model through expectation-maximization (EM) algorithm, working properly regardless of underlying SNRs. Based on this SNR robust modeling, we can make any statistical inference, e.g., hypothesis testing for activation detection. Details of estimating parameters of interest and constructing a detection statistic from Rician distributed model are provided in Chapter 7. In addition, we analyze the asymptotic power of the proposed detection statistic for activation derived from a Rician distributed model there as well in Chapter 7.

1.3 Signal and Noise Model Formulation

We consider a real-valued measurement model which has the following additive form at a time point t and a voxel location v:

$$y_{t,v} = d_{t,v} + s_{t,v} + w_{t,v}, \tag{1.5}$$

where $d_{t,v}$ denotes drift slowly varying along time, $s_{t,v}$ means a signal component which we are interested in, the BOLD response, and $w_{t,v}$ denotes a zero mean random field, assumed to be a spatially and temporally correlated stationary Gaussian noise. The BOLD response $s_{t,v}$ models the brain response to the given temporal stimuli during the experiment. For integer-valued t and v, we assume that $y_{t,v}$ is observed from a rectangular region of interest, $\{0, \ldots, T-1\} \times \{0, \ldots, M-1\}$, where T is the number of time points and M is the number of voxels in a ROI. Detailed models of these three terms will be given in the following sections.

1.3.1 Drift Model

We first model the slowly varying drift. The simplest model using a temporal polynomial consists of two pieces,

$$d_{t,v} = m_v + b_v t, \tag{1.6}$$

where m_v is baseline and b_vt is linear drift, which are needed to model uncorrected motion artifacts and magnetic field inhomogeneity in a MR scanner. This can be easily extended to more complicated models. For example, a temporal polynomial with a high order or temporal sinusoids with low temporal frequencies can be considered. We have found that the linear drift is reasonably adequate [39]. In practice, it is known that the baseline dominates the temporal linear drift.

1.3.2 BOLD Response Model

Since the BOLD response is our main concern, an exact and simple model for it is necessary to model the observed signal as accurately as possible and to reduce computations. However, it's usually difficult to achieve these two goals simultaneously. A simple model tends to give biased estimates, and complex models are likely to give estimates with large variance and heavy computational load. Thus, a reasonable tradeoff is necessary. Our main focus will be on linear models for the BOLD response, i.e., the BOLD contrast is assumed linearly related to given stimuli. It is shown that the linear modeling is accurate to the first order [5]. We now review several linear models for the BOLD response.

The Parametric Approach

In the past, several approaches were proposed for modeling the BOLD response, $s_{t,v}$. The simplest one is so called the parametric approach in which the BOLD response is simply represented as $s_{t,v} = (h * c)_t f_v$ with a canonical hemodynamic response function (HRF), h_t and a scalar activation amplitude, f_v [11]. c_t denotes a temporal stimulus and h_t is fully specified and based loosely on experimental studies, namely;

$$h_t = k \; t^{8.6} exp(-t/0.546), \tag{1.7}$$

where k is a scaling factor to satisfy $\sum h_t^2 = 1$. This parsimonious parametric approach does not have enough flexibility. If h_t is not well matched to a particular dataset, the parametric approach shows poor performance, giving biased estimates of f_v and misleading activation detection. In fact, it is known that the canonical HRF is usually biased whereas nonparametric basis methods have much lower bias.

FIR Basis

For more flexible BOLD response modeling, subspace modeling has been used. In subspace modeling, the BOLD response is assumed to live in a space spanned by a set of basis functions. FIR approach is an example which has been used in FMRI [26, 9]. In the FIR approach, $s_{t,v}$ is modeled as the output of a FIR filter of a given order excited by an given input stimulus. The mathematical expression of the FIR approach is given by

$$s_{t,v} = \sum_{i=0}^{p-1} c_{t-i} f_{i,v}, \tag{1.8}$$

where $f_{i,v}$ denotes the i-th coefficient of a FIR filter and p is the order of the FIR filter at each voxel. In other words, the BOLD response is represented as a weighted linear combination of shifted temporal stimuli. Although the FIR approach provides flexibility and reduces biases, a high order FIR filter is typically necessary, decreasing the accuracy of parameter estimation [9]. In addition, as the number of voxels in a ROI increases, computational demands for parameter estimation and model selection to determine p sharply increases.

Laguerre Basis

An alternative method belonging to subspace modeling was recently developed for the BOLD response modeling. Laguerre polynomials were suggested to provide a set of basis functions to make an accurate and compact modeling of the BOLD response [56].

The Laguerre modeling is expressed as follows.

$$s_{t,v} = \sum_{i=1}^{q} \xi_{i,t} f_{i,v}, \tag{1.9}$$

where $\xi_{i,t}$ is the i-th temporal basis function obtained from the i-th Laguerre function, $f_{i,v}$ denotes the i-th weight associated with $\xi_{i,t}$ at a particular voxel v, and q is the model order. The method based on Laguerre polynomials usually needs 2-3 coefficients while the FIR approach requires 15-20 coefficients. Thus, using Laguerre basis, we can have an accurate and compact model for the BOLD response. The Laguerre functions are generated by the following inverse **Z** transformation:

$$h_{i,t}^{a} \triangleq \mathbf{Z}^{-1}\left[\frac{z^{-1}}{1-az^{-1}} \cdot \left(\frac{z^{-1}-a}{1-az^{-1}}\right)^{i-1}\right], \tag{1.10}$$

where a denotes a time constant to control decaying of Laguerre functions. Then, the temporal basis function $\xi_{i,t}$ can be obtained by the convolution of $h_{i,t}^{a}$ with temporal stimulus, that is $\xi_{i,t} \triangleq h_{i,t}^{a} * c_t$. Two crucial features of the basis functions obtained from Laguerre polynomials are that it ensures temporal causality and allows linear fitting [48]. For detailed discussions about the system identification with Laguerre polynomials, the reader is referred to [61].

A Generalized Expression

A generalized representation of $s_{t,v}$ covering the above mentioned three methods, the parametric approach, FIR approach, and Laguerre modeling, is given by

$$s_{t,v} = \left(\sum_{i=1}^{L} h_{i,t} f_{i,v}\right) * c_t \triangleq \sum_{i=1}^{L} \xi_{i,t} f_{i,v}, \tag{1.11}$$

where L is the number of basis functions, $h_{i,t}$ means the i-th temporal basis function, and $f_{i,v}$ represents the associated activation amplitude. Plugging (1.6) and (1.11) into (1.5) yields a compact and parameterized measurement model,

$$\begin{aligned} y_{t,v} &= m_v + b_v t + \sum_{i=1}^{L} \xi_{i,t} f_{i,v} + w_{t,v}, \\ &= X_t^T \beta_v + \xi_t^T f_v + w_{t,v}, \end{aligned} \tag{1.12}$$

where $X_t \triangleq [1, t]^T$, $\beta_v \triangleq [m_v, b_v]^T$, $\xi_t \triangleq [\xi_{1,t}, \ldots, \xi_{L,t}]^T$, and $f_v \triangleq [f_{1,v}, \ldots, f_{L,v}]^T$. The first term contains nuisance signal components such as temporally varying drift and models temporal non-stationary behaviors of FMRI measurements from the human brain. The second term represents the signal component, e.g., the BOLD response we are mainly interested in. These two terms are assumed deterministic but unknown. The last term denotes a spatiotemporal colored random noise that we discuss in the following section. In this thesis, we use this generalized model formulation.

1.3.3 Noise Model

Based on empirical evidences from the rest condition of the brain, the noise has been assumed to be generated from two independent sources. The first one is background noise from an MR scanner and the second one is physiological noise from low frequency hemodynamic fluctuations in the brain, which are not fully understood in neuroscience until now. This physiological noise is possibly related to background processes in the brain, as well as cardiac and respiratory fluctuations. In most cases, the cardiac and respiratory influences account for a small portion of the observed variabilities, because these are usually filtered out by preprocessing, e.g., high-pass filtering. The MR scanner noise can be modeled as a white noise and the physiological noise can be modeled as a spatiotemporally colored noise. There could be other sources of spatiotemporal correlation. For example, the image reconstruction performed in discrete k-space and techniques used to correct motion artifacts can induce spatial correlation.

Conventional Noise Model under Spatial Independence

By assuming spatial independence of noise in FMRI measurements, a substantial amount of simplification is made. Now, we can separately model and analyze each voxel time course using techniques developed in the time series literature. Under spatial independence, noise models based on autoregressive (AR) processes are widely used in FMRI [39, 8, 33]. For example, [39] exploits an AR(1) process plus white noise, equivalent to an ARMA(1,1) process, which is described as follows. For a given voxel position v,

$$w_{t,v} = u_{t,v} + \epsilon_{t,v}, \tag{1.13}$$

$$u_{t,v} = \varphi_v u_{t-1,v} + \delta_{t,v}, \tag{1.14}$$

where $\epsilon_{t,v} \sim \mathcal{N}(0, \sigma^2_{\epsilon_v})$, $\delta_{t,v} \sim \mathcal{N}(0, \sigma^2_{\delta_v})$, and $\epsilon_{t,v}$ and $\delta_{t,v}$ are statistically independent. φ_v is a coefficient for an AR(1) process. Note that this noise modeling is also under the assumption of space-time separability as other conventional noise modelings. By Wold's decomposition theorem, it can be shown that an ARMA(1,1) process has an equivalent expression as an AR(p) process of high order $p > 1$ [31].

Noise Model with Spatial and Temporal Correlations

Conventional approaches to describe noise in FMRI involve AR-based models as mentioned above, whereas a different class of models are used in this thesis to efficiently deal with spatial and temporal correlations without space-time separability. Our noise modeling approach is unusual and based on a truncated cepstrum expansion, but it allows several advantages over the conventional AR-based methods [55]. For example, the parametric cepstrum allows dramatic reduction of computations in modeling fitting. The parametric cepstrum is defined by truncating a Fourier series expansion of the logarithm of PSD as follows.

For temporal frequency $k(= 0, \ldots, T-1)$ and spatial wave-number $l(= 0, \ldots, M-1)$,

$$\log F_{k,l} = \sum_{t=-n}^{n} \sum_{v=-p}^{p} \theta_{t,v} e^{-j(\omega_k t + \lambda_l v)} \triangleq x^T_{k,l}\theta, \tag{1.15}$$

where $F_{k,l}$ is PSD, $\omega_k \triangleq \frac{2\pi k}{T}$, and $\lambda_l \triangleq \frac{2\pi l}{M}$. $\theta_{t,v}$ is called cepstrum, also known as cepstral coefficient at a time t and a spatial location v. In the last equality, θ denotes a vector containing lexicographically ordered cepstral coefficients and $x_{k,l}$ is a vector consisting of the associated cosine terms. Details of noise modeling by the parametric cepstrum are given in Chapter 4, in which the model order will be determined by model selection criterion.

1.4 List of Relevant Publications

In this section, we provide relevant publications to the discussions presented in this thesis.

Journal Papers

[J1] J. Noh and V. Solo, "FMRI with Joint Spatial and Temporal Correlations, Part I : Activation Detection", IEEE *Trans. on Medical Imaging*, Submitted.

[J2] J. Noh and V. Solo, "FMRI with Joint Spatial and Temporal Correlations, Part II : Signal and Noise Modeling", IEEE *Trans. on Medical Imaging*, Submitted.

[J3] J. Noh and V. Solo, "An EM Algorithm for Rician FMRI", *NeuroImage*, In preparation.

Conference Papers

[C1] J. Noh and V. Solo, "A True Spatiotemporal Approach for Activation Detection in Functional MRI", Proceedings of the 2006 IEEE International Conference on Acoustic, Speech, and Signal Processing, Toulouse, France, May 2006.

[C2] J. Noh and V. Solo, "A True Spatio-Temporal Test Statistic for Activation Detection in FMRI by Parametric Cepstrum", Proceedings of the 2007 IEEE International Conference on Acoustic, Speech, and Signal Processing, Honolulu, Hawaii, USA, April 2007.

[C3] J. Noh and V. Solo, "Testing for Space-Time Separability in Functional MRI", Proceedings of the 2007 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Washington DC, USA, April 2007.

[C4] V.Solo and J. Noh, "An EM Algorithm for Rician FMRI Activation Detection", Proceedings of the 2007 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Washington DC, USA, April 2007.

[C5] J. Noh and V. Solo, "Power Analysis for Testing Space-Time Separability in Functional MRI", Proceedings of the 2008 IEEE International Conference on Acoustic, Speech, and Signal Processing, Las Vegas, USA, March 2008, Submitted.

CHAPTER 2

Background : Standard Approach and Previous Works on Spatio-Temporal Modeling

In this chapter, we review conventional data analysis methods widely used in FMRI and some previous works on spatiotemporal modeling. The conventional data analysis methods involve spatial smoothing by a Gaussian kernel (SSK) and dynamic linear model (DLM), also known as general linear model (GLM). Section 2.1 describes a underlying model, methods to estimate parameters of interest, and how to construct a detection statistic for activation in standard approaches. In section 2.2, a standard method to control FWE rate with random field theory (RFT) is briefly reviewed. Section 2.3 has some previous works on spatiotemporal modeling. Chapter conclusions will be drawn in section 2.4.

2.1 Classical Statistics for Activation Detection : SSK-DLM

Widely used conventional approaches involving SSK and DLM have a common underlying model different from our model formulation in (4.2). The underlying model is given by, for a time point t and a voxel location v,

$$y_{t,v} = X_t^T \beta_v + \xi_t^T f_v^G + \eta_{t,v}, \tag{2.1}$$

where f_v^G is a Gaussian amplitude activation and $\eta_{t,v}$ is a temporally correlated but spatial independent stationary Gaussian field [67]. Other terms are the same as those in (4.2). Two main differences from our model formulation should be emphasized here. Firstly, our model formulation does not impose any assumption for activation amplitude f_v, whereas the conventional model assumes that activation amplitude has a Gaussian shape, f_v^G, artificially forcing smooth activations. Secondly, our model in (4.2) consider spatial correlation without space-time separability, whereas the conventional model ignores spatial correlation caused by physiological noise with the implicitly assumed space-time separability. These two key assumptions on the activation amplitude and on spatial independence play important roles to derive a detection statistic in conventional approaches as we now discuss details.

2.1.1 Spatial Smoothing by Gaussian Kernel (SSK)

The SSK is one of the main techniques used to create a detection statistic in conventional data analysis methods. The SSK simply means application of a Gaussian amplitude kernel to observed FMRI images, which is sometimes though of as a pre-processing step before DLM performed to build up a detection statistic for activation. As a matter of fact, it will be shown below that SSK is needed to increase detectibility of assumed Gaussian amplitude activations f_v^G. In D-dimensional space, an isotropic Gaussian amplitude kernel is defined as follows.

Definition 2.1 (D-dimensional isotropic Gaussian amplitude kernel)**.**

$$\phi_\sigma(\mathbf{x}) \triangleq (2\pi\sigma^2)^{-D/2} exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2}\right), \tag{2.2}$$

where $\|\mathbf{x}\|$ *is the* 2*-norm of a* D*-dimensional vector* $\mathbf{x}$ *and* σ *determines the width of the kernel.* ■

The width of a Gaussian amplitude kernel is usually expressed in terms of full width half maximum (FWHM) defined as a distance between two points whose values of a kernel are the half of the value at the origin. We have a simple relation between σ and FWHM, that is, FWHM $= \sigma\sqrt{8\log 2}$ in one direction. For an anisotropic Gaussian kernel, a FWHM in each direction can be similarly defined by an associated width. A 3D Gaussian amplitude kernel is used for volumetric data and 2D kernel is for cross-sectional slices. The Gaussian kernel has an attractive property, that the convolution of two Gaussian kernels gives another Gaussian kernel, namely,

$$\phi_{\sigma_1}(\mathbf{x}) \; * \; \phi_{\sigma_2}(\mathbf{x}) = \phi_{\sqrt{\sigma_1^2+\sigma_2^2}}(\mathbf{x}).$$

Activation Amplitude Model

Using a Gaussian kernel, a very simple activation amplitude model was suggested [67]. Based on this activation amplitude model, ignoring nuisance signal components, e.g., baseline and linear drift, a measurement at a fixed time point was defined as follows. For a spatial location $\mathbf{x}$,

$$\overbrace{y(\mathbf{x})}^{observation} = [\underbrace{h^G \cdot (2\pi\sigma_S^2)^{D/2}\phi_{\sigma_S}(\mathbf{x})}_{point\ spread\ function} * \overbrace{\phi_{\sigma_A}(\mathbf{x})}^{anatomical\ variability} + \underbrace{\frac{\varepsilon(\mathbf{x})}{\sqrt{n}}}_{white\ noise}]* \overbrace{\phi_{\sigma_R}(\mathbf{x})}^{reconstruction}, \tag{2.3}$$

where h^G is an activation amplitude and n is the number of subjects who participate in a FMRI experiment. $\phi_{\sigma_A}(\mathbf{x})$ describes anatomical variabilities of different subjects. Since we are only interested in data analysis with a single subject, $\sigma_A = 0$ and $n = 1$ are set. The important characteristics of the conventional observation model in (2.3) are summarized as follows.

- The Gaussian amplitude activation, $h^G \cdot (2\pi\sigma_S^2)^{D/2}\phi_{\sigma_S}(\mathbf{x})$ represents a spatial point spread function for a simple spark-like stimulus to subject's brain. Spatial sampling of this point spread function defines f_v^G in (2.1).

- The noise $\varepsilon(\mathbf{x})$ is assumed to be a stationary white Gaussian field. Effects of image reconstruction and additional pre-processing procedures, e.g., motion artifact correction and image registration, are described by $\phi_{\sigma_R}(\mathbf{x})$.
- Spatial correlation is induced by the convolution with $\phi_{\sigma_R}(\mathbf{x})$. Therefore, in fact, the observation is not under spatially independence any more.
- The model implicitly supposes that an activation amplitude, $h^G \cdot (2\pi\sigma_S^2)^{D/2}\phi_{\sigma_S}(\mathbf{x})$ has spatial continuity and derivatives with any order, which impose very strong assumptions on f_v^G in (2.1).

Activation Detectibility

For a single subject experiment, it is straightforward to compute signal component of the observed signal $y(\mathbf{x})$, resulting in

$$h^G \cdot (2\pi\sigma_S^2)^{D/2}\phi_{\sqrt{\sigma_S^2+\sigma_R^2}}(\mathbf{x}).$$

The noise component gives us an autocorrelation function which is proportional to

$$\phi_{\sigma_R}(\mathbf{x}) * \phi_{\sigma_R}(\mathbf{x}) = \phi_{\sqrt{2}\sigma_R}(\mathbf{x}).$$

According to [67], the SNR at the origin, the center of an activated voxel, is then proportional to the ratio of the values of signal component and square root of noise variance evaluated at the origin, thus

$$\text{SNR} \quad \propto \quad h^G \cdot \left(\frac{\sigma_S^2\sigma_R^2}{\sigma_S^2+\sigma_R^2}\right)^{D/2}, \tag{2.4}$$

where, SNR is maximized when the reconstruction width matches to the width of the assumed Gaussian amplitude activation, namely when $\sigma_R = \sigma_S$. Therefore, assuming σ_S is larger than σ_R, for the best detectibility of the Gaussian amplitude activation, we need to additionally smooth the observed images with a Gaussian amplitude kernel to match widths. SSK is based on this argument, basically the same idea as matched filter theorem (MFT) [44]. The same conclusion can be drawn using a spatial likelihood ratio test (LRT) under the assumption that noise is spatially independent and Gaussian distributed as in (2.3). In [53] which is one of the main papers for this activation amplitude model, using a spatial LRT, authors showed that a detection statistic requires the same spatial kernel as an assumed point spread function. In practice, another purpose of SSK is to obtain sufficient smoothness to apply RFT. In standard approaches involving SSK, to approximate a maximum distribution of detection statistics, RFT is often used in FWE rate control. Detailed reviews of FWE rate control using RFT will be given in section 2.2.

Discussions on SSK

Since a Gaussian amplitude activation is assumed in (2.1), a SSK is needed to increase activation detectibility. However, there are no experimental evidences for Gaussian amplitude activations. In other words, the selection of the Gaussian shape activation is ad hoc. In an extreme case, if we assume a sinc shape activation, MFT leads us to a spatial smoothing by a sinc amplitude kernel. In facts, evidences from the neuroscience literature are against the Gaussian amplitude activation assumption, especially, spatial continuity of activation amplitudes. For example, [25] showed that a SSK can artificially shift functional localizations and recommended to avoid a SSK for clinical purposes. We introduce a proper spatial kernel based on spatial correlation of observed images without any assumption on the activation amplitude in Chapter 3. Note that SSK dealt with spatial aspects of measurements in FMRI and DLM deals with temporal characteristics of observations. We now discuss details of DLM.

2.1.2 Dynamic Linear Model (DLM)

After applying a SSK to observed FMRI measurements, to build up a detection statistic for activation at a voxel location we are interested in, a widely used method is based on linear regression and somewhat inaccurately termed, general linear model (GLM) in which spatial independence of voxels is assumed. In this thesis, we prefer the more descriptive terminology, dynamic linear model (DLM). In the framework of DLM, each voxel time course is modeled and fitted separately, thus called a univariate approach. The DLM leads to many standard tests such as t-test and F-test.

The DLM for an observed voxel time series is built up as follows. For a voxel location v,

$$y_v = \mathbf{G} b_v + w_v, \tag{2.5}$$

where $v = 0, \ldots, M-1$, a $T \times p$ matrix $\mathbf{G}$ is called design matrix and w_v is a $T \times 1$ noise vector which obeys $\mathcal{N}(0, \Sigma_{w_v})$. A $p \times 1$ vector b_v contains parameters to be estimated, e.g., baseline and activation amplitudes. The design matrix $\mathbf{G}$ contains known information such as temporal basis functions for the BOLD response modeling. We take a simple example with a canonical HRF of the parametric approach in (1.7). From (4.2), for simplicity, we assume that baseline and linear drift are pre-filtered out. Then we have

$$y_{t,v} = s_{t,v} + w_{t,v} = \xi_t f_v + w_{t,v}, \tag{2.6}$$

leading to a DLM form in (2.5),

$$y_v = \begin{pmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_{T-1} \end{pmatrix} f_v + \begin{pmatrix} w_{0,v} \\ w_{1,v} \\ \vdots \\ w_{T-1,v} \end{pmatrix} = \xi f_v + w_v, \tag{2.7}$$

where $\xi \triangleq [\xi_0, \ldots, \xi_{T-1}]^T$ and the number of parameters p is 1.

Detection Statistics for Activation

From (2.5), we can construct t-test or F-test. Firstly, we review how t-test is built up. With the covariance matrix Σ_{w_v}, the generalized least square (GLS) estimate of b_v is given by

$$\widehat{b}_v = \left(\mathbf{G}^T \Sigma_{w_v}^{-1} \mathbf{G}\right)^{-1} \mathbf{G}^T \Sigma_{w_v}^{-1} y_v, \tag{2.8}$$

and it is well known that

$$\widehat{b}_v \sim \mathcal{N}(b_v, \left(\mathbf{G}^T \Sigma_{w_v}^{-1} \mathbf{G}\right)^{-1}). \tag{2.9}$$

Then, t-statistic at a voxel v is given by

$$\mathcal{T}_v \triangleq \frac{c^T \widehat{b}_v}{\sqrt{c^T \left(\mathbf{G}^T \widehat{\Sigma}_{w_v}^{-1} \mathbf{G}\right)^{-1} c}}, \tag{2.10}$$

where a $p \times 1$ column vector c is called contrast used to choose interesting parameters to be tested. Since the covariance matrix Σ_{w_v} is not known in practice, the statistic $\mathcal{T}_v$ requires to estimate b_v and Σ_{w_v} simultaneously. An iterative algorithm to do that is proposed in [57] for a temporal AR(1) process, which is not the same way used in SPM.

For a temporal AR(1) noise, SPM assumes covariance matrix has a form of $\Sigma_{w_v} = \sigma_{w_v}^2 \cdot \mathbf{V}$, which means the variance part of Σ_{w_v} depends on the voxel position v and the correlation part of Σ_{w_v} is common over all voxels. Since all voxel time series data can be used to estimate the matrix $\mathbf{V}$, it can be estimated more accurately than $\sigma_{w_v}^2$. Firstly, in SPM, the matrix $\mathbf{V}$ is estimated by restricted maximum likelihood (ReML) method [42]. And then, its estimate is used to estimate other parameters. These two steps are repeated to obtain the accurate estimates. It can be shown that $\mathcal{T}_v$ obeys a t-distribution with $T - p$ degrees of freedom, t_{T-p} under the null hypothesis H_0, that is, no activation.

To build up $\mathcal{T}_v$, [66] used ordinary least square (OLS) estimate of b_v instead of GLS estimate. The OLS estimate of b_v is given by

$$\widehat{b}_v = \left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T y_v \ \triangleq \ \mathbf{G}^- y_v, \tag{2.11}$$

and it is well known that

$$\mathrm{Var}(\widehat{b}_v) = \mathbf{G}^-\Sigma_{w_v}\mathbf{G}^{-T}. \tag{2.12}$$

Then, $\mathcal{T}_v$ is given by

$$\mathcal{T}_v \ \triangleq \ \frac{c^T\widehat{b}_v}{\sqrt{c^T\left(\mathbf{G}^-\Sigma_{w_v}\mathbf{G}^{-T}\right)c}} \quad \sim \quad t_\nu, \tag{2.13}$$

where an effective degrees of freedom ν is determined by Satterthwaite approximation [66].

To construct a F-test, we consider the following hypotheses. The two competing models are given by, for a voxel location v,

$$\begin{aligned} H_0 \ &: \ y_v = \mathbf{G}_r b_v^r + w_v, \\ H_1 \ &: \ y_v = \mathbf{G}_f b_v^f + w_v, \end{aligned} \tag{2.14}$$

where a $T \times q$ matrix $\mathbf{G}_r$ represents the design matrix of reduced model and a $T \times p$ matrix $\mathbf{G}_f$ means the design matrix of full model. By setting all activation to zeros, the design matrix of reduced model contains terms only related to nuisance signal components such as baseline and temporal linear drift, thus $q < p$. From (2.14), F-statistic, $\mathcal{F}_v$ is defined as

$$\mathcal{F}_v = \frac{(RSS^r - RSS^f)/p - q}{RSS^f/T - p} = \frac{\mathcal{H}_v/df_h}{\mathcal{E}_v/df_e}, \tag{2.15}$$

where RSS^r is the residual sum of squares (RSS) of the reduced model and RSS^f is the RSS of the full model. $\mathcal{H}_v$ means the hypothesis sum of squares with degrees of freedom df_h, $\mathcal{E}_v$ represents error sum of squares with degrees of freedom df_e. It is known that, when the null hypothesis H_0 is true, $\mathcal{F}_v$ obeys a F-distribution with df_h and df_e degrees of freedom.

Discussions on DLM

Although spatial correlation is induced by performing SSK, this induced spatial correlation is not properly considered to develop detection statistics in DLM. In other words, t-statistic or F-statistic is obtained as if all voxels in a ROI are statistically independent. Therefore, the detection statistic, $\mathcal{T}_v$ or $\mathcal{F}_v$, has spatial correlation, which will be reconsidered to control FWE rate in an ad hoc way in the next section.

Here, we need to make two more important comments on SSK-DLM. Firstly, no spatial correlation caused by physiological noise is considered in SSK-DLM. Secondly, SSK that is a purely spatial operation, and DLM that is a purely temporal operation are performed separately and sequentially under the assumption of space-time separability.

2.2 Family-Wise Error Rate Control by Random Field Theory (RFT)

Based on a detection statistic obtained in previous sections, $\mathcal{T}_v$ or $\mathcal{F}_v$, to create a binary activation map, we should consider all detection statistics in a ROI and threshold them with an appropriate cutoff point. Following the simplest method to control FWE rate, Bonferroni correction, we need to find a threshold $\gamma(\alpha)$ satisfying the last equality in

$$\text{FWE} \triangleq \Pr\left(\bigcup_{v=1}^{M}\{\mathcal{T}_v > \gamma(\alpha)\} \;\middle|\; H_0\right) \leq \sum_{v=1}^{M}\Pr\left(\mathcal{T}_v > \gamma(\alpha) \;\middle|\; H_0\right) = \alpha, \tag{2.16}$$

under the null hypothesis, i.e., no activation in a ROI. The inequality is called Bonferroni inequality. By selecting $\gamma(\alpha)$ satisfying $\Pr(\mathcal{T}_v > \gamma(\alpha)) = \alpha/M$ for all voxels, we can control the FWE less than α. The corrected significance level $\alpha_B = \alpha/M$ is called the Bonferroni correction. The main problem of this Bonferroni correction is that it provides too conservative results even for the case of independently and identically distributed samples. This problem becomes more severe, when $\mathcal{T}_v$s are not statistically independent or there are a large amount of voxels in a ROI. Since a detection statistic developed by SSK-DLM has spatial correlation, truly activated voxels can be ruled out by conservative Bonferroni correction.

To control FWE rate more accurately than Bonferroni correction, many alternatives have been suggested until now. Most alternatives depend on the maximum of detection statistics in a ROI due to the following identity:

$$\Pr\left(\bigcup_{v=1}^{M}\{\mathcal{T}_v > \gamma(\alpha)\} \;\middle|\; H_0\right) = \Pr\left(\max_v \mathcal{T}_v > \gamma(\alpha) \;\middle|\; H_0\right) = \alpha. \tag{2.17}$$

If we know a distribution of $\max_v \mathcal{T}_v$ under the null hypothesis H_0, $\gamma(\alpha)$ can be determined and the FWE is controlled. Based on this idea, a widely used method in FMRI is from random field theory (RFT) to approximate the $\max_v \mathcal{T}_v$ for a high $\gamma(\alpha)$. RFT is implemented in SPM and gives less conservative results than Bonferroni correction if a spatial plot of $\mathcal{T}_v$ is smooth. When the threshold $\gamma(\alpha)$ is high and a spatial plot of $\mathcal{T}_v$ is sufficiently smooth, according to RFT, we have the following approximation:

$$\text{FWE} = \Pr\left(\max_v \mathcal{T}_v > \gamma \;\middle|\; H_0\right) \approx \mathbf{E}[EC] \approx \sum_{d=0}^{D} Resels_d EC_d(\gamma), \tag{2.18}$$

where $Resels_d$ is the number of RESolution ELements, EC_d is Euler characteristic density in d dimensions, and D is the number of dimensions in a ROI [65]. The Euler characteristic density is a function of d and γ for a particular type of fields, e.g., Z-field, t-field and χ^2-field. When a threshold γ is high, the Euler characteristic can be thought as the counts of the number of activated clusters in which the values of detection statistics are larger than the threshold γ. For a ROI of large size, in the rightmost equation of (2.18), it is known that the D-th order term is dominant and the associated $Resels_D$ is given by

$$Resels_D = \frac{V}{\mathrm{FWHM}_e{}^D}, \tag{2.19}$$

where V is the volume of the ROI and FWHM_e is the effective full width half maximum of a Gaussian amplitude kernel used in SSK. In three-dimensional space, an Euler characteristic density for a t-distributed field with ν degrees of freedom is given by

$$EC_3(\gamma) \triangleq \frac{(4\log_e 2)^{3/2}}{(2\pi)^2}\left(\frac{\nu-1}{\nu}\gamma^2 - 1\right)\left(1+\frac{\gamma^2}{\nu}\right)^{-\frac{1}{2}(\nu-1)}. \tag{2.20}$$

Thus, for a given significance level α, with an estimate for an effective FWHM, a threshold γ can be determined by solving (2.20). For a small ROI, since other terms become important in the rightmost equation of (2.18), a correction is required.

Discussions on RFT

Since RFT is under the assumption that observed images are lattice representations of a underlying continuous random field, it requires sufficient smoothness to control FWE rate properly. For example, in a Gaussian distributed random field, a FWHM should be two or three times of a voxel size in each direction to obtain less conservative results than Bonferroni correction. In the case of t-distributed field with low degrees of freedom, a FWHM which is more than 10 times of a voxel size in a direction is required [21]. If images are not sufficiently smooth, the estimate would be biased and corrections are necessary. In addition, the approximation by RFT to $\max_v \mathcal{T}_v$ is valid only for a high γ. For more details of RFT, the reader is referred to [21] focusing on applications of RFT to FWE controls in FMRI and to [2] containing more mathematical materials about RFT, especially Gaussian RFT.

Recall that RFT is introduced to control FWE rate in FMRI data analysis when spatial correlation of detection statistics for activation exists. Conversely speaking, if we have spatial independence for finite samples or infinite samples (asymptotically), then the RFT is not necessary any more. In fact, the proposed detection statistic in Chapter 3 has asymptotic spatial independence under the null hypothesis H_0, allowing an easy method to control FWE rate unlike the standard approaches.

Permutation Test

Another method to control FWE rate used in FMRI is permutation test (PT), which does not require strong assumptions about a underlying distribution of the observed data unlike RFT [29]. Since PT is under the assumption that a voxel time course at a voxel v can be shuffled over time, the voxel time course should be temporally whitened before the application of PT. Since PT obtains an approximate distribution of $\max_v \mathcal{T}_v$ using simulations, it usually requires a tremendous amount of computations to compute a threshold for FWE rate control. PT is usually used with false discovery rate (FDR) instead of FWE. Some comparative studies between RFT and PT are provided in [38], in which PT shows better performance than RFT does when the assumptions of RFT are in doubt.

2.3 Previous Works on Spatio-Temporal Modeling

In this section, we discuss some previous works considering spatiotemporal dependence between voxels in FMRI data analysis. There have been made a much smaller amount of works to take care of spatiotemporal dependence than approaches to consider only temporal dependence as in the standard approaches. In addition to approaches discussed in section 2.3.1 (LRST modeling) and 2.3.2 (NN-ARx modeling), there are some recent works with discrete wavelet transforms [60, 34]. In [18], Markov random filed (MRF) is applied to the spatiotemporal modeling of FMRI measurements, in which an overall MRF prior is applied in space and time. However, since it is known that FMRI data have different spatial propeties from temporal ones, it does not seem to be attractive. Here, we need to emphasize that all these earlier works including the LRST and NN-ARx models do not provide an appropriate detection statistic for activation with the full consideration of spatial and temporal correlation, which is the main difference from our proposed detection statistic in Chapter 3.

2.3.1 Locally Regularized Spatio-Temporal (LRST) Model

In [57, 39], an approach to consider the spatial dependence of FMRI measurements is provided, which is called locally regularized spatio-temporal (LRST) modeling. The main idea of LRST modeling is to use an empirically known property of physiological noise, spatial continuity of noise. Note that this spatial continuity is completely different from the spatial continuity of activation amplitudes assumed in the underlying model for SSK-DLM. To recognize the spatial continuity of noise, [57] and [39] apply local spatial regularization to parameters for noise modeling. For temporal modeling, an AR(1) process in white noise is used. Therefore, the LRST model is a semiparametric model in the sense that the temporal specification is parametric, whereas the spatial specification is nonparametric.

For the local spatial regularization of the likelihood function of a voxel time course, a pre-determined spatial kernel whose form can be pyramid or cone, is applied. For example, a separable two-dimensional kernel which is supported on a finite neighborhood is given by

$$K_q^h \triangleq K\left(\frac{q_1}{M_1 h}\right) K\left(\frac{q_2}{M_2 h}\right) \frac{c_M}{M_1 M_2 h^2}, \tag{2.21}$$

where c_M is a correction term to make the sum of kernel weights unity and $q \triangleq (q_1,\ q_2)$ means pixel coordinates in the two-dimensional space. h is a smoothness parameter controlling the width of kernel. In [57, 39], Epanechnikov weighting kernel of quadratic polynomial form, namely, $K(u) \triangleq 0.75(1-u^2)_+$ is used. For three-dimensional space, this kernel formula can be easily extended.

To clearly explain the procedure used for LRST modeling, we first start the discussion without local spatial regularization. Assuming a Gaussian distributed noise, the following voxel-wise form of the negative log-likelihood function is made in the temporal frequency domain [7]: at a voxel location v,

$$J_v^0(\theta_v) = \sum_k \frac{I_{k,v}(\beta_v)}{2T \cdot F_k(\alpha_v)} + \sum_k \frac{\log F_k(\alpha_v)}{2}, \tag{2.22}$$

where the superscript 0 means no regularization, α_v is a column vector containing noise parameters, β_v is a column vector containing signal parameters, and $\theta_v^T \triangleq [\alpha_v^T\ \beta_v^T]$. F_k is a discrete PSD with temporal frequency index k and T is the number of time points. Here, the Periodogram $I_{k,v}(\beta_v)$ is defined as

$$I_{k,v}(\beta_v) \triangleq \frac{|\tilde{y}_{k,v} - \tilde{\mu}_{k,v}(\beta_v)|^2}{T}, \tag{2.23}$$

where $\tilde{y}_{k,v}$ is the temporally DFT transformed $y_{t,v}$ and $\tilde{\mu}_{k,v}(\beta_v)$ is the temporally DFT transformed signal component, $\mu_{t,v}(\triangleq m_v + b_v t + s_{t,v})$. Then, a weighted average of

$J_v^0(\theta_v)$ in (2.22) is obtained using the spatial kernel, yielding the following locally and spatially regularized log-likelihood function:

$$J_v(\theta_v) = \overbrace{\sum_q K_q^h J_{v-q}^0(\theta_v)}^{local\ regularization}. \tag{2.24}$$

Minimizing $J_v(\theta_v)$ with respect to parameters allows an estimate of θ_v at a voxel v. The minimization can be done by cyclic descent, yielding an algorithm that iterates between estimating signal parameters and noise parameters. This allows the noise and signal parameter subsets regularized separately. In [57, 39], the local spatial regularization is imposed on the AR(1) parameters only, using the spatial continuity to improve the estimation.

In LRST modeling, the information of local log-likelihood functions from neighboring voxels is used to estimate parameters of a given voxel v. Then, this parameter estimation repeats for the whole ROI one voxel by one voxel at a time. In [57], an activation map is quantified in terms of the square root of the weighted 2-norm of the estimated activation signal,

$$T_v \triangleq \sqrt{\sum_k \frac{|\tilde{s}_{k,v}|^2}{F_{k,v}}}, \tag{2.25}$$

where $\tilde{s}_{k,v}$ is the DFT transformed $s_{t,v}$ and $F_{k,v}$ denotes temporal PSD at a voxel v. For more details, the reader is referred to [57] and [39].

2.3.2 Near Neighborhood AutoRegressive Model with Exogenous Variable (NN-ARx)

Another approach to take account of the spatial dependence of FMRI measurements is suggested in [41], called near neighborhood autoregressive model with exogenous variable (NN-ARx). In the NN-ARx model, the neuronal process in the brain is assumed to comprise two components, deterministic evoked transient activity by pre-specified stimuli and spontaneous activity at the level of synapses. Originally motivated from a model presented in [23], the NN-ARx model is defined as, for a time point t and a voxel location t,

$$y_{t,v} = \overbrace{d_{t,v}}^{drift} + \underbrace{\sum_{k=1}^{p} \phi_{k,v} y_{t-k,v}}_{AR(p)} + \overbrace{\mathbf{X}_v \xi_{t-\Delta,v}}^{from\ neighborhood} + \underbrace{\sum_{k=0}^{r} \theta_{k,v} c_{t-k-d}}_{evoked\ transient\ activity} + \overbrace{\varepsilon_{t,v}}^{intrinsic\ activity}, \tag{2.26}$$

where $d_{t,v}$ denotes drift and the second term provides an AR(p) process for the modeling of the hemodynamic response of the brain. The third term describes the contributions from neighboring voxels, which is determined by

$$\xi_{t-\Delta,v} \triangleq \{y_{t-\Delta,v'}, v' \in \Omega_v\}, \tag{2.27}$$

where Ω_v is a pre-determined near neighborhood set of the v-th voxel, and a vector $\mathbf{X}_v$ describing the contributions of $\xi_{t-\Delta,v}$ to the voxel of interest. The fourth term represents the evoked activity by stimuli and the last term denotes the intrinsic and spontaneous activity in the brain. Δ is a mean delay of the contributions from neighboring voxels and d is a delay of a stimulus process c_t. In (2.26), thus, the AR(p) process is externally perturbed by the contributions of neighboring voxels and evoked transient activities by a given temporal stimulus.

If a temporal polynomial is used to model the drift, that is, $d_{t,v} = \sum_{k=0}^{\delta} \gamma_{k,v} t^k$, the parameters, $\Xi_v \triangleq \{\phi_{k,v}, \theta_{k,v}, \gamma_{k,v}, \sigma_v, \mathbf{X}_v\}$ must be estimated voxel by voxel from the data. The model selection consists of determining both model orders and delays which are global parameters, $\Lambda \triangleq (p, r, d, \Delta, \delta)$. In [41], the detection statistic[1] at the voxel v is defined as

$$\theta_v \triangleq \sum_{k=0}^{r} \theta_{k,v}, \tag{2.28}$$

meaning the spatial distribution of the brain synaptic sensitivity to pre-specified stimuli.

2.4 Conclusions

To construct detection statistics from FMRI measurements, the current standard method, SSK-DLM is built up on a underlying model which has two assumptions, that activation amplitudes are assumed to have Gaussian shapes and that noise is assumed to have spatial independence with space-time separability. These assumptions allow a substantial amount of simplifications in creating a detection statistic and implementing it, but are against evidences in physiology and neuroscience. In addition, although spatial correlation is induced by SSK, a detection statistic is first constructed under the assumption of spatial independence. Then, to control FWE rate, the induced spatial correlation is reconsidered in RFT, which imposes additional ad hoc assumptions on FMRI measurements.

The following chapter introduces a method to construct a properly formulated detection statistic for activation considering intrinsic spatial and temporal correlation from physiological sources without space-time separability and any specific assumption on activation amplitudes. It will be shown that, since our newly proposed method involves spatiotemporal whitening, FWE rate control can be easily performed and a complicated method such as RFT is not necessary.

---

[1] Authors call a spatial plot of detection statistics theta map

CHAPTER 3

True Spatio-Temporal Detection and Estimation I : Detection Statistics

A significant task in FMRI data analysis is creating an activation map to show localized brain responses to pre-specified temporal stimuli. An activation map is just a spatial plot of a detection statistic. These detection statistics have usually been derived from a spatial-wise independence assumption with implicit space-time separability. We develop, for the first time in FMRI, a properly formulated spatiotemporal detection statistic based on a spatially and temporally correlated noise model without space-time separability. We develop these new methods in this chapter and the following chapter. In this chapter, we develop the detection statistic and illustrate it in a simulation. In the subsequent chapter, we develop the joint signal and noise modeling necessary for full implementation, and the proposed methods are applied to a real human.

3.1 Introduction

3.1.1 Overview

Until recently, although many methods for data analysis in FMRI have been proposed, with rare exceptions, a voxel-wise model-driven approach of one-dimensional (1D) time series without a proper use of spatial correlation has been dominant in

FMRI [8, 33, 66, 22]. In other words, most attention has focused on the temporal characteristics of voxel time series. Under the assumptions of **spatial independence** and **space-time separability**, a widely used method is based on linear regression and somewhat inaccurately termed the general linear model (GLM) [66]. We prefer the more descriptive terminology, dynamic linear model (DLM). In the framework of DLM, each voxel time series is modeled and fitted separately. After interesting model parameters are estimated and test statistics, e.g., F-statistic, are computed over all voxels in a region of interest (ROI), it is determined which voxels are activated by a thresholding whose level is decided in a number of ways. The most traditional method is Bonferroni correction but an alternative method is random field theory (RFT) [65] or permutation test (PT) [38]. Note that the two main assumptions: spatial independence and space-time separability, have been dominantly assumed in data-driven methods as well.

In addition to temporal correlation, there have been some works attempting to use spatial dependence in the modeling. Locally regularized spatio-temporal (LRST) model was proposed for a temporal AR(1) process in white noise, equivalently a temporal ARMA(1,1) process, in [57, 39]. The suggested model is semi-parametric in the sense that the temporal specification is parametric while the spatial one is nonparametric. The information of local likelihood functions from neighboring voxels is used to estimate signal parameters and noise parameters of an ARMA(1,1) process at a given voxel. Near-neighborhood autoregressive with exogenous (NN-ARx) variable model was proposed in [41]. The model defined at a particular voxel has a term reflecting local dynamic contributions of neighboring voxels with an average delay. For each voxel, a set of near-neighborhoods is determined and the contributions of neighboring voxels in that set are estimated. Some details of these two methods were given in Chapter 2. Space-time simultaneous autoregressive (STSAR) model with a Bayesian prior was used in [64]. The proposed STSAR model combines a temporal AR(p) with $p > 1$ and a spatial AR(1) noise structure, which is heavily constrained. Some works with a spatial wavelet transform can be found in [60] and [34], and a method with a spatiotemporal Markov random field (MRF) was suggested in [18]. However, these earlier methods did not provide an appropriate detection statistic for activation with the full consideration of spatial and temporal correlation, but only focused on spatiotemporal modeling.

3.1.2 The Main Tasks and Organization

There are two main tasks to be carried out. Firstly, we derive a detection statistic in a setting allowing both spatial and temporal correlations without the assumption of space-time separability. As a byproduct, we are able to develop a test procedure for space-time separability whose details will be discussed in Chapter 5. Secondly, we develop a joint signal and noise modeling framework to allow complete implementation of the proposed detection statistic for activation. This is a considerable task in itself and is accomplished in Chapter 4 via the parametric cepstrum.

Throughout the remainder of this chapter, the organization is as follows. In section 3.2, we first review the widely used detection statistic derived from DLM and involving spatial smoothing by Gaussian kernel (SSK). Then we develop a new detection statistic derived from a spatially and temporally correlated noise model without space-time separability. In addition, a method of family-wise error (FWE) rate control is given. Under space-time separability, in section 3.3, a compact version of the developed detection statistic is shown, which allows more direct interpretations. We perform simulations to compare the new detection statistic with the widely used F-

statistic in section 3.4. Conclusions are drawn in section 7.7. Details of mathematical derivations are given in Appendix 3.6−3.8.

3.1.3 Acronyms and Notations

We collect here acronyms and notations which are frequently used in the rest of this chapter. CGD means complex-valued Gaussian distribution; CLT is central limit theorem; DFT means discrete Fourier transform; DLM is dynamic linear model; LRT means likelihood ratio test; MLE is maximum likelihood estimate; PSD means power spectral density; RFT is random field theory; ROI means region of interest; SSK is spatial smoothing by Gaussian kernel; ST-LRT means spatio-temporal LRT; STWK is spatio-temporal whitening kernel; SWK means spatial whitening kernel.

In the spatiotemporal frequency domain, integer k is used to denote the index of temporal frequency and integer l is assigned to the index of spatial wave-number. Thus, $k = 0, \ldots, T-1$ and $l = 0, \ldots, M-1$. $\tilde{y}_{k,l}$ denotes the spatiotemporal DFT of $y_{t,v}$. The definition of the DFT is given by, for $(\omega_k, \lambda_l) \triangleq (\frac{2\pi k}{T}, \frac{2\pi l}{M})$,

$$\tilde{y}_{k,l} \triangleq \sum_{t=0}^{T-1} \sum_{v=0}^{M-1} y_{t,v} \; e^{-j(\omega_k t + \lambda_l v)}. \tag{3.1}$$

Discrete spatiotemporal PSD is defined as $F_{k,l} \triangleq F(\omega_k, \lambda_l)$ from continuous PSD. Under the assumption of space-time separability, namely $F_{k,l} = F_k G_l$, F_k represents a purely temporal PSD and G_l means a purely spatial PSD.

For activation detection, the null hypothesis H_0 means no activations in a ROI and H_1 denotes the alternative hypothesis. Under H_j for $j = 0, 1$, $\widehat{F}_{j,k,l}$ denotes the estimate of $F_{k,l}$, allowing a STWK $g_{j,t,v}$. $\mathcal{F}_v$ represents the widely used F-statistic derived from SSK-DLM. $\mathcal{N}(\,\cdot\,,\,\cdot\,)$ means a real-valued Gaussian distribution, while $\mathcal{N}_c(\,\cdot\,,\,\cdot\,)$ denotes a complex-valued Gaussian distribution. The operator $**$ denotes a spatiotemporal linear convolution. $*$ is used for a temporal convolution and $*_s$ is for a spatial convolution. The circled asterisk, $\circledast\circledast$ represents a spatiotemporal circular convolution. Superscripts $(\,\cdot\,)^F$, $(\,\cdot\,)^*$, $(\,\cdot\,)^T$, and $(\,\cdot\,)^H$ denote a temporally filtered signal, a complex conjugate, a transpose, and a Hermitian transpose, respectively.

3.2 Detection of Activations

3.2.1 Classical Detection Statistic : SSK-DLM

For the modeling of activation amplitude f_v, a Gaussian point spread function (PSF) was proposed in [65]. Based on this Gaussian PSF and under a spatial white noise assumption, [53] showed that likelihood ratio test leaded to a matched filter involving spatial smoothing by Gaussian kernel (SSK), which is used in several software packages for FMRI data analysis, e.g., statistical parametric mapping (SPM). However, the modeling with Gaussian PSF which implicitly supposes spatial continuity of activations is not supported empirically. For instance, empirical evidence that SSK could artificially shift real activations was reported by [25]. Also the assumption of spatially white noise is not practically satisfactory.

After applying SSK to a collected dataset, at a given voxel v, the approach to time series decomposition is built up on the work of several researchers. DLM allowing the t or F-statistic was suggested by [66]. For example, the F-statistic is defined as

$$\mathcal{F}_v \triangleq \frac{\mathcal{H}_v/df_h}{\mathcal{E}_v/df_e}, \tag{3.2}$$

where $\mathcal{H}_v$ is the hypothesis sum of squares or signal sum of squares and $\mathcal{E}_v$ is error sum of squares; df_h and df_e denote associated degrees of freedoms. Remark that this SSK-DLM was developed under an implicit assumption of space-time separability.

The assumption of Gaussian PSF and DLM allow the use of RFT to control FWE rate [65]. Although the theory behind RFT is very complicated, it provides an approximate threshold, allowing a thresholded activation map. However, RFT provides more conservative results than does Bonferroni correction or permutation test for t-distributed fields with low smoothness and low degrees of freedom [38].

More details of SSK-DLM is provided in Chapter 2.

Given these drawbacks of SSK-DLM, in this chapter, we introduce an empirically more satisfactory activation detection considering temporal and spatial correlations of physiological noise without any specific assumption on spatial continuity of activation f_v and without space-time separability. It turns out that the new approach allows a simple analytical method to determine a threshold for a given significance level α, not requiring a complicated method as RFT.

3.2.2 New Detection Statistic : ST-LRT

The consideration of the spatiotemporal structure of the noise in the temporal frequency and spatial wave-number domains allows a substantial amount of simplification in building up an activation statistic. We proceed in 3 steps; first a general development; then an observation that the statistic can be decomposed into signal and noise parts; then a spatial decomposition.

General Development

Taking spatial and temporal DFTs of the FMRI measurement model in (4.2) gives an equivalent expression in the temporal frequency domain and the spatial wave-number domain,

$$\tilde{y}_{k,l} = \tilde{X}_k^T \tilde{\beta}_l + \tilde{\xi}_k^T \tilde{f}_l + \tilde{w}_{k,l}, \tag{3.3}$$

where, e.g., $\tilde{\xi}_k$ is the temporally DFT transformed ξ_t and $\tilde{f}_l$ represents the spatially DFT transformed f_v. For large T and M, under a spatiotemporal stationarity assumption and some other regularity conditions involving joint cumulants, $\tilde{w}_{k,l}$ obeys a CLT [7, Chap. 4], asymptotically leading to

$$\frac{1}{\sqrt{TM}} \cdot \tilde{w}_{k,l} \quad \sim \quad \mathcal{N}_c\left(0,\ F_{k,l}\right) \tag{3.4}$$

and $(\tilde{w}_{k,l})_{\substack{k=0,\dots,A \\ l=0,\dots,B}}$ are jointly CGD and independent for any (A, B) which belongs to the following region:

$$\begin{aligned} \{1 \leq A \leq (T-1)/2,\ 0 \leq B \leq M-1; \\ A = 0,\ 1 \leq B \leq (M-1)/2\}\,, \end{aligned} \tag{3.5}$$

where T and M are assumed odd for simplicity. In practice, to make T and M odd, we drop observations at the beginning of scans and on the edges of images. Note that the asymptotic distribution given by the CLT at the origin obeys a real-valued Gaussian distribution, that is, $\tilde{w}_{0,0} \sim \mathcal{N}(0,\ TM \cdot F_{0,0})$.

To develop an activation map, we consider the following hypotheses which are about a whole ROI:

$$\begin{aligned} H_0 &: f_v = 0 \qquad for\ all \quad v, \\ H_1 &: f_v \neq 0 \qquad for\ some \quad v. \end{aligned} \tag{3.6}$$

In the temporal frequency and spatial wave-number domains, we have equivalent hypotheses,

$$\begin{aligned} H_0 &: \tilde{f}_l = 0 \qquad for\ all \quad l, \\ H_1 &: \tilde{f}_l \neq 0 \qquad for\ some \quad l, \end{aligned} \tag{3.7}$$

where H_0 means no activations in the ROI and H_1 denotes the alternative. From the equivalent hypotheses in (3.7), with asymptotically independent CGD in (3.4), we can immediately construct a LRT statistic $\mathbf{\Lambda}$; its logarithm is given by

$$\begin{aligned} -2\log\mathbf{\Lambda} \;&\triangleq\; \sum_{k=0}^{T-1}\sum_{l=0}^{M-1} \log\widehat{F}_{1,k,l} - \log\widehat{F}_{0,k,l} \\ &+ \sum_{k=0}^{T-1}\sum_{l=0}^{M-1} \frac{\left|\tilde{e}_{1,k,l}\right|^2}{TM\cdot\widehat{F}_{1,k,l}} - \frac{\left|\tilde{e}_{0,k,l}\right|^2}{TM\cdot\widehat{F}_{0,k,l}}, \end{aligned} \tag{3.8}$$

where $\tilde{e}_{k,l}$s denote residuals; $\tilde{e}_{1,k,l} \triangleq \tilde{y}_{k,l} - \tilde{X}_k^T\widehat{\tilde{\beta}}_{1,l} - \tilde{\xi}_k^T\widehat{\tilde{f}}_l$ and $\tilde{e}_{0,k,l} \triangleq \tilde{y}_{k,l} - \tilde{X}_k^T\widehat{\tilde{\beta}}_{0,l}$ under H_0 and H_1, respectively. Under H_j for $j = 0,\,1$, $\widehat{F}_{j,k,l}$ is the MLE of $F_{k,l}$ and $\widehat{\tilde{\beta}}_{j,l}$ is the MLE of $\tilde{\beta}_l$; $\widehat{\tilde{f}}_l$ is the MLE of $\tilde{f}_l$ defined only under H_1. Note that the LRT in (3.8) involves the real-valued Gaussian distribution at the origin. The MLEs of parameters are given by

$$\widehat{\tilde{\beta}}_{1,l} = \widehat{\tilde{\beta}}_{10,l} - \mathbf{S}_{X\xi,l}\cdot\widehat{\tilde{f}}_l, \tag{3.9}$$

$$\widehat{\tilde{\beta}}_{0,l} = \left(\sum_{k=0}^{T-1}\frac{\tilde{X}_k^*\cdot\tilde{X}_k^T}{\widehat{F}_{0,k,l}}\right)^{-1}\left(\sum_{k=0}^{T-1}\frac{\tilde{y}_{k,l}\tilde{X}_k^*}{\widehat{F}_{0,k,l}}\right), \tag{3.10}$$

$$\widehat{\tilde{f}}_l = \left(\sum_{k=0}^{T-1}\frac{\tilde{\xi}_{1,k,l}^*\cdot\tilde{\xi}_{1,k,l}^T}{\widehat{F}_{1,k,l}}\right)^{-1}\left(\sum_{k=0}^{T-1}\frac{\tilde{y}_{1,k,l}\tilde{\xi}_{1,k,l}^*}{\widehat{F}_{1,k,l}}\right). \tag{3.11}$$

where

$$\widehat{\tilde{\beta}}_{10,l} \triangleq \left(\sum_{k=0}^{T-1}\frac{\tilde{X}_k^*\cdot\tilde{X}_k^T}{\widehat{F}_{1,k,l}}\right)^{-1}\left(\sum_{k=0}^{T-1}\frac{\tilde{y}_{k,l}\tilde{X}_k^*}{\widehat{F}_{1,k,l}}\right), \tag{3.12}$$

$$\mathbf{S}_{X\xi,l} \triangleq \left(\sum_{k=0}^{T-1}\frac{\tilde{X}_k^*\cdot\tilde{X}_k^T}{\widehat{F}_{1,k,l}}\right)^{-1}\left(\sum_{k=0}^{T-1}\frac{\tilde{X}_k^*\cdot\tilde{\xi}_k^T}{\widehat{F}_{1,k,l}}\right), \tag{3.13}$$

$$\tilde{y}_{1,k,l} \triangleq \tilde{y}_{k,l} - \tilde{X}_k^T\widehat{\tilde{\beta}}_{10,l}, \qquad \tilde{\xi}_{1,k,l} \triangleq \tilde{\xi}_k - \mathbf{S}_{X\xi,l}^T\tilde{X}_k. \tag{3.14}$$

Clearly, $\tilde{y}_{1,k,l}$ is just $\tilde{y}_{k,l}$ adjusted for drift under the alternative hypothesis, H_1. $\mathbf{S}_{X\xi,l}$ reflects the interaction between $\tilde{X}_k$ and $\tilde{\xi}_k$ under H_1. For detailed derivations of the MLEs, the reader is referred to Appendix 3.6.

Decomposition of LRT into Signal and Noise Pieces

Plugging (3.9) and (3.11) into (3.8) and canceling some common terms, allows a decomposition of $LRT (\triangleq 2 \log \mathbf{\Lambda})$ into two pieces; a noise piece (LRT^N) and a signal piece (LRT^S),

$$LRT \triangleq LRT^N + LRT^S, \tag{3.15}$$

where

$$LRT^N \triangleq \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} log \left(\frac{\widehat{F}_{0,k,l}}{\widehat{F}_{1,k,l}} \right) + \frac{\left| \tilde{y}_{0,k,l} \right|^2}{TM\widehat{F}_{0,k,l}} - \frac{\left| \tilde{y}_{1,k,l} \right|^2}{TM\widehat{F}_{1,k,l}}, \tag{3.16}$$

$$LRT^S \triangleq \frac{1}{TM} \cdot \sum_{l=0}^{M-1} \left(\sum_{k=0}^{T-1} \frac{\tilde{y}_{1,k,l} \tilde{\xi}_{1,k,l}^H}{\widehat{F}_{1,k,l}} \right) \left(\sum_{k=0}^{T-1} \frac{\tilde{\xi}_{1,k,l} \cdot \tilde{\xi}_{1,k,l}^H}{\widehat{F}_{1,k,l}} \right)^{-1} \left(\sum_{k=0}^{T-1} \frac{\tilde{y}_{1,k,l}^* \tilde{\xi}_{1,k,l}}{\widehat{F}_{1,k,l}} \right), \tag{3.17}$$

and

$$\tilde{y}_{0,k,l} \triangleq \tilde{y}_{k,l} - \tilde{X}_k^T \widehat{\tilde{\beta}}_{0,l}. \tag{3.18}$$

Note that, if $F_{k,l}$ is known, the noise piece LRT^N which reflects the difference between estimates of $F_{k,l}$s under H_0 and H_1 becomes zero, resulting in $LRT = LRT^S$. This follows because then $F_{0,k,l} = F_{1,k,l}$, $\tilde{\beta}_{10,l} = \tilde{\beta}_{0,l}$, and $\tilde{y}_{0,k,l} = \tilde{y}_{1,k,l}$. The signal piece LRT^S is computed only under H_1 and can be reformulated as follows.

$$LRT^S = \frac{1}{M} \cdot \sum_{l=0}^{M-1} \tilde{\mathbf{s}}_{y\xi,l}^H \mathbf{S}_{\xi\xi,l}^{-1} \tilde{\mathbf{s}}_{y\xi,l} \ , \tag{3.19}$$

where

$$\tilde{\mathbf{s}}_{y\xi,l} \triangleq \frac{1}{T} \cdot \sum_{k=0}^{T-1} \frac{\tilde{y}_{1,k,l}^* \tilde{\xi}_{1,k,l}}{\widehat{F}_{1,k,l}}, \quad \mathbf{S}_{\xi\xi,l} \triangleq \frac{1}{T} \cdot \sum_{k=0}^{T-1} \frac{\tilde{\xi}_{1,k,l} \cdot \tilde{\xi}_{1,k,l}^H}{\widehat{F}_{1,k,l}}. \tag{3.20}$$

By the spectral decomposition of $\mathbf{S}_{\xi\xi,l}$, we obtain

$$LRT^S = \frac{1}{M} \cdot \sum_{l=0}^{M-1} \left(\mathbf{S}_{\xi\xi,l}^{-\frac{1}{2}} \tilde{\mathbf{s}}_{y\xi,l} \right)^H \left(\mathbf{S}_{\xi\xi,l}^{-\frac{1}{2}} \tilde{\mathbf{s}}_{y\xi,l} \right), \tag{3.21}$$

where, for a unitary matrix $\mathbf{U}_l$ and a diagonal matrix $\mathbf{D}_l$,

$$\mathbf{S}_{\xi\xi,l} = \mathbf{U}_l^H \mathbf{D}_l \mathbf{U}_l, \qquad \mathbf{S}_{\xi\xi,l}^{\frac{1}{2}} \triangleq \mathbf{U}_l^H \mathbf{D}_l^{\frac{1}{2}} \mathbf{U}_l.$$

To provide a compact representation and to presage subsequent modeling, we introduce the zero-lag cepstral coefficient under H_1,

$$\widehat{\theta}_{1,0,0} = \frac{1}{TM} \cdot \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} \log \widehat{F}_{1,k,l}, \tag{3.22}$$

where $\widehat{\theta}_{0,0,0}$ is similarly defined from $\widehat{F}_{0,k,l}$ under H_0. In the sequel, we make much use of the STWK, $\tilde{g}_{1,k,l}$ which requires temporal causality and is defined as follows.

$$g_{1,t,v} \overset{DFT}{\longleftrightarrow} \tilde{g}_{1,k,l}, \qquad |\tilde{g}_{1,k,l}|^2 = \frac{1}{\widehat{F}_{1,k,l}}. \tag{3.23}$$

With $\widehat{F}_{0,k,l}$, we similarly define $g_{0,t,v}$ under H_0.

Spatial Decomposition of LRT

Using (3.21)-(3.23), we obtain an equivalent expression of LRT in the time and space domains through Parseval's relation. To be more specific, by defining $\tilde{\mathbf{s}}_{y,l} \triangleq \mathbf{S}_{\xi\xi,l}^{-\frac{1}{2}} \tilde{\mathbf{s}}_{y\xi,l}$, one arrives at the following representations:

$$\begin{aligned} LRT^N &= TM\left(\widehat{\theta}_{0,0,0} - \widehat{\theta}_{1,0,0}\right) + \sum_{t=0}^{T-1} \sum_{v=0}^{M-1} \left(\varepsilon_{0,t,v}^2 - \varepsilon_{1,t,v}^2\right), \\ LRT^S &= \sum_{v=0}^{M-1} \mathbf{s}_{y,v}^H \mathbf{s}_{y,v}, \end{aligned}$$

where $\varepsilon_{j,t,v}$ is spatiotemporally whitened $y_{j,t,v}$ under H_j for $j = 0, 1$, namely

$$\varepsilon_{j,t,v} \triangleq (g_{j,t,v} \circledast \circledast y_{j,t,v}), \tag{3.24}$$

and $\mathbf{s}_{y,v}$ denotes the inverse DFT of $\tilde{\mathbf{s}}_{y,l}$. Recall that $\circledast\circledast$ denotes a spatiotemporal circular convolution.

This enables us to decompose the statistic spatially as $LRT \triangleq \sum_v LRT_v$. We call LRT_v spatio-temporal LRT (ST-LRT) whose form is defined as

$$LRT_v \triangleq LRT_v^N + LRT_v^S, \tag{3.25}$$

where two pieces are defined as

$$LRT_v^N \triangleq T\left(\widehat{\theta}_{0,0,0} - \widehat{\theta}_{1,0,0}\right) + \sum_{t=0}^{T-1}\left(\varepsilon_{0,t,v}^2 - \varepsilon_{1,t,v}^2\right), \tag{3.26}$$

$$LRT_v^S \triangleq \mathbf{s}_{y,v}^H \mathbf{s}_{y,v}, \tag{3.27}$$

and

$$\mathbf{s}_{y,v} \overset{DFT}{\longleftrightarrow} \tilde{\mathbf{s}}_{y,l}, \qquad \tilde{\mathbf{s}}_{y,l} = \mathbf{S}_{\xi\xi,l}^{-\frac{1}{2}} \tilde{\mathbf{s}}_{y\xi,l}.$$

If the PSD, $F_{k,l}$ is known, then $F_{j,k,l}$'s are identical for all j and LRT_v^N vanishes, resulting in $LRT_v = LRT_v^S$, where $j = 0, 1$. For a voxel v, the spatially decomposed LRT_v represents a local contribution of that voxel to LRT testing hypotheses for a whole ROI in (3.6). Thus, by spatial decomposition, LRT_v is reasonably defined and has a nice intuitive interpretation.

In (3.25), the newly developed ST-LRT shows what kind of spatiotemporal operations are needed to build up a proper statistic for activation detection based on spatial and temporal correlations. Firstly, spatiotemporal whitening of $y_{j,t,v}$ under H_j is necessary, producing $\varepsilon_{j,t,v} \triangleq (g_{j,t,v} \circledast \circledast y_{j,t,v})$ for $j = 0, 1$. Notice that the STWK, $g_{j,t,v}$ is temporally causal and spatially non-causal. Secondly, according to the definition of $\tilde{\mathbf{s}}_{y\xi,l}$ and $\mathbf{S}_{\xi\xi,l}$, one needs the application of the STWK, $g_{1,t,v}$ to the purely temporal signal ξ_t. Thus, the filtered ξ_t is dependent on the index of spatial wave-number l. Since $g_{j,t,v}$ can be decomposed into a purely temporal filter and a purely spatial kernel under space-time separability, these interpretations becomes more direct and are discussed in section 3.3. Especially, a required purely spatial kernel K_v shows the characteristics of spatial operation needed for a proper detection statistic more clearly. It will turn out that spatial whitening is necessary instead of SSK which has been widely used in FMRI.

3.2.3 Family-Wise Error Rate Control

By computing ST-LRTs over all voxels within a ROI and thresholding them with a pre-determined cutoff point, we create a thresholded (binary) activation map. In a thresholded map, 1 is assigned to activated voxels and 0 is set to non-activated ones. To determine a threshold, we need to know the null distribution of ST-LRT. Under H_0, if $\tilde{\beta}_l$ and $F_{k,l}$ are known, we find the following asymptotic properties:

(P1) $\varepsilon_{0,t,v} \triangleq (g_{0,t,v} \circledast \circledast y_{0,t,v})$ is a spatial and temporal white noise with $\mathcal{N}(0,1)$ distribution.

(P2) ST-LRTs at different locations are independent, i.e., $LRT_u \overset{i.d.}{\sim} LRT_v$ for all $u \neq v$.

(P3) ST-LRT, $LRT_v \sim \chi^2_L$, i.e., a chi-square distribution with L degrees of freedom.

Detailed proofs of (P1)-(P3) are provided in Appendix 3.7. Since there are usually lots of voxels in a ROI, controlling overall error rate is a multiple comparison problem (MCP). A widely used measure to deal with the MCP is FWE (Type I error) rate whose definition and an equivalent expression are given by, under H_0,

$$\text{FWE} \triangleq \Pr\left(\bigcup_{v=0}^{M-1} \{LRT_v > \gamma\} \;\middle|\; H_0 \right) = \Pr\left(\max_v LRT_v > \gamma \;\middle|\; H_0 \right), \qquad (3.28)$$

where γ is a threshold to be determined by a pre-specified significance level α, typically set to 0.05 in FMRI. False discovery rate (FDR) can be an alternative choice to FWE [38]. Since, under H_0 by (P2) and (P3), LRT_v is spatially independent and follows χ^2_L, a threshold γ is determined analytically. For a given level α, a threshold $\gamma(\alpha)$ is given by

$$\gamma(\alpha) = \Psi_L^{-1}\left(\sqrt[M]{1-\alpha} \right), \qquad (3.29)$$

where $\Psi_L(t)$ denotes the cumulative density function of χ^2_L.

3.3 Space-Time Separability

In this section, we show how the activation statistic, ST-LRT simplifies when space-time separability holds. In Chapter 5, a test procedure for testing the space-time separability will be developed and its asymptotic power will be analyzed.

3.3.1 ST-LRT under Space-Time Separability

General Development

Under the assumption of space-time separability which is defined as, in the temporal frequency and spatial wave-number domains,

$$F_{k,l} = F_k G_l, \tag{3.30}$$

the ST-LRT in (3.25) has a simplified form. Specifically, plugging (3.30) into (3.16) and (3.17) and following a similar procedure to that which produced (3.25) yields a new ST-LRT,

$$LRT_v \triangleq LRT_v^N + LRT_v^S, \tag{3.31}$$

where

$$\begin{aligned} LRT_v^N &\triangleq T\left(\widehat{\theta}_{0,0,0} - \widehat{\theta}_{1,0,0}\right) + \sum_{t=0}^{T-1}\left(\varepsilon_{0,t,v}^2 - \varepsilon_{1,t,v}^2\right), \\ LRT_v^S &\triangleq \left(\sum_{t=0}^{T-1}\varepsilon_{1,t,v}(\xi_{1,t}^F)^T\right)\left(\sum_{t=0}^{T-1}\xi_{1,t}^F\cdot(\xi_{1,t}^F)^T\right)^{-1}\left(\sum_{t=0}^{T-1}\varepsilon_{1,t,v}\xi_{1,t}^F\right), \end{aligned} \tag{3.32}$$

and now under H_j for $j = 0, 1$,

$$\varepsilon_{j,t,v} \;\triangleq\; K_{j,v} \circledast_s (q_{j,t} \circledast y_{j,t,v}). \tag{3.33}$$

Here, $\widehat{\theta}_{j,0,0}$ is similarly defined under separability as in (3.22) and, since $\tilde{\xi}_{1,k,l}$ does not depend on the index of wave-number l any more, we drop the index l, thus $\tilde{\xi}_{1,k} \triangleq \tilde{\xi}_{1,k,l}$. The temporally whitened $\xi_{1,t}$ is defined as $\xi_{1,t}^F(\triangleq q_{1,t} \circledast \xi_{1,t})$, where $q_{j,t}$ is a causal temporal whitening filter, while $K_{j,v}$ is a non-causal SWK, which are given by

$$q_{j,t} \overset{DFT}{\longleftrightarrow} \tilde{q}_{j,k}, \qquad |\tilde{q}_{j,k}|^2 = \frac{1}{\widehat{F}_{j,k}}, \qquad K_{j,v} \overset{DFT}{\longleftrightarrow} \frac{1}{\sqrt{\widehat{G}_{j,l}}}, \tag{3.34}$$

under H_j. For details of the derivation, the reader is referred to Appendix 3.8.

In this simplified setting, more direct interpretations of LRT_v^S is available. Firstly, the temporal whitening of $\xi_{1,t}$ and the observed data $y_{1,t,v}$ are performed. Secondly, the application of a SWK to the temporally filtered data is necessary. Thirdly, cross correlation of the spatiotemporally whitened $y_{1,t,v}$ and the temporally whitened $\xi_{1,t}$ is required. Finally, a weighted norm of the computed cross correlation is formed.

New Spatial Kernel

Notice that the idea of a temporal whitening filter has been suggested by several researchers [8, 33]. However, the idea of SWK is introduced here for the first time in FMRI. This is totally different from the ad hoc approach of SSK. As a matter of fact, it can be easily checked that the SWK, K_v is not a smoothing kernel and is rather more like a spatial differentiator. Therefore, it is shown that ST-LRT requires an opposite operation, spatial differentiation, to SSK in the standard approach which is essentially spatial integration.

3.4 Simulation Studies

Simulations are carried out to compare the two competing detection statistics, the newly proposed ST-LRT, LRT_v and $\mathcal{F}_v$. $\mathcal{F}_v$ is the widely used F-statistic for activation detection from SSK-DLM. To simplify the discussion, without baseline and linear drift terms, simulated FMRI signals are constructed from (4.2), namely $y_{t,v} = \xi_t^T f_v + w_{t,v}$. We have four different cases determined by two types of f_v (activation amplitude) and two types of spatial correlation of $w_{t,v}$ (background noise). The common setups applicable to all four cases are described as follows.

(S1) At a given time t, a two-dimensional (2D) slice is considered, thus giving $v \triangleq (v_1, v_2)$.

(S2) The number of time points is assigned to $T = 99$ and the number of voxels in a ROI is set to $M = 63 \times 63$. The size of voxels is assumed 3.125×3.125 (mm^2) for a slice. These setup are basically matched to those of the real dataset used in Chapter 4.

(S3) The parametric approach of the BOLD response is used.

(S4) The noise obeys space-time separability, allowing $F_{k,l} = F_k G_l$.

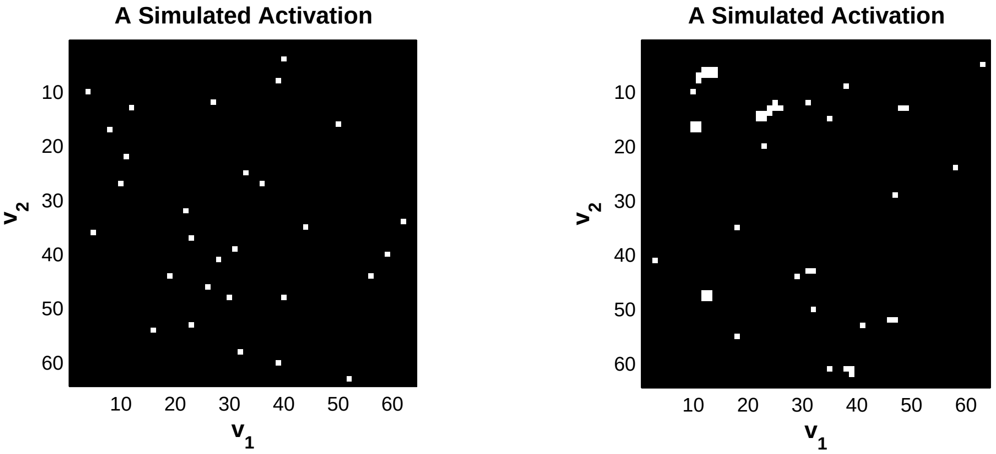
(S5) Spatially correlated and temporally white stationary Gaussian noise with known spectra is assumed.

To specify f_v, two types of activation amplitude are assumed, one of which is called random-position-fixed-shape (RPFS) activation whose position is random and shape is voxel-wise. The other is called random-position-random-shape (RPRS) activation whose position and shape are both random. Activated voxels have a fixed amplitude specified by BOLD-to-noise ratio (BNR) which is defined as BNR $\triangleq f_v/\sigma$, where σ^2

denotes the variance of $w_{t,v}$. Locations of RPFS activation are randomly selected from a 2D uniform distribution. RPRS activation is generated from a 2D standard Gaussian colored noise thresholded by an arbitrarily selected cut-off point. Examples of RPFS activation and RPRS activation are given on Fig.3.1.

To specify the spatial structure of $w_{t,v}$, both spatially white noise with $\sigma^2 = 1$ and spatially colored noise with a known autocorrelation function, namely $\gamma_v = exp(-v^2/2.254)$, are used. Since temporal whitening is also used in some previous approaches, for simplicity, temporally white noise is assumed in all four cases.

To perform SSK for $\mathcal{F}_v$, the full-width-half-maximum (FWHM) of Gaussian kernel is set as 2.5 times of voxel size along each axis as recommended in SPM. Based on the simulation setups, for a given significant level α, a threshold for LRT_v and a threshold for $\mathcal{F}_v$ are determined by (3.29) and RFT [65], respectively.



(a) RPFS activation

(b) RPRS activation

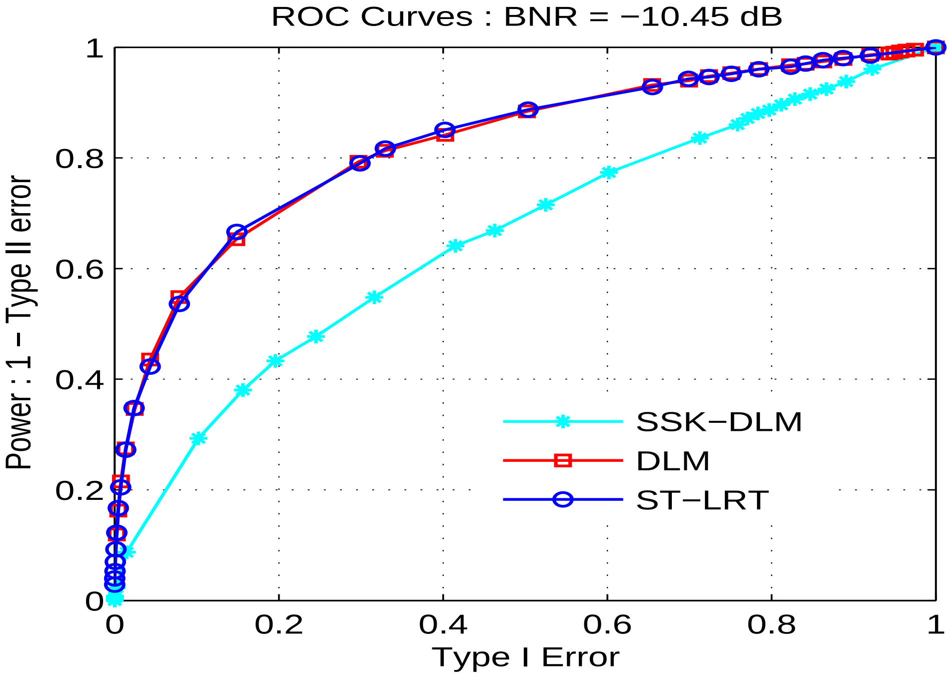
Figure 3.1: Examples of two types of activation amplitude: random-position-fixed-shape (RPFS) activation and random-position-random-shape (RPRS) activation. White dots and black dots indicate activated voxels and non-activated voxels, respectively.

The well-known receiver operating characteristic (ROC) curves are used as performance measure. In each ROC curve, for a pre-specified FWE (Type I error), we count the number of correct detections in an activation map generated by each detection statistic by comparing it with a true activation map. For one value on a ROC curve, 1.5×10^4 randomly chosen true activation spots are used. Then, we repeat this procedure for several values of FWE to obtain a ROC curve. In Fig.3.2 and Fig.3.3, LRT_v shows better performances than $\mathcal{F}_v$ with SSK for all cases and LRT_v is much better for spatially colored noise. In each graph, results from $\mathcal{F}_v$ without SSK are also shown for reference. As shown in Fig.3.2, $\mathcal{F}_v$ without SSK shows competeing performances with LRT_v in the case of spatially white noise. As a matter of fact, it can be shown that they are equivalent when the noise is temporally and spatially independent. Remark that, since sufficient smoothness is not guaranteed without SSK, we lose justifications for RFT to control FWE rate in unsmoothed $\mathcal{F}_v$.

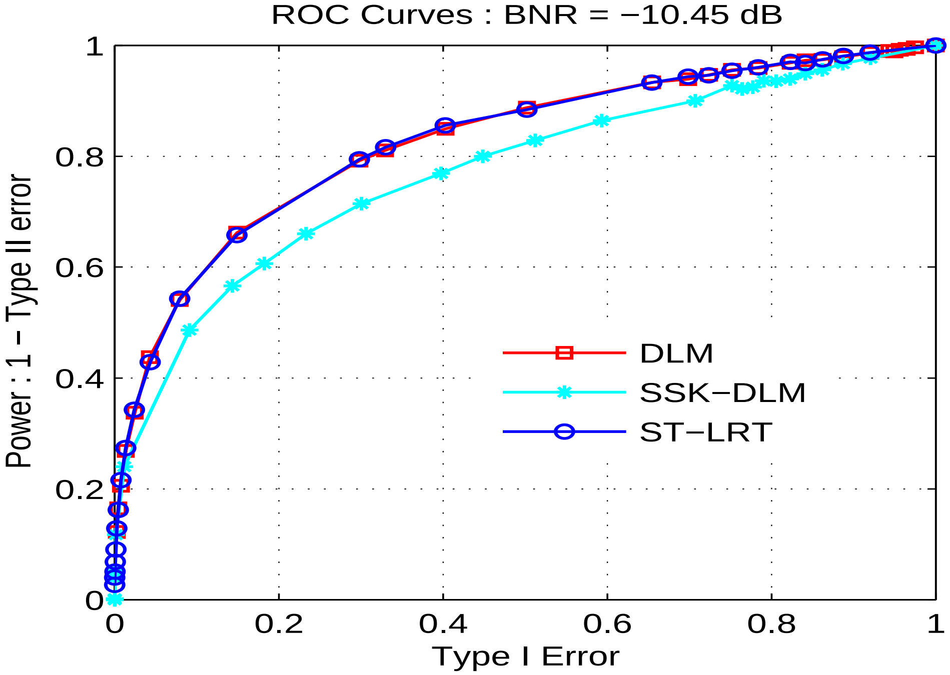
A method for joint signal and noise modeling using the parametric cepstrum is discussed in Chapter 4. An application of ST-LRT to a real human dataset involving the estimation of $F_{k,l}$ and a comparison of ST-LRT and F-statistic with Akaike information criterion (AIC) are performed there.

3.5 Conclusions

In this chapter, we built up a new detection statistic for activation with full consideration of spatial and temporal correlations of background noise without space-time separability. The developed ST-LRT required spatiotemporal whitening for non-separable elliptic field to create a properly formulated activation map. Under space-time separability, it was shown that spatial whitening kernel was necessary instead of Gaussian amplitude kernel for spatial smoothing. In simulations, according to comparisons through ROC curves, better performances of ST-LRT than the standard F-statistic were shown for several scenarios.

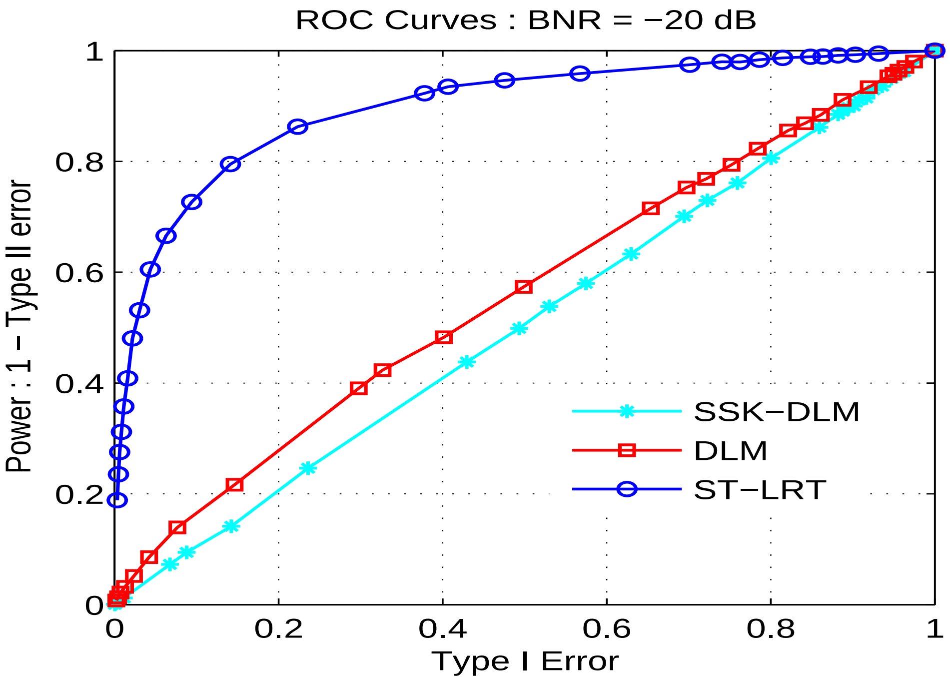


(a) RPFS and WN

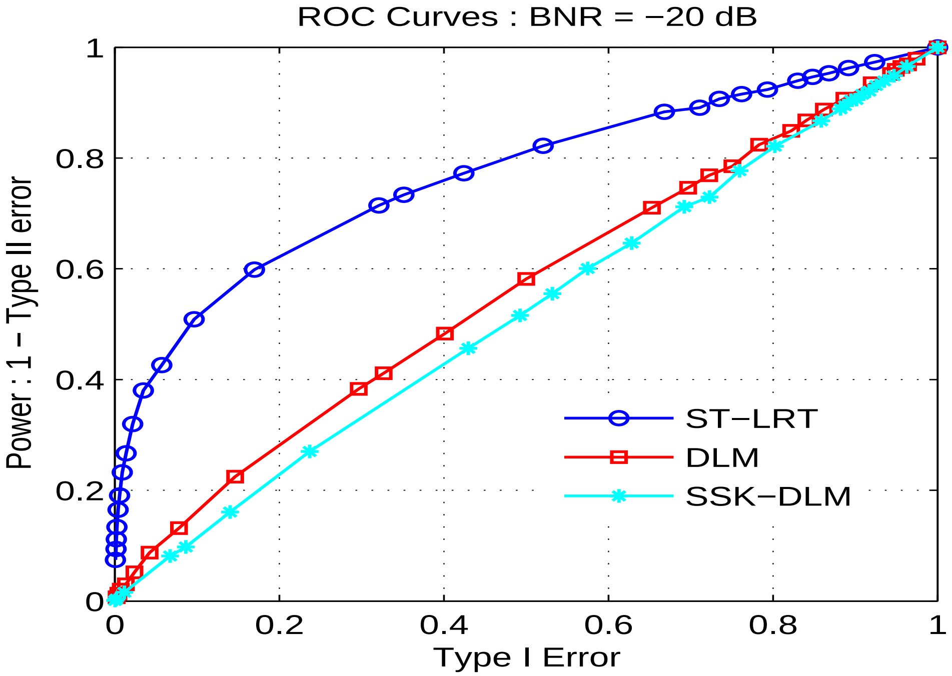


(b) RPRS and WN

Figure 3.2: ROC curves of LRT_v and $\mathcal{F}_v$ with and without SSK under the assumptions of random-position-fixed-shape (RPFS) and random-position-random-shape (RPRS) activations in spatially white noise (WN). BNR= -10.45 dB (i.e. BNR= 0.3) is set.



(a) RPFS and CN



(b) RPRS and CN

Figure 3.3: ROC curves of LRT_v and $\mathcal{F}_v$ with and without SSK under the assumptions of random-position-fixed-shape (RPFS) and random-position-random-shape (RPRS) activations in spatially colored noise (CN). BNR= −20 dB (i.e. BNR= 0.1) is set.

3.6 Appendix I : Derivations of MLEs

Taking the derivatives of $-2\log\mathbf{\Lambda}(\tilde{\beta}_{0,l},\tilde{\beta}_{1,l},\tilde{f}_l;\tilde{y}_{k,l})$ in (3.8) with respect to $\tilde{\beta}_{0,l}$, $\tilde{\beta}_{1,l}$, and $\tilde{f}_l$ and setting them zeros allow the desired estimates. Firstly, for $\tilde{\beta}_{0,l}$, one obtains an Euler equation,

$$\sum_{k=0}^{T-1}\frac{1}{F_{0,k,l}}\left(\tilde{y}_{k,l}-\tilde{X}_k^T\tilde{\beta}_{0,l}\right)^*\tilde{X}_k=0 \tag{3.35}$$

and then (3.10) after rearranging terms.

Secondly, for $\tilde{\beta}_{1,l}$, one similarly obtains an Euler equation,

$$\sum_{k=0}^{T-1}\frac{1}{F_{1,k,l}}\left(\tilde{y}_{k,l}-\tilde{X}_k^T\tilde{\beta}_{1,l}-\tilde{\xi}_k^T\tilde{f}_l\right)^*\tilde{X}_k=0. \tag{3.36}$$

By defining $\tilde{\beta}_{10,l}$ and $\mathbf{S}_{X\xi,l}$ whose definitions are given in (3.12) and (3.13), it is straightforward to show that (3.36) leads to

$$\widehat{\tilde{\beta}}_{1,l}=\tilde{\beta}_{10,l}-\mathbf{S}_{X\xi,l}\cdot\tilde{f}_l, \tag{3.37}$$

where $\mathbf{S}_{X\xi,l}$ reflects the interaction between $\tilde{X}_k$ and $\tilde{\xi}_k$ under the alternative hypothesis H_1.

Finally, for $\tilde{f}_l$, plugging (3.37) into $-2\log\mathbf{\Lambda}$ and following the same procedure as above allow an Euler equation,

$$\sum_{k=0}^{T-1}\frac{1}{F_{1,k,l}}\left(\tilde{y}_{k,l}-\tilde{X}_k^T\tilde{\beta}_{10,l}-\tilde{\xi}_{1,k,l}^T\tilde{f}_l\right)^*\tilde{\xi}_{1,k,l}=0 \tag{3.38}$$

by defining $\tilde{y}_{1,k,l}$ and $\tilde{\xi}_{1,k,l}$ in (3.14). Then rearranging terms gives the estimate of $\tilde{f}_l$ in (3.11). ■

3.7 Appendix II : Properties of Null Distribution of LRT_v

We provide mathematical proofs of three properties of LRT_v under H_0, (P1)−(P3) in this section. Remark that H_0 means there is no activations in a ROI and we assume that $\tilde{\beta}_l$ and $F_{k,l}$ are known, resulting in $LRT_v = LRT_v^S$ and $y_{j,t,v}$s are all identical regardless of j.

Proof of (P1)

Under H_0, $y_{0,t,v}$ has only the noise part which is stationary Gaussian field with zero mean. Thus, it is obvious that $\varepsilon_{0,t,v} \triangleq (g_{0,t,v} \circledast \circledast y_{0,t,v})$ is Gaussian and zero mean. Since $g_{0,t,v}$ is a spatiotemporal whitening filter, the output is spatiotemporal white noise. For computing the variance, one notices that

$$var(g_{0,t,v} \circledast \circledast y_{0,t,v}) = \frac{var(\tilde{g}_{0,k,l}\tilde{y}_{0,k,l})}{TM} = \frac{|\tilde{g}_{0,k,l}|^2 var(\tilde{y}_{0,k,l})}{TM} = \frac{var(\tilde{y}_{0,k,l})}{TM \cdot F_{k,l}} = 1, \quad (3.39)$$

where the last equality holds due to CLT. Therefore, it is shown that $(g_{0,t,v} \circledast \circledast y_{0,t,v}) \sim \mathcal{N}(0,1)$ asymptotically. ■

Proof of (P2)

It is an equivalent to show $\mathbf{s}_{y,u}$ and $\mathbf{s}_{y,v}$ are independent for all $u \neq v$. We consider $cov(\mathbf{s}_{y,u}, \mathbf{s}_{y,v})$ which is

$$cov(\mathbf{s}_{y,u}, \mathbf{s}_{y,v}) = \frac{1}{M^2} \sum_{l_1,l_2} cov(\tilde{\mathbf{s}}_{y,l_1}, \tilde{\mathbf{s}}_{y,l_2}) e^{j\frac{2\pi}{M}(ul_1 - vl_2)}. \quad (3.40)$$

Using $\tilde{\mathbf{s}}_{y,l} \triangleq \mathbf{S}_{\xi\xi,l}^{-\frac{1}{2}} \tilde{\mathbf{s}}_{y\xi,l}$ and (3.20), it can be shown that

$$cov(\tilde{\mathbf{s}}_{y,l_1}, \tilde{\mathbf{s}}_{y,l_2}) = \mathbf{S}_{\xi\xi,l_1}^{-\frac{1}{2}} cov(\tilde{\mathbf{s}}_{y\xi,l_1}, \tilde{\mathbf{s}}_{y\xi,l_2}) \mathbf{S}_{\xi\xi,l_2}^{-\frac{H}{2}} \quad (3.41)$$

$$cov(\tilde{\mathbf{s}}_{y\xi,l_1}, \tilde{\mathbf{s}}_{y\xi,l_2}) = \frac{1}{T^2} \sum_{k_1,k_2} \left(\tilde{g}_{1,k_1,l_1}\tilde{\xi}_{1,k_1,l_1}\right) cov(\tilde{\varepsilon}^*_{1,k_1,l_1}, \tilde{\varepsilon}^*_{1,k_2,l_2}) \left(\tilde{g}_{1,k_2,l_2}\tilde{\xi}_{1,k_2,l_2}\right)^H. \quad (3.42)$$

By CLT, one arrives at

$$cov(\tilde{\varepsilon}^*_{1,k_1,l_1}, \tilde{\varepsilon}^*_{1,k_2,l_2}) = TM \cdot \delta_{k_1-k_2, l_1-l_2}. \quad (3.43)$$

Plugging (7.53), (7.63) and (3.43) into (7.52) gives

$$cov(\mathbf{s}_{y,u}, \mathbf{s}_{y,v}) = \delta_{u-v} \cdot \mathbf{I}_{L\times L}, \quad (3.44)$$

where $\mathbf{I}_{L\times L}$ is an identity matrix whose size is $L \times L$. ■

Proof of (P3)

According to the definition in (3.20), $\tilde{\mathbf{s}}_{y\xi,l}$ is CGD with zero mean under H_0. By the intermediate result in *Proof of* (P2), it can be easily checked that

$$cov(\tilde{\mathbf{s}}_{y,l_1}, \tilde{\mathbf{s}}_{y,l_2}) = \delta_{l_1-l_2} \cdot M \cdot \mathbf{I}_{L\times L}. \tag{3.45}$$

Therefore, $\tilde{\mathbf{s}}_{y,l} \sim \mathcal{N}_c(0, M\cdot \mathbf{I}_{L\times L})$ and $\tilde{\mathbf{s}}_{y,0} \sim \mathcal{N}(0, M\cdot \mathbf{I}_{L\times L})$, which are equivalent to $\mathbf{s}_{y,v} \sim \mathcal{N}(0, \mathbf{I}_{L\times L})$. Then, one arrives at $LRT_v = \mathbf{s}_{y,v}^H \mathbf{s}_{y,v} \sim \chi_L^2$. ■

3.8 Appendix III : Derivation of ST-LRT under Separability

Since the noise piece, LRT_v^N has the same form as that from the non-separable case, we mainly concentrate on the signal piece, LRT_v^S. Under separability, since two $\widehat{G}_{1,l}$s in (3.13) are canceled out, $\mathbf{S}_{X\xi,l}$ is independent of the index of spatial wavenumber l. By dropping the index l, this gives $\tilde{\xi}_{1,k} \triangleq \tilde{\xi}_k - \mathbf{S}_{X\xi}^T \tilde{X}_k$. Then, by canceling out $\widehat{G}_{1,l}$ in the center term, (3.17) has the following form :

$$LRT^S = \frac{1}{TM} \cdot \sum_{l=0}^{M-1} \left(\sum_{k=0}^{T-1} \frac{\tilde{y}_{1,k,l}\tilde{\xi}_{1,k}^H}{\widehat{F}_{1,k}\sqrt{\widehat{G}_{1,l}}} \right) \left(\sum_{k=0}^{T-1} \frac{\tilde{\xi}_{1,k}\cdot \tilde{\xi}_{1,k}^H}{\widehat{F}_{1,k}} \right)^{-1} \left(\sum_{k=0}^{T-1} \frac{\tilde{y}_{1,k,l}^{*}\tilde{\xi}_{1,k}}{\widehat{F}_{1,k}\sqrt{\widehat{G}_{1,l}}} \right), \tag{3.46}$$

where notice that the center term which is defined as $\mathbf{S}_{\xi\xi}^{-1}$ is also independent of l. With an abuse of notations, we redefine $\tilde{\mathbf{s}}_{y\xi,l}$ and $\mathbf{S}_{\xi\xi}$ as

$$\tilde{\mathbf{s}}_{y\xi,l} \triangleq \frac{1}{T}\sum_{k=0}^{T-1} \frac{\tilde{y}_{1,k,l}^{*}\tilde{\xi}_{1,k}}{\widehat{F}_{1,k}\sqrt{\widehat{G}_{1,l}}}, \qquad \mathbf{S}_{\xi\xi} \triangleq \frac{1}{T}\sum_{k=0}^{T-1} \frac{\tilde{\xi}_{1,k}\cdot\tilde{\xi}_{1,k}^H}{\widehat{F}_{1,k}}. \tag{3.47}$$

Then, by Parseval's relation, it can easily be found that

$$LRT^S = \sum_{v=0}^{M-1} \mathbf{s}_{y\xi,v}^T \mathbf{S}_{\xi\xi}^{-1} \mathbf{s}_{y\xi,v}, \tag{3.48}$$

leading to $LRT_v^S \triangleq \mathbf{s}_{y\xi,v}^T \mathbf{S}_{\xi\xi}^{-1} \mathbf{s}_{y\xi,v}$ by a spatial decomposition. Again, by Parseval's relation, one can obtain expressions for $\mathbf{s}_{y\xi,v}$ and $\mathbf{S}_{\xi\xi}$ in the time domain, allowing the desired expression for LRT_v^S under the separability. ■

CHAPTER 4

True Spatio-Temporal Detection and Estimation II : Signal and Noise Modeling

In Chapter 3, for the first time in FMRI, we developed a detection statistic, fully considering spatial and temporal correlations without space-time separability. In this chapter, we develop joint signal and noise model fitting necessary to implement the proposed detection statistic, spatio-temporal likelihood ratio test (ST-LRT). Our noise modeling approach is unusual, being based on a truncated cepstrum expansion, but it allows dramatic reduction of computations in model fitting and a very simple method to obtain a desired spatiotemporal whitening operator. In addition, a method of model comparison is developed to compare ST-LRT and an existing approach, F-statistic derived from SSK-DLM. The developed techniques are applied to a human dataset.

4.1 Introduction

The implementation of ST-LRT requires joint signal and noise modeling. Noise modeling in multi dimensions, for example, three dimensions (3D) or four dimensions (4D) in FMRI, requires a non-trivial statistical model estimation and involves additional issues such as edge effects [13], compared to that of time series in one dimension (1D). Therefore, the noise modeling itself is a considerable task in FMRI. We start our discussion from the literature review of noise modeling in multi dimensions.

4.1.1 Literature Review

For a Gaussian distributed field under the assumption of stationarity, a widely used method for noise modeling is AR-based parametric spectral estimation. In 1D, for example, the AR model and its extensions have a lot of applications in the time series literature [52]. In two-dimensional (2D) space, a spatial AR model of semi-causality [62] and one of non-causality with some boundary conditions [51] were suggested. [30] is a readable reference for three types of 2D spatial AR-based approaches, which are causal, semi-causal, and non-causal. In 2D spatial and 3D/4D spatiotemporal spectral estimation, however, the fundamental theorem of algebra (FTA) does not hold as in 1D temporal case. Therefore, since polynomials do not generally factor, the asymptotic likelihood equations for fitting a spatial AR or ARMA model can not be solved linearly as in the 1D. As a matter of fact, the existence of a 2D spatial solution is not guaranteed as shown in [19]. Assuming a known 2D spatial spectrum, an approach to solve the factorization problem through homomorphic transform was suggested in [20]. However, the 2D spatial spectrum should be estimated before applying the suggested method.

The parametric cepstrum was proposed to solve the 2D spatial spectral estimation problem by [55]. It was shown that modeling by 2D spatial parametric cepstrum had several advantages over spatial AR-based modeling of random fields, particularly providing very fast and mostly linear computations for model fitting by fast Fourier transforms.

4.1.2 The Main Tasks and Organization

Two main issues to implement ST-LRT are addressed in this chapter. The first issue is a statistically non-trivial joint signal and noise modeling in space and time. For noise model fitting in 3D, namely, 2D space and 1D time, the spatiotemporal parametric cepstrum which allows quick and easy model fitting is proposed. The originally proposed 2D spatial parametric cepstrum by [55] is extended to 3D spatiotemporal one in FMRI. Secondly, a new method of model comparison based on the proposed cepstral noise modeling is discussed. We construct an Akaike information criterion (AIC) and show how it can be decomposed spatially to provide an AIC map. The AIC map can be used to compare models, specifically we compare the ST-LRT approach with a standard approach based on SSK-DLM.

The rest of this chapter is organized as follows. Section 4.2 briefly describes two detection statistics, F-statistic from SSK-DLM and the newly developed ST-LRT. In section 4.3, a method to implement ST-LRT with the parametric cepstrum is introduced. To be more specific, a method of noise model fitting and an iterative algorithm combining the estimation of signal parameters and noise parameters are given. Section 4.4 discusses an approach to compare ST-LRT model with SSK-DLM using a spatially decomposed AIC. The application of the developed techniques to a real FMRI dataset is discussed in section 4.5. Finally, some conclusions are drawn in section 7.7.

4.1.3 Acronyms and Notations

For convenience, we partly repeat acronyms and notations from Chapter 3 and add some new ones which are frequently used in this chapter. AIC means Akaike's information criterion; BOLD is blood oxygenation level dependent; DFT means discrete Fourier transform; DLM is dynamic linear model; FFT means fast Fourier transform; LRT is likelihood ratio test; LSE means least square estimate; MLE is maximum likelihood estimate; OLS means ordinary least square; PSD is power spectral density; ROI means region of interest; SSK is spatial smoothing by Gaussian kernel; ST-LRT means spatio-temporal likelihood ratio test; STWK is saptio-temporal whitening kernel; SWK means spatial whitening kernel.

For integer-valued t and $v = (v_1, \ldots, v_d)$, we consider a signal $y_{t,v}$ observed on a rectangle $\{0, \ldots, T-1\} \times \{0, \ldots, M_1 - 1\} \times \ldots \times \{0, \ldots, M_d - 1\}$ of sample size $TM \triangleq TM_1 \ldots M_d$, where d represents spatial dimensions. β_v contains nuisance signal components containing baseline and temporally varying drift, and f_v represents an activation amplitude at voxel v. In the spatiotemporal frequency domain, an integer k is used to denote the index of temporal frequency and an integer-valued vector $l = (l_1, \ldots, l_d)$ is assigned to the index of spatial wave-number. $\tilde{y}_{k,l}$ denotes the DFT of $y_{t,v}$ whose definition is given by, for $\omega_k \triangleq \frac{2\pi k}{T}$ and $\lambda_l = (\lambda_{l_1}, \ldots, \lambda_{l_d}) \triangleq (\frac{2\pi l_1}{M_1}, \ldots, \frac{2\pi l_d}{M_d})$,

$$\tilde{y}_{k,l} \triangleq \sum_{t=0}^{T-1} \sum_{v=0}^{M-1} y_{t,v} \; e^{-j(\omega_k t + \lambda_l \cdot v)}, \tag{4.1}$$

where $\sum_{v=0}^{M-1} \triangleq \sum_{v_1=0}^{M_1-1} \cdots \sum_{v_d=0}^{M_d-1}$ for a short notation. Notice that the periodicity of $\tilde{y}_{k,l}$, e.g., $\tilde{y}_{k,l_1} = \tilde{y}_{k+T,l_1+M_1}$ for any integer k and l_1. Discrete spatiotemporal PSD is defined as $F_{k,l} \triangleq F(\omega_k, \lambda_l)$ from a continuous PSD, $F(\omega, \lambda)$. Under space-time separability, namely $F_{k,l} = F_k G_l$, F_k is a purely temporal PSD and G_l is a purely spatial PSD.

For activation detection, the null hypothesis H_0 means there is no activations, and the alternative hypothesis H_1 represents activations may exist in a ROI. Under H_j for $j = 0, 1$, $\widehat{F}_{j,k,l}$ denotes the estimate of $F_{k,l}$, allowing a STWK $g_{j,t,v}$. $\theta_{t,v}$ represents a cepstral coefficient at (t, v), θ denotes a vector containing lexicographically ordered $\theta_{t,v}$s, and $\widehat{\theta}_{j,t,v}$ is the estimate of $\theta_{t,v}$ under H_j for $j = 0, 1$. $\mathcal{F}_v$ represents the widely used F-statistic derived from SSK-DLM. The operator $**$ denotes a spatiotemporal linear convolution; $*$ is used for a temporal convolution; $*_s$ is for a spatial convolution. The spatiotemporal circular convolution is denoted by circled asterisk, $\circledast\circledast$. Superscripts $(\,\cdot\,)^*$, $(\,\cdot\,)^T$ and $(\,\cdot\,)^H$ denote a complex conjugate, a transpose, and Hermitian transpose, respectively.

4.2 Detection Statistics for Activation - Revisited

We consider a real-valued model of FMRI measurements which has the following linear and additive form:

$$y_{t,v} = X_t^T \beta_v + \xi_t^T f_v + w_{t,v}, \tag{4.2}$$

where $X_t \triangleq [1, t]^T$, $\beta_v \triangleq [m_v, b_v]^T$, $\xi_t \triangleq [\xi_{1,t}, \dots, \xi_{L,t}]^T$, and $f_v \triangleq [f_{1,v}, \dots, f_{L,v}]^T$ are defined. m_v denotes baseline, b_v is the coefficient of temporal linear drift, and $s_{t,v}$ means the BOLD response. $w_{t,v}$ is spatiotemporally stationary Gaussian noise with zero mean whose PSD is denoted as $F_{k,l}$ in the temporal frequency and spatial wave-number domains. L denotes the number of basis functions used to model the BOLD response. The detailed discussions of the signal model formulation and two detection statistics are presented in Chapter 2 and 3, respectively.

4.2.1 Classical Detection Statistic : SSK-DLM

The SSK-DLM used in statistical parametric mapping (SPM) has a different model from (4.2),

$$y_{t,v} = X_t^T \beta_v + \xi_t^T f_v^G + \eta_{t,v}, \tag{4.3}$$

where f_v^G denotes a Gaussian amplitude activation and $\eta_{t,v}$ is a zero mean stationary Gaussian field which obeys space-time separability. $\eta_{t,v}$ is assumed spatially independent but temporally correlated. In SPM, after SSK and temporal filtering, e.g., temporal whitening or coloring, are applied to FMRI measurements, a voxel-wise the t or F-statistic is built up from OLS [66]. For example, the F-statistic at v is defined as

$$\mathcal{F}_v \triangleq \frac{\mathcal{H}_v / df_h}{\mathcal{E}_v / df_e}, \tag{4.4}$$

where $\mathcal{H}_v$ is the hypothesis sum of squares or signal sum of squares and $\mathcal{E}_v$ is error sum of squares; df_h and df_e denote associated degrees of freedoms. Then, random field theory (RFT) is used to control an overall error rate in a ROI, e.g., family-wise error (FWE). Therefore, the standard approach, SSK-DLM is based on two main assumptions: spatial independence and space-time separability.

4.2.2 New Detection Statistic : ST-LRT

We developed a detection statistic for activation based on (4.2), not requiring two main assumptions made in SSK-DLM. We called it spatio-temporal likelihood ratio test (ST-LRT) which, when spatially decomposed, consists of two pieces; a signal piece and a noise piece,

$$LRT_v \triangleq LRT_v^N + LRT_v^S, \tag{4.5}$$

where two pieces are defined as

$$\begin{aligned} LRT_v^N &\triangleq T \left(\widehat{\theta}_{0,0,0} - \widehat{\theta}_{1,0,0} \right) + \sum_{t=0}^{T-1} \left(\varepsilon_{0,t,v}^2 - \varepsilon_{1,t,v}^2 \right), && (4.6) \\ LRT_v^S &\triangleq \mathbf{s}_{y,v}^H \mathbf{s}_{y,v}. && (4.7) \end{aligned}$$

For example, under H_1, $\widehat{\theta}_{1,0,0}$ denotes the estimate of cepstral coefficient evaluated at the origin, $(t, v) = (0, 0)$ and

$$\varepsilon_{1,t,v} \triangleq (g_{1,t,v} \circledast \circledast y_{1,t,v}), \tag{4.8}$$

where $y_{1,t,v}$ denotes $y_{t,v}$ adjusted for baseline and drift, and $g_{1,t,v}$ is a temporally causal and spatially non-causal STWK defined by DFT as follows.

$$g_{1,t,v} \overset{DFT}{\longleftrightarrow} \tilde{g}_{1,k,l}, \qquad |\tilde{g}_{1,k,l}|^2 = \frac{1}{\widehat{F}_{1,k,l}}, \tag{4.9}$$

where $\widehat{F}_{1,k,l}$ is the estimator of $F_{k,l}$.

The LRT_v^S is specified via $\mathbf{s}_{y,v} \overset{DFT}{\longleftrightarrow} \tilde{\mathbf{s}}_{y,l} \triangleq \mathbf{S}_{\xi\xi,l}^{-\frac{1}{2}} \tilde{\mathbf{s}}_{y\xi,l}$,

$$\tilde{\mathbf{s}}_{y\xi,l} \triangleq \frac{1}{T} \cdot \sum_{k=0}^{T-1} \frac{\tilde{y}_{1,k,l}^{*} \tilde{\xi}_{1,k,l}}{\widehat{F}_{1,k,l}}, \qquad \mathbf{S}_{\xi\xi,l} \triangleq \frac{1}{T} \cdot \sum_{k=0}^{T-1} \frac{\tilde{\xi}_{1,k,l} \cdot \tilde{\xi}_{1,k,l}^{H}}{\widehat{F}_{1,k,l}}, \tag{4.10}$$

where

$$\tilde{y}_{1,k,l} \triangleq \tilde{y}_{k,l} - \tilde{X}_k^T \widehat{\tilde{\beta}}_{10,l}, \qquad \tilde{\xi}_{1,k,l} \triangleq \tilde{\xi}_k - \mathbf{S}_{X\xi,l}^T \tilde{X}_k, \tag{4.11}$$

$$\widehat{\tilde{\beta}}_{10,l} \triangleq \left(\sum_{k=0}^{T-1} \frac{\tilde{X}_k^* \cdot \tilde{X}_k^T}{\widehat{F}_{1,k,l}} \right)^{-1} \left(\sum_{k=0}^{T-1} \frac{\tilde{y}_{k,l} \tilde{X}_k^*}{\widehat{F}_{1,k,l}} \right), \tag{4.12}$$

$$\mathbf{S}_{X\xi,l} \triangleq \left(\sum_{k=0}^{T-1} \frac{\tilde{X}_k^* \cdot \tilde{X}_k^T}{\widehat{F}_{1,k,l}} \right)^{-1} \left(\sum_{k=0}^{T-1} \frac{\tilde{X}_k^* \cdot \tilde{\xi}_k^T}{\widehat{F}_{1,k,l}} \right). \tag{4.13}$$

$\widehat{\theta}_{0,0,0}$, $\varepsilon_{0,t,v}$, and $\tilde{y}_{0,k,l}$ can be similarly defined under H_0. For details of the derivation of ST-LRT, see Chapter 3. In this chapter, since H_1 is more complicated than H_0, we discuss signal and noise modeling mainly with H_1.

4.3 Joint Signal and Noise Modeling

The parametric cepstrum is proposed to solve the spatio-temporal spectral estimation problem in FMRI. The parametric cepstrum approach allows a dramatic reduction of computation in the fitting process with FFTs and a very simple relation between the required STWK and estimated spatiotemporal cepstral coefficients. In addition, we can describe the space and time separability condition easily (linearly) through the cepstrum, allowing the development of a test procedure for it in Chapter 5. We extend the purely 2D spatial setup in [55] to a 3D spatiotemporal situation requiring temporal causality which is typical in FMRI and suggest a way to obtain 3D STWK, $g_{t,v}$ from the estimated cepstral coefficients. Since only noise modeling in 2D space is described in [55], we have also extended the approach to include estimation of signal parameters such as β_v (nuisance signal components) and f_v (activation amplitudes) in an iterative way. The suggested framework can be easily extended to higher dimensions, e.g., 4D, that is 1D for time and 3D for space.

4.3.1 Parametric Cepstrum

Non-Separable Field

In 3D (1D is for time and 2D are for space), the parametric cepstrum is obtained by truncating the cepstral coefficient array, $\theta_{tv_1v_2}$ which is just the array of Fourier coefficients in a Fourier series expansion for the logarithm of spectrum; for $-\pi \leq \omega, \lambda_1, \lambda_2 \leq \pi$,

$$\widetilde{F}(\omega, \lambda_1, \lambda_2) \triangleq \log F(\omega, \lambda_1, \lambda_2) = \sum_{t=-n}^{n} \sum_{v_1=-p}^{p} \sum_{v_2=-q}^{q} \theta_{tv_1v_2} e^{-j(\omega t + \lambda_1 v_1 + \lambda_2 v_2)}, \qquad (4.14)$$

where $F(\omega, \lambda_1, \lambda_2)$ is a continuous spatiotemporal PSD, n represents the temporal order, and (p, q) means the spatial orders of the cepstrum model. Note that the cepstrum has a symmetry, that is $\theta_{tv_1v_2} = \theta_{(-t)(-v_1)(-v_2)}$ for $\forall(t, v_1, v_2)$ because PSD is real-valued. For the non-separable elliptic field, there are $R_{all} \triangleq (2p+1)(2q+1)n + 2pq + p + q + 1$ cepstral coefficients to be estimated.

Space-Time Separable Field

An advantage of the cepstrum is that hypotheses such as space-time separability are easily described. The space-time separable field is defined by the following condition in 3D:

$$F(\omega, \lambda_1, \lambda_2) = F(\omega)G(\lambda_1, \lambda_2), \tag{4.15}$$

where $F(\omega)$ is a continuous temporal PSD and $G(\lambda_1, \lambda_2)$ is a continuous PSD associated with space. Equivalently, taking logarithms gives

$$\widetilde{F}(\omega, \lambda_1, \lambda_2) = \widetilde{F}(\omega) + \widetilde{G}(\lambda_1, \lambda_2). \tag{4.16}$$

In the cepstral domain, therefore, the space-time separability condition in (4.15) entails the following linear condition: for $\forall(t, v_1, v_2) \neq (0, 0, 0)$,

$$\theta_{tv_1v_2} = \theta_{t00}\delta_{0v_1v_2} + \theta_{0v_1v_2}\delta_{t00}, \tag{4.17}$$

that is, in the 3D cepstral domain, cepstral coefficients only on the central v_1v_2-plane located at $t = 0$ and along the t-axis located at $(v_1, v_2) = (0, 0)$ have non-zero values. Under the space and time separability, the cepstral model in (4.14) is reduced to

$$\begin{aligned} \widetilde{F}(\omega, \lambda_1, \lambda_2) = \sum_{t=-n}^{n} \sum_{v_1=-p}^{p} \sum_{v_2=-q}^{q} \theta_{tv_1v_2} e^{-j(\omega t + \lambda_1 v_1 + \lambda_2 v_2)} = \theta_{000} \qquad (4.18) \\ + 2\sum_{t=1}^{n} \theta_{t00} \cos(\omega t) + 2\sum_{v_1=1}^{p} \theta_{0v_10} \cos(\lambda_1 v_1) + 2\sum_{v_2=1}^{q} \theta_{00v_2} \cos(\lambda_2 v_2) \\ + 2\sum_{v_1=1}^{p}\sum_{v_2=1}^{q} \theta^{(+)}_{0v_1v_2} \cos(\lambda_1 v_1 + \lambda_2 v_2) + 2\sum_{v_1=1}^{p}\sum_{v_2=1}^{q} \theta^{(-)}_{0v_1v_2} \cos(\lambda_1 v_1 - \lambda_2 v_2), \end{aligned}$$

where sine terms in the complex exponentials of (4.14) are canceled out due to the symmetry of the cepstrum. There are $R_s \triangleq 2pq + p + q + n + 1$ cepstral coefficients to be estimated for the separable field. In noise model fitting, therefore, space-time separability allows a substantial amount of reductions in the number of parameters.

4.3.2 Space-Time Filters in the Cepstral Domain

Non-Separable Field

According to the definition in (4.14), the parametric cepstrum allows two computationally nice properties. One is that the convolution of two functions in the space and time domains is equivalent to the addition of their cepstral coefficients in the cepstral domain. The other is that causality in the temporal cepstral domain entails causality in the time domain. Thus, after estimating spatiotemporal cepstral coefficients and determining model orders (n, p, q), it can be easily shown that the following relation is equivalent to (4.9) based on these two properties,

$$g_{j,t,v} \overset{CT}{\longleftrightarrow} \left(-\widehat{\theta}_{j,t,v}\right)^{+}, \tag{4.19}$$

where CT denotes cepstral transform and $(\widehat{\theta}_{j,t,v})^{+}$ represents the temporally causal part of $\widehat{\theta}_{j,t,v}$ in the cepstral domain under H_j for $j = 0, 1$. The causality of $\widehat{\theta}_{j,t,v}$ is required due to the temporally causal characteristic of STWK, $g_{j,t,v}$.

Space-Time Separable Field

Under space-time separability, one easily finds that the following two relations are equivalent to (4.9):

$$K_{j,v} \overset{CT}{\longleftrightarrow} -\frac{1}{2}\widehat{\theta}_{j,0,v}, \qquad\qquad q_{j,t} \overset{CT}{\longleftrightarrow} \left(-\widehat{\theta}_{j,t,0}\right)^{+}, \tag{4.20}$$

where $K_{j,v}$ denotes a SWK and $q_{j,t}$ denotes a temporally causal whitening filter under H_j. The causality of $\widehat{\theta}_{j,t,0}$ is required due to that of $q_{j,t}$. Therefore, regardless of the space and time separability, we easily obtain a STWK by FFTs from the estimated cepstral coefficients.

4.3.3 Noise Model Fitting

For a 3D stationary random field $w_{tv_1v_2}$ with zero mean, the periodogram is defined as

$$I_{kl_1l_2} \triangleq I(\omega_k, \lambda_{l_1}, \lambda_{l_2}) = \frac{|\tilde{w}(\omega_k, \lambda_{l_1}, \lambda_{l_2})|^2}{T \cdot M_1 \cdot M_2}, \tag{4.21}$$

where $\tilde{w}(\omega_k, \lambda_{l_1}, \lambda_{l_2})$ is the DFT of $w_{tv_1v_2}$. M_1 and M_2 represent the number of voxels along v_1-axis and v_2-axis, respectively. According to [6], under certain regularity conditions involving joint cumulants, we have an asymptotic distribution as $\min(T, M_1, M_2) \to \infty$, which is given by

$$\frac{I_{kl_1l_2}}{F_{kl_1l_2}} \sim \frac{\chi_2^2}{2}, \tag{4.22}$$

for

$$\forall (k, l_1, l_2) \in \Omega_h, \qquad \Omega_h \triangleq \cup_{i=1}^{3} \Omega_i, \tag{4.23}$$

where Ω_h denotes the half of the whole index region, Ω_f which is a 3D rectangle. Periodograms at different ordinates are asymptotically independent and χ_2^2 denotes a chi-square distribution with two degrees of freedom. The sets of indices Ω_f and Ω_i, where $i = 1, 2, 3$ have the following ranges;

$$\Omega_f = \{|k| \leq k^m, |l_1| \leq l_1^m, |l_2| \leq l_2^m\}, \tag{4.24}$$

$$\Omega_1 = \{1 \leq k \leq k^m, -l_1^m \leq l_1 \leq l_1^m, -l_2^m \leq l_2 \leq l_2^m\}, \tag{4.25}$$

$$\Omega_2 = \{k = 0, 1 \leq l_1 \leq l_1^m, -l_2^m \leq l_2 \leq l_2^m\}, \tag{4.26}$$

$$\Omega_3 = \{k = 0, l_1 = 0, 1 \leq l_2 \leq l_2^m\}, \tag{4.27}$$

where the union of Ω_2 and Ω_3 represents the half of the central l_1l_2-plane located at $k = 0$, excluding the origin. Here, $k^m \triangleq [(T-1)/2]$, $l_1^m \triangleq [(M_1-1)/2]$, and

$l_2^m \triangleq [(M_2-1)/2]$. To simplify the discussion, T, M_1 and M_2 are assumed odd and taken odd in the sequel. When the proposed techniques in this thesis are applied to a real human dataset in section 4.5, to make them odd, observations at the beginning of scans and on the boundaries of images are dropped. Notice that periodicity of $I_{kl_1l_2}$ and $F_{kl_1l_2}$, that is $I_{kl_1l_2} = I_{(k+T)(l_1+M_1)(l_2+M_2)}$ and $F_{kl_1l_2} = F_{(k+T)(l_1+M_1)(l_2+M_2)}$ for any integer k, l_1, and l_2.

Linear Estimator for Cepstral Coefficients

The framework of classical linear regression with the logarithm of $I_{kl_1l_2}$ as response variable is used for the linear estimation of cepstral coefficients, $\theta_{tv_1v_2}$. In details, taking the logarithm of (4.22) yields, defining pseudo-residual $Y_{kl_1l_2}$,

$$Y_{kl_1l_2} \triangleq \log I_{kl_1l_2} - \psi(1) = \log F_{kl_1l_2} + \epsilon_{kl_1l_2}, \tag{4.28}$$

where $\psi(1) = -0.5772$ is called Euler-Mascheroni constant.[1] $\epsilon_{kl_1l_2}$s are asymptotically independent and obey an extreme value distribution (EVD) with zero mean and $\psi'(1) = 1.6449$ variance. By replacing the logarithm of PSD in (4.28) with the parametric cepstrum model in (4.14), a classical linear regression equation is obtained, whose form is as follows.

For $\forall(k, l_1, l_2) \in \Omega_h$,

$$Y_{kl_1l_2} = x_{kl_1l_2}^T\theta + \epsilon_{kl_1l_2}, \tag{4.29}$$

where $x_{kl_1l_2}$ is a vector representing lexicographically ordered cosine terms and θ means a vector containing associated $\theta_{tv_1v_2}$ terms in (4.14). Note that sine terms are canceled due to the symmetry of cepstrum. For separable fields, we obtain the

---

[1] Let $U \sim \chi_2^2/2$, then $\mathbf{E}(\log U) = \psi(1)$ and $Var(\log U) = \psi'(1)$.

same regression equation with $x_{kl_1l_2}$ and θ of much smaller sizes than those for non-separable fields.

Due to the non-Gaussianity of $\epsilon_{kl_1l_2}$, least square estimate (LSE) is not statistically efficient in this case. However, it is very attractive since it is unbiased and linear. We have a LSE whose form[2] is given by

$$\widehat{\theta}_{LSE} = \left(X^T X\right)^{-1} X^T Y, \tag{4.30}$$

where

$$X^T X \triangleq \sum_{\Omega_h} x_{kl_1l_2} x_{kl_1l_2}^T, \qquad X^T Y \triangleq \sum_{\Omega_h} x_{kl_1l_2} Y_{kl_1l_2}. \tag{4.31}$$

Moreover, since $X^T X$ is almost diagonal[3], a modification of (4.30) to reduce computations is possible. Following [55], a biased estimator, $\widehat{\theta}_b$ can be obtained by taking just one FFT of the pseudo-residual, $Y_{kl_1l_2}$ and it turns out that a simple adjustment scheme exists to get the exact LSE from $\widehat{\theta}_b$. To be specific, $\widehat{\theta}_{LSE}$ is obtained by

$$\widehat{\theta}_{LSE} = \widehat{\theta}_b + \text{adjustment}, \qquad \widehat{\theta}_b \triangleq \begin{bmatrix} 2 & 0 \\ 0 & \mathbf{I} \end{bmatrix} X^T Y \big/ \, TM, \tag{4.32}$$

where **I** is an identity matrix. The adjustment term is computed by simple regional summations of $\widehat{\theta}_b$ in the cepstral domain. To overcome the inefficiency of $\widehat{\theta}_{LSE}$, in the following section, we consider an iterative method for MLEs of cepstral coefficients, which is based on the asymptotic log-likelihood equation and scoring algorithm. There, $\widehat{\theta}_{LSE}$ initialize the algorithm.

Asymptotic ML Estimator for Cepstral Coefficients

The performance of the LSE can be improved by one-step estimation (OSE). According to [59, pp. 71-75], if a preliminary estimator $\widetilde{\theta}$ is a $\sqrt{TM}$-consistent,

---

[2] For separable fields, $X^T X$ is a $R_s \times R_s$ matrix, which typically gives a huge matrix size. For non-separable fields, the size is much larger and $R_{all} \times R_{all}$.

[3] It is because that the design matrix X consists of only cosine terms.

then the scoring algorithm yields an estimator $\widehat{\theta}$ as efficient as MLE in the sense of asymptotics by the following one-step computation:

$$\widehat{\theta} = \widetilde{\theta} + \mathcal{I}^{-1}(\widetilde{\theta}) \cdot \nabla \mathscr{L}(\widetilde{\theta}), \tag{4.33}$$

where $\mathcal{I}$ means the Fisher information matrix (FIM) and $\nabla\mathscr{L}$ is the gradient of $\mathscr{L}$, Whittle's asymptotic log-likelihood equation [62]. For $\lambda = (\lambda_1, \ldots, \lambda_d)$, $\mathscr{L}$ is given by

$$\mathscr{L} \triangleq \iint \log F(\omega, \lambda) + \frac{I(\omega, \lambda)}{F(\omega, \lambda)}.$$

In 2D space, [55] shows that $\mathcal{I}$ is an identity matrix with a negative sign and $\mathscr{L}$ is convex with respect to θ, which are still true and can be easily shown in our 3D spatiotemporal case. Thus, (4.33) boils down to a very simple form involving no matrix inversion,

$$\widehat{\theta} = \widetilde{\theta} - h(\widetilde{\theta}), \tag{4.34}$$

where

$$h(\widetilde{\theta}) = \left[h_{tv_1v_2}(\widetilde{\theta}) \right], \qquad \psi_{tv_1v_2} \triangleq \cos(\omega t + \lambda_1 v_1 + \lambda_2 v_2), \tag{4.35}$$

$$\begin{aligned} h_{tv_1v_2}(\widetilde{\theta}) &\triangleq \iiint \frac{I(\omega, \lambda_1, \lambda_2)}{F(\omega, \lambda_1, \lambda_2 | \widetilde{\theta})} \cdot \psi_{tv_1v_2} \, \frac{\mathrm{d}\omega}{2\pi} \frac{\mathrm{d}\lambda_1}{2\pi} \frac{\mathrm{d}\lambda_2}{2\pi} \\ &\simeq \sum_{\Omega_f} \frac{I_{kl_1l_2}}{F_{kl_1l_2}(\widetilde{\theta})} \cdot \frac{\psi_{tv_1v_2}(k, l_1, l_2)}{TM_1M_2}, \end{aligned} \tag{4.36}$$

$$\psi_{tv_1v_2}(k, l_1, l_2) \triangleq \cos\left(\omega_k t + \lambda_{l_1} v_1 + \lambda_{l_2} v_2\right). \tag{4.37}$$

Note that $h_{tv_1v_2}$ can be obtained by taking the real part of the inverse FFT of $I_{kl_1l_2}/F_{kl_1l_2}$ for a given $\widetilde{\theta}$ in (4.36). Therefore, the asymptotic MLE is computed very quickly from $\widehat{\theta}_{LSE}$ by a small number of FFTs, resulting in a procedure for estimating $F_{kl_1l_2}$ that is performed linearly.

4.3.4 Noise plus Signal Modeling : Implementing ST-LRT

Our discussions were focused on the modeling of the noise $w_{t,v}$ in the previous sections. To implement ST-LRT in practice, we need to estimate signal parameters and noise parameters simultaneously under each hypothesis. Since the estimation of $\tilde{\beta}_l$ and $\tilde{f}_l$ (signal parameters) and the estimation of $F_{k,l}$, equivalently $\theta_{t,v}$ (noise parameters) are related to each other as shown in Chapter 3, an iterative algorithm to combine these two estimation procedures is necessary. This iterative algorithm allows us a practically efficient detection statistic.

From the DFT of the FMRI measurement model in (4.2),

$$\tilde{y}_{k,l} = \tilde{X}_k^T \tilde{\beta}_l + \tilde{\xi}_k^T \tilde{f}_l + \tilde{w}_{k,l}, \tag{4.38}$$

defining $\tilde{y}_l = [\tilde{y}_{0,l}, \ldots, \tilde{y}_{T-1,l}]^T$ and $\tilde{w}_l = [\tilde{w}_{0,l}, \ldots, \tilde{w}_{T-1,l}]^T$ allows a regression equation, $\tilde{y}_l = \tilde{\mathbf{Z}}\tilde{\alpha}_l + \tilde{w}_l$, where $\tilde{\alpha}_l^T \triangleq [\tilde{\beta}_{1,l}^T \ \tilde{f}_l^T]$ and matrix $\tilde{\mathbf{Z}}$ is defined as

$$\tilde{\mathbf{Z}}^T \triangleq \begin{bmatrix} \tilde{X}_0 & \tilde{X}_1 & \cdots & \tilde{X}_{T-1} \\ \tilde{\xi}_0 & \tilde{\xi}_1 & \cdots & \tilde{\xi}_{T-1} \end{bmatrix}. \tag{4.39}$$

Thus, the LSE of $\tilde{\alpha}_l$ is given by $\widehat{\tilde{\alpha}}_{LSE,l} = (\tilde{\mathbf{Z}}^H\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^H\tilde{y}_l$. We now describe a cyclic ascent algorithm for maximizing the likelihood and so calculating the ST-LRT. Maximization must be done under H_0 and H_1. We just describe the more complicated H_1 case. The cyclic ascent algorithm consists of two steps : a noise (parameter) step and a signal (parameter) step. We initialize with $\widehat{\tilde{\alpha}}_{LSE,l}$. Then, the algorithm is described as follows.

- **Noise-Step**: given $\tilde{\alpha}_l$, get θ.

  Given $\widehat{\tilde{\alpha}}_l$, compute the residuals, $\tilde{e}_{1,k,l} = \tilde{y}_{k,l} - \tilde{\mathbf{Z}}\widehat{\tilde{\alpha}}_l$ and use the method of section 4.3.3 to estimate θ from $\tilde{e}_{1,k,l}$.

- **Signal-Step**: given θ, get $\tilde{\alpha}_l$.

  Given $\widehat{\theta}$, calculate $\widehat{F}_{1,k,l}$ and so compute the quantities $\widehat{\tilde{\beta}}_{1,l}$ and $\widehat{\tilde{f}}_l$ from (3.9) and (3.11)$-$(3.14).

- Iterate the Noise-Step and Signal-Step until the algorithm converges. Then, we can compute the ST-LRT from section 4.2.2.

4.4 Joint Model Selection and Model Comparison

It is very important to emphasize here something which seems to be poorly understood in the FMRI literature. Namely, that competing models can not be compared by looking at activation maps; rather a properly constructed model comparison criterion is required that is typically a sum of a term measuring model fit and a term measuring model complexity. Note that the model fit term measures discrepancy between fitted model and data, whereas the overall model comparison criterion measures discrepancy between fitted model and underlying noise free unknown truth [57, 32]

The comparison of statistical models with a finite number of parameters includes model order selection as a special case. Using Akaike information criterion (AIC), we can not only compare models with different structures such as ST-LRT model and SSK-DLM but also models with the same structure to determine a proper order for a model. First, we consider a model selection for ST-LRT. Then, the model comparison of ST-LRT model and SSK-DLM will be discussed.

4.4.1 AIC Map for ST-LRT Model

General Development of AIC

To select a proper model for ST-LRT, we need to choose L, the number of basis functions for the modeling of the BOLD response in (4.2) and (n, p, q), the orders of parametric cepstrum for the modeling of PSD in (4.14). Thus, we perform a 4D search for (L, n, p, q). The AIC is defined as

$$AIC \triangleq -2 \cdot max \left(l(\tilde{\beta}, \tilde{f}, \theta; \tilde{y}) \right) + 2 \cdot (R + L), \tag{4.40}$$

where $l(\tilde{\beta}, \tilde{f}, \theta; \tilde{y})$ denotes a log-likelihood function for a given dataset $\tilde{y}$ and R is the number of cepstral coefficients [32]. To be specific, $R = R_{all}$ for non-separable fields and $R = R_s$ for separable fields. Thus, under H_1, we need to minimize

$$AIC = \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} \log \widehat{F}_{1,k,l} \; + \; \frac{|\tilde{e}_{1,k,l}|^2}{TM \cdot \widehat{F}_{1,k,l}} \; + 2 \cdot (R + L), \tag{4.41}$$

where residuals are defined as $\tilde{e}_{1,k,l} \triangleq \tilde{y}_{k,l} - \tilde{X}_k^T \widehat{\tilde{\beta}}_{1,l} - \tilde{\xi}_k^T \widehat{\tilde{f}}_l$ and all estimates are MLEs. Instead of AIC, an alternative method can be used, e.g., Bayesian information criterion (BIC) or minimum description length (MDL), which typically gives a smaller model than does AIC. Under H_0, a different set of noise model orders, (n, p, q) can be similarly determined by AIC.

Spatial Decomposition of AIC

As in the procedure to obtain ST-LRT in Chapter 3, we provide a spatial decomposition of AIC. To be specific, with Parseval's relation, the spatial decomposition of (4.41) gives an AIC map for the ST-LRT model which is defined as

$$AIC_{LRT,v} \; \triangleq \; T \cdot \widehat{\theta}_{1,0,0} + \sum_{t=0}^{T-1} (g_{1,t,v} \circledast \circledast e_{1,t,v})^2 + \frac{2(R+L)}{M}, \tag{4.42}$$

where $e_{1,t,v}$ is the inverse DFT of $\tilde{e}_{1,k,l}$ and $AIC = \sum_v AIC_{LRT,v}$. For a voxel v, therefore, the AIC map shows a contribution of that voxel to AIC for a whole ROI, indicating how close a fitted model is to the underlying but unknown truth. If values in an AIC map from one model are substantially lower than those from another model, then one is on average much closer to the truth than the other. Thus, by measuring relative distances to underlying noise free truth in the sense of average, ST-LRT model and SSK-DLM can be compared.

4.4.2 AIC Map for SSK-DLM

After SSK and temporal filtering, e.g., temporal whitening or coloring, are performed in (4.3), one arrives at

$$\bar{y}_{t,v} = \bar{X}_t^T \bar{\beta}_v + \bar{\xi}_t^T \bar{f}_v + \bar{\eta}_{t,v}, \tag{4.43}$$

where, e.g., $\bar{y}_{t,v} \triangleq \phi_t * (K_v^G *_s y_{t,v})$, K_v^G denotes a Gaussian kernel, and ϕ_t is a temporal filter. Then, for each voxel time course, the F-statistic is built up from (4.43) using OLS. Thus, one finds parameters which give the minimum value of the following objective function:

$$\begin{aligned} J &= \sum_{v=0}^{M-1} \sum_{t=0}^{T-1} \left(\bar{y}_{t,v} - \bar{X}_t^T \bar{\beta}_v - \bar{\xi}_t^T \bar{f}_v\right)^2, \\ &= \frac{1}{TM} \cdot \sum_{l=0}^{M-1} \sum_{k=0}^{T-1} \left|\tilde{y}_{k,l} - \tilde{X}_k^T \tilde{\beta}_l - \tilde{\xi}_k^T \tilde{f}_l\right|^2 |\tilde{\phi}_k|^2 |\tilde{K}_l^G|^2, \end{aligned} \tag{4.44}$$

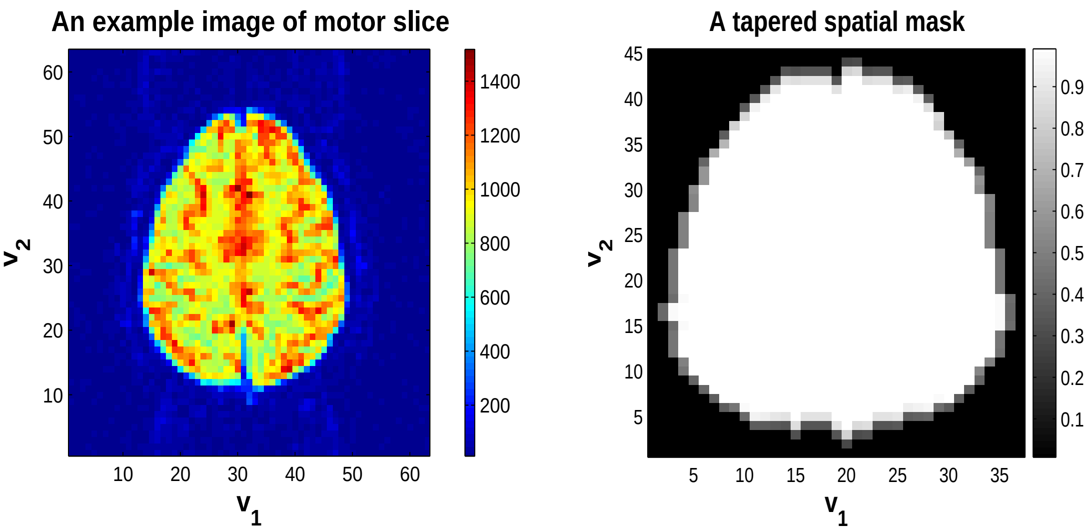
where the last equality is from Parseval's relation with zero padding. $\tilde{\phi}_k$ is the temporal DFT of ϕ_t and $\tilde{K}_l^G$ is the spatial DFT of K_v^G. If we compare (4.44) with (4.41), the simplest interpretation of $\tilde{\phi}_k$ and $\tilde{K}_l^G$ is that they are whitening operators (for a similar interpretation on temporal filtering, refer to [57]). Therefore, $|\tilde{\phi}_k|^2$ and $|\tilde{K}_l^G|^2$ take the place of $1/\widehat{F}_{1,k,l}$ in (4.41). If ϕ_t is a temporal whitening filter, this allows an AIC map for SSK-DLM,

$$AIC_{DLM,v} \triangleq \sum_{k=0}^{T-1} log\widehat{\check{F}}_k - 2T\theta_0^G + \sum_{t=0}^{T-1} \bar{e}_{t,v}^2 + \frac{2(n_p + L)}{M}, \tag{4.45}$$

where θ_0^G is the value of cepstral coefficient of K_v^G at the origin and $\bar{e}_{t,v} \triangleq \bar{y}_{t,v} - \bar{X}_t^T \widehat{\bar{\beta}}_v - \bar{\xi}_t^T \widehat{\bar{f}}_v$. n_p is the number of parameters for noise modeling. For AR(1) fitting of $\check{\eta}_{t,v} (\triangleq K_v^G *_s \eta_{t,v})$, we have $n_p = 2$ and $\check{F}_k = \frac{\check{\sigma}^2}{|1 - \check{\varphi} e^{-j\omega_k}|^2}$, where $\check{\sigma}^2 = var(\check{\eta}_{t,v})$ and $\check{\varphi}$ is an AR(1) coefficient. Now, we can compare the AIC map from ST-LRT model in (4.42) and one from SSK-DLM in (4.45) for a real FMRI dataset.

4.5 Application to a Human Dataset

We apply the newly developed techniques in the previous chapter and this chapter to a real dataset from the AFNI homepage (http://afni.nimh.nih.gov/afni/). The dataset is collected while a human subject is performing right-hand sequential finger-thumb opposition in the presence of a given motor stimulus signal. The on-off pattern



(a)An example image of motor slice (b)A tapered spatial mask

Figure 4.1: Real FMRI measurements from the AFNI homepage : (a) A 2D spatial plot of the collected dataset related to motor stimulus picked at an arbitrarily selected time point, (b) A brain-shaped mask with the tapered boundary by 2D spatial Tukey-Hanning window.

of the stimulus is shown on the bottom of Fig. 4.3 (black solid line). The experiment is performed under a 3T MRI scanner and TR is set to 2 seconds. At a given time t, for simplicity, a 2D axial slice is analyzed instead of a 3D volume. T is assigned to 100 and each axial slice consists of 64 × 64 voxels. Each voxel has dimensions of $3.125 \times 3.125 \times 5$ (mm^3). From the original 21 slices, a slice expected to contain motor reactions to the given motor stimulus is selected for analysis. Data taken at the beginning of the experiment are dropped, thus T is 99 and a spatial mask is applied to remove signals outside the subject's brain, thus M is reduced but odd. An example image of the motor slice picked at an arbitrarily selected time point is shown on Fig.4.1(a).

To reduce a bias from edge effects which are not negligible in 3D spatiotemporal spectral estimation [13, 27], we perform a spatiotemporal tapering using a 3D Tukey-Hanning window. Using a 2D Tukey-Hanning window, a tapered spatial mask is

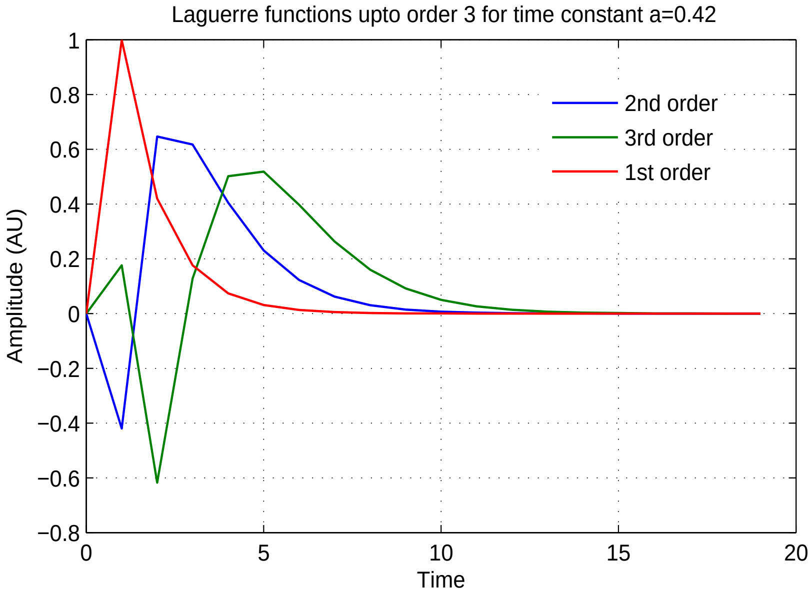


Figure 4.2: Laguerre functions up to order 3 for time constant $a = 0.42$.

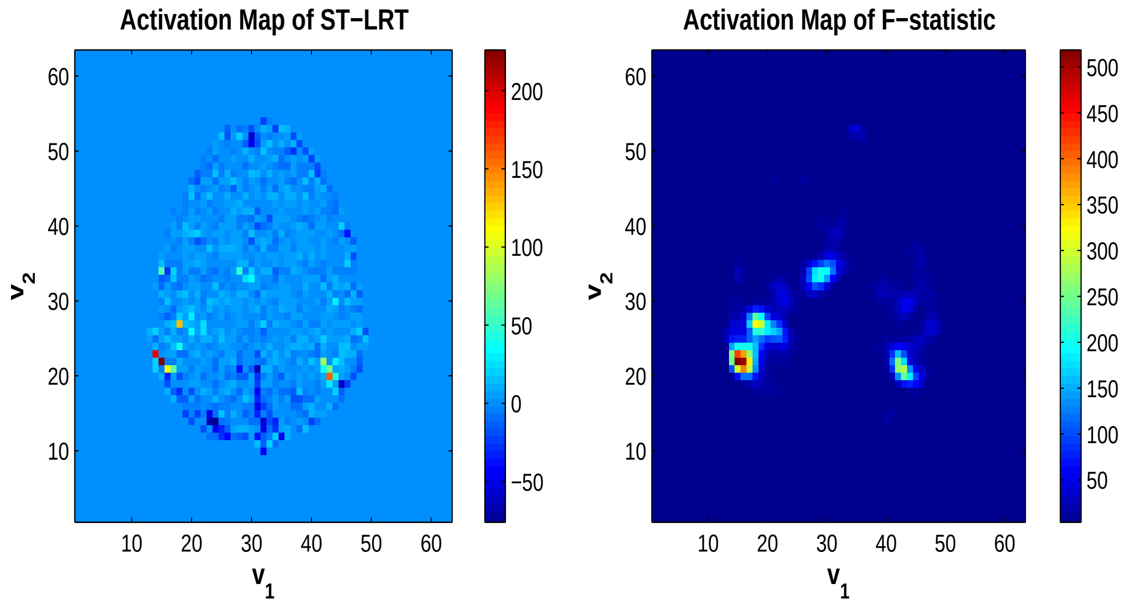
Figure 4.3: Fitted time courses at the voxel where LRT_v is maximized with different BOLD response modelings. Red solid line represents the observed FMRI time course. Blue solid line means the fitted time course with Laguerre modeling and cyan dashed line is from the parametric approach. Black solid line is the associated temporal stimulus where TR= 2 seconds.

shown in Fig.4.1(b). In addition to bias reduction in multi-dimensional spectral estimation, tapering generally provides two more advantages, one of which is to allow a better estimation of peaky spectrum. The other is to alleviate any non-stationary effect which can exist on the boundary of subject's brain [27].

For modeling of the BOLD response, a set of Laguerre functions is used [56]. Following our discussions about model selection, model orders are determined by AIC. Under H_1, the selected orders are $(L, n, p, q) = (3, 10, 7, 8)$, which means that three Laguerre functions whose orders are up to 3 are used to construct a subspace representing the BOLD responses. These three Laguerre functions with time constant $a = 0.42$ are shown on Fig.4.2. It turns out that BIC gives a smaller model than does AIC, but the model selected by BIC shows the similar results to those by AIC. The time constant $a = 0.42$ for the set of Laguerre functions is selected by the investigation of goodness of fit at the voxel where LRT_v has the maximum value. On Fig.4.3, at that voxel, fitted time courses using Laguerre modeling and the parametric modeling of the BOLD response are given. According to Fig.4.3, Laguerre modeling provides the better descriptions of undershoot and overshoot at the beginning and end of one period of the given stimulus. It is empirically known that typical BOLD responses have such undershoots and overshoots.

4.5.1 Activation Detection

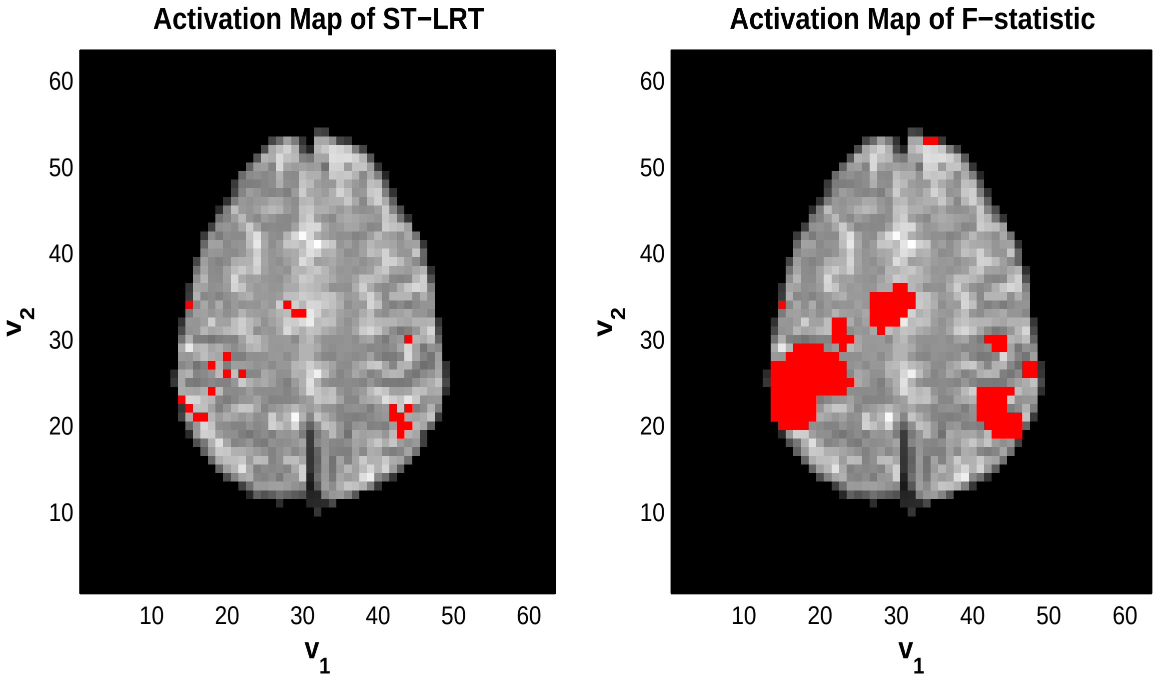
Before thresholding, an activation map associated with the motor responses generated from ST-LRT is given on Fig.4.4(a). It shows that the subject's right-hand movements make strong activation spots and weak spots around those strong spots in the primary motor cortex of the left hemisphere. The supplementary motor cortex in the center region of the brain has weak activations as well. There are some


Activation Map of ST−LRT

Activation Map of F−statistic

(a) Activation map by ST-LRT (b) Activation map by SSK-DLM

Figure 4.4: Unthresholded activation maps : (a) ST-LRT and (b) F-statistic, $\mathcal{F}_v$ derived from SSK-DLM.


Activation Map of ST−LRT

Activation Map of F−statistic

(a) Thresholded activation map by ST-LRT (b) Thresholded activation map by SSK-DLM

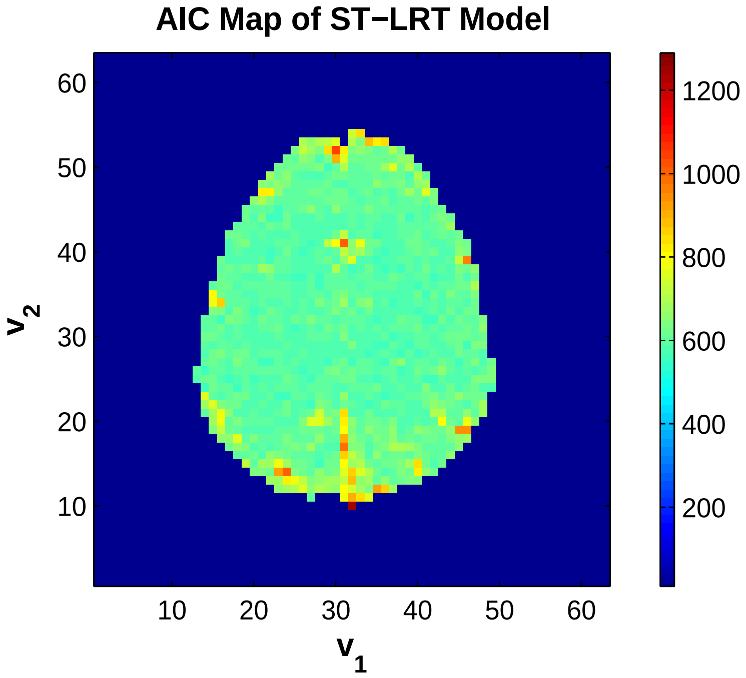
Figure 4.5: Thresholded activation map : (a) ST-LRT and (b) F-statistic, $\mathcal{F}_v$ derived from SSK-DLM. Red spots indicate activated voxels.

unexpected bright spots in the right hemisphere of the subject's brain. By thresholding ST-LRTs on Fig.4.4(a) with a threshold determined from (3.29) for significance level $\alpha = 0.05$, we obtain a thresholded activation map from ST-LRT shown on Fig.4.5(a). Before thresholding, an activation map from SSK-DLM, a spatial plot of F-statistics is shown on Fig.4.4(b). To create this unthresholded activation map, a Gaussian kernel whose full-width-half-maximum (FWHM) is set as 2.5 times of voxel size along each axis, as recommended in SPM, is used for SSK. A temporal AR(1) model with φ and σ^2 is fitted to spatially smoothed voxel time courses inside the subject's brain. To describe the BOLD response, since the Laguerre modeling was recently proposed by [56] and is not being used in softwares for FMRI data analysis like SPM, the parametric approach is applied to the data. Then, for each time series, temporal whitening is performed and the F-statistic, $\mathcal{F}_v$ is computed by (4.4). By thresholding these F-statistics on Fig.4.4(b), we obtain a thresholded activation map from SSK-DLM shown on Fig.4.5(b).

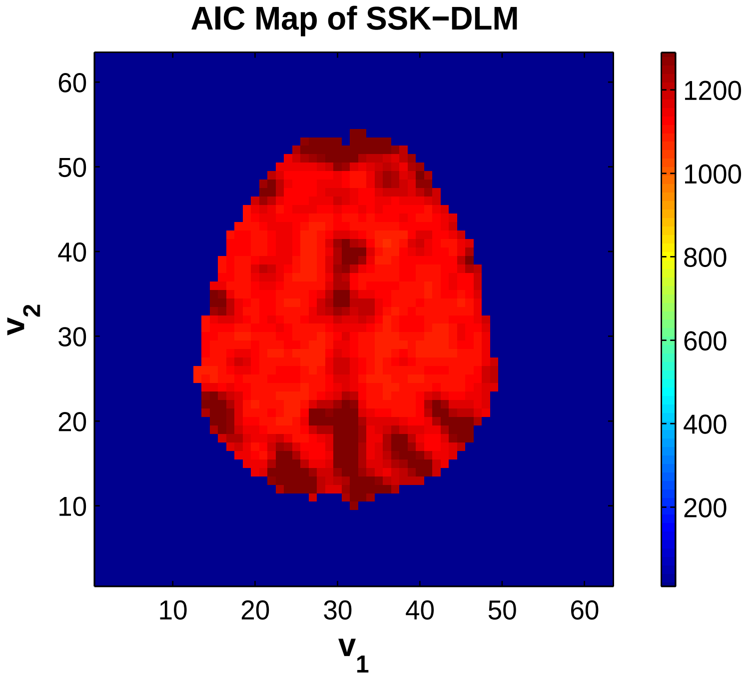
By comparing Fig.4.4(a) and Fig.4.4(b), we recognize that the activation map of ST-LRT shows shaper and more well-defined activated regions than the spatial plot of $\mathcal{F}_v$ derived from SSK-DLM in the whole brain. Besides that, in the primary motor cortex of the left hemisphere, the activation map by SSK-DLM on Fig.4.4(b) shows blurred activation regions and suggests wider activation spots than that by ST-LRT on Fig.4.4(a), indicating that SSK could artificially shift true activations as empirically reported in [25].

4.5.2 Model Comparison

AIC maps from ST-LRT model and SSK-DLM are compared on Fig.4.6(a) and Fig.4.6(b). For a proper comparison, the same scaling for values of the AIC maps



(a) AIC map of ST-LRT model



(b) AIC map of SSK-DLM

Figure 4.6: AIC maps : (a) ST-LRT model and (b) SSK-DLM.

are used for Fig.4.6(a) and Fig.4.6(b). According to two AIC maps, AIC values from ST-LRT model are substantially lower than those from SSK-DLM, thus indicating that ST-LRT model is on average much closer to the underlying unknown truth than SSK-DLM. By comparing two AIC maps, we made a comparison of two competing models, ST-LRT model and SSK-DLM based on a properly formulated criterion.

4.6 Conclusions

We have implemented the ST-LRT fully considering the spatiotemporal correlation of background noise developed in Chapter 3. With joint signal and noise modeling based on the parametric cepstrum, the implementation of ST-LRT was performed linearly and quickly. A new technique of model comparison based on a spatially decomposed AIC was introduced for the comparison of ST-LRT model and the existing SSK-DLM. The application of ST-LRT to a human dataset provided sharper and more well-defined activated regions than did F-statistic derived from SSK-DLM. The comparison of AIC maps indicated that ST-LRT model was on average much closer to the underlying noise free truth than SSK-DLM.

CHAPTER 5

A Procedure for Testing Space-Time Separability

One of the main tasks of FMRI data analysis is to determine which regions of the human brain are activated by presented temporal stimuli, equivalently, to build up an activation map from measurements in the experiment. In most studies in FMRI, space-time separability has been implicitly assumed and accepted without a proper justification. This is probably because the assumption of space and time separability allows a substantial amount of simplifications in modeling and analysis for the activation study. For the first time in FMRI, in this chapter, we propose a test procedure for testing the separability of space and time in FMRI time series by fully considering spatial and temporal correlations. In addition to that, we analyze the asymptotic power of the proposed testing procedure for space-time separability. The developed test statistic is tested by means of simulation and applied to a human dataset.

5.1 Introduction

Demands for FMRI are dramatically increasing due to its applicability to several fields and abilities in the investigation of the human brain. Creating an activation map to show localized activations in the human brain induced by pre-specified stimuli is one of the main purposes of data analysis in FMRI. Since the assumption of space-time separability allows researchers lots of simplifications for the activation study, from the very beginning of FMRI, it has been assumed and applied to most of developed techniques for data analysis without an appropriate justification. In fact, it has been assumed in other scientific studies, for example, to simplify procedures for solving Maxwell's equations in electromagnetism or Schrödinger's equation in quantum mechanics, the space-time separability has been assumed and yielded solutions satisfactorily explaining physical phenomena.

Conceptually, space-time separability in FMRI data analysis implies pure spatial operations and pure temporal operations can be separately and sequentially performed to achieve proper activation detection, where determining the order of spatial and temporal operations is a totally different problem. For example, in softwares such as statistical parametric mapping (SPM), widely used dynamic linear model (DLM) is only based on temporal aspects of voxel time courses and spatial smoothing by a Gaussian kernel (SSK) is purely built up on spatial assumptions of activation amplitudes and noise. For detailed reviews on SSK-DLM, refer to Chapter 2. For a given dataset, however, the validity of the space-time separability assumption is not known without an appropriate test procedure for it. Therefore, determining if a given dataset is space-time separable is an important issue not only for computational reduction but also for accurate modeling. To the best of author's knowledge, any testing procedure for space and time separability in FMRI has so far not been treated. As a matter of fact, a method proposed in this chapter is not only the first such testing procedure in FMRI but also a novel approach applicable to spatiotemporal modelings in other literatures. For some studies in other literatures, the reader is referred to [24, 36] which have non-parametric approaches, whereas our proposed test is made in the framework of the parametric cepstrum.

In Chapter 3, we developed a new statistic for activation detection by fully considering spatial and temporal correlations without space-time separability and showed the proposed detection statistic, spatio-temporal likelihood ratio test (ST-LRT) simplified under space-time separability as we expected. In Chapter 4, to implement ST-LRT, noise modeling for a general non-separable elliptic random field was required and the parametric cepstrum was proposed as an attractive solution to it. Thus, as a byproduct of discussions in two previous chapters, we are able to develop a method for testing space-time separability in this chapter.

5.2 Development of Test Procedure

Before we start our discussion to develop a test procedure for space-time separability, we need to emphasize here an important point to make the discussion straightforward. According to signal and noise modeling used in Chapter 3 and 4, we had two types of parameters, one for signal components and the other for noise components. The signal components contained nuisance parameters for drift and parameters of interest for activation amplitudes. The noise components had cepstral coefficients for the modeling of power spectral density (PSD). Since a test for space-time separability is purely made on noise parameters, cepstral coefficients, parameters associated with signal components are treated as nuisance parameters in this chapter. Therefore, assuming that activations may exist in a region of interest (ROI), the space and time separability test is built up and performed. Partitioning the whole parameter space into two parts, one for nuisance parameters and the other for parameters of interest, is particularly important when we analyze the asymptotic power of the proposed test for space-time separability in section 5.4.

5.2.1 Signal Model Formulation

We consider the following FMRI measurement model obtained at a time point t and a voxel location v:

$$y_{t,v} = X_t^T \beta_v + \xi_t^T f_v + w_{t,v}, \tag{5.1}$$

where $t = 0, \ldots, T-1$ and $v = 0, \ldots, M-1$. Equivalently, in the temporal frequency and spatial wave-number domains, discrete Fourier transform (DFT) yields

$$\tilde{y}_{k,l} = \tilde{X}_k^T \tilde{\beta}_l + \tilde{\xi}_l^T \tilde{f}_l + \tilde{w}_{k,l}, \tag{5.2}$$

where $k = 0, \ldots, T-1$ and $l = 0, \ldots, M-1$. β_v contains parameters for nuisance signal components such as drift and f_v denotes parameters for activation amplitudes. The noise $w_{t,v}$ is assumed spatiotemporally correlated and a stationary random field with zero mean. The PSD of $w_{t,v}$ is denoted as $F_{k,l}$. For details of this model formulation, refer to Chapter 1.

The **space-time separability** is now defined in the temporal frequency and spatial wave-number domains as follows. For $\forall(k,l)$,

$$F_{k,l} = F_k G_l, \tag{5.3}$$

where F_k is purely temporal PSD and G_l is purely spatial PSD. In other words, space and time separability, which can be equivalently defined in space and time domains using spatiotemporal auto-covariance function, means the spatiotemporal PSD can be decomposed into its purely temporal piece and purely spatial piece.

5.2.2 Test Statistic

Therefore, to test the condition in (5.3), we consider the following hypotheses:

$$\begin{aligned} H_{S0} &: F_{k,l} = F_k G_l \qquad for\ all\ \ (k,l), \\ H_{S1} &: F_{k,l} \neq F_k G_l \qquad for\ some\ \ (k,l), \end{aligned} \tag{5.4}$$

where H_{S0} says that $w_{t,v}$ is a separable random field and H_{S1} says that $w_{t,v}$ is a non-separable one. For the activation study in Chapter 3 and 4, the null hypothesis denoted as H_0 means there is no activation in a ROI and the alternative hypothesis denoted as H_1 means there are activations in the ROI as shown in (3.6). We differentiate two important testing problems in this thesis with a subscript S, meaning "separability". Note that the space-time separability test is built up under $H_0 \cup H_1$, thus activations in a ROI may exist, indicating f_v contains nuisance parameters.

By the same method used to obtain the likelihood ratio test (LRT) for the activation study in (3.8), we develop another LRT statistic, $\mathcal{L}_S (\triangleq 2\log \mathbf{\Lambda}_S)$ defined by

$$\begin{aligned} -2\log\mathbf{\Lambda}_S \triangleq & \sum_{k=0}^{T-1}\sum_{l=0}^{M-1} \log\widehat{F}_{k,l} - \log\widehat{F}_k\widehat{G}_l \\ & + \sum_{k=0}^{T-1}\sum_{l=0}^{M-1} \frac{\left|\tilde{e}_{S1,k,l}\right|^2}{TM\cdot\widehat{F}_{k,l}} - \frac{\left|\tilde{e}_{S0,k,l}\right|^2}{TM\cdot\widehat{F}_k\widehat{G}_l}, \end{aligned} \tag{5.5}$$

where residuals are defined as $\tilde{e}_{Sj,k,l} \triangleq \tilde{y}_{k,l} - \tilde{X}_k^T\widehat{\tilde{\beta}}_{Sj,l} - \tilde{\xi}_k^T\widehat{\tilde{f}}_{Sj,l}$ and all estimates are MLEs. The mathematical expressions of $\widehat{\tilde{\beta}}_{Sj,l}$ and $\widehat{\tilde{f}}_{Sj,l}$ are given by (3.9) and (3.11) under H_{Sj} for $j = 0, 1$. Under space-time separability, $\widehat{F}_k\widehat{G}_l$ is the estimate of $F_{k,l}$.

The examples of null and alternative distributions of $\mathcal{L}_S$ in (5.5) are shown on Fig.5.1, in which the red curve in the left indicates a distribution of $\mathcal{L}_S$ under H_{S0}, a central chi-square distribution and the blue curve in the right indicates a distribution of $\mathcal{L}_S$ under H_{S1}, a non-central chi-square distribution. In fact, the curves on Fig.5.1 are based on the discussions in section 5.3 (null distribution) and 5.4 (alternative distribution). Since the cepstrum allows a linear description for the condition of separability in (5.3), an idea of thresholding $\mathcal{L}_S$ in (5.5), i.e., the null distribution is discussed in the following section in the framework of the parametric cepstrum.

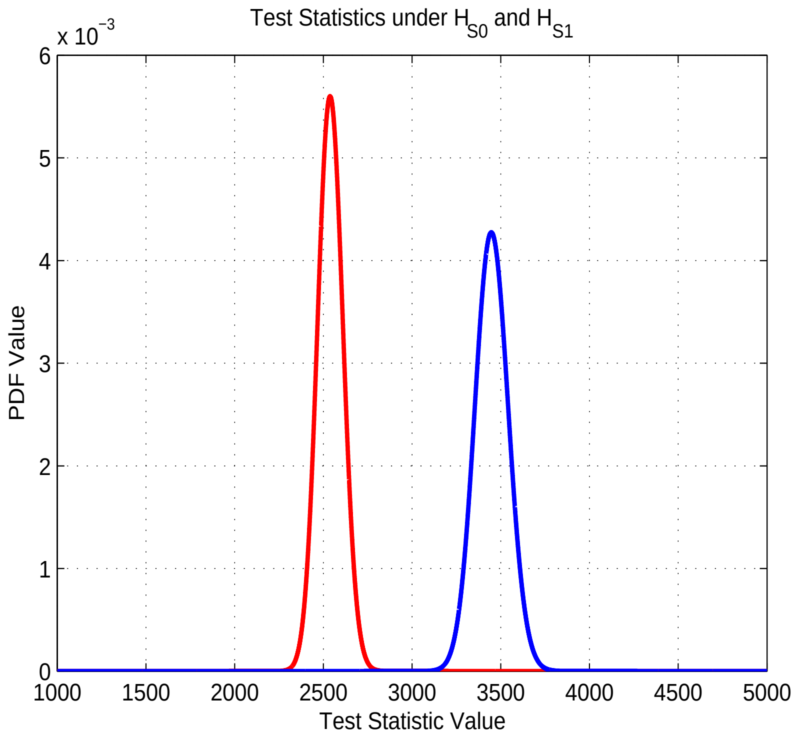


Figure 5.1: The examples of null (red curve in the left) and alternative (blue curve in the right) distributions of the test statistic $\mathcal{L}_S$ for space-time separability. From the discussion in section 5.3 and 5.4, the red curve indicates a central chi-square distribution and the blue curve indicates a non-central chi-square distribution. The separation of two curves is made by non-centrality parameter.

5.3 Reformulation in the Cepstrum Domain

To make a valid test procedure for space-time separability, we should be able to control false positive rate (FPR), the probability of type I error, within a pre-specified level. It is also known as false alarm rate in radar processing. To control FPR, the distribution of $\mathcal{L}_S$ under H_{S0}, called null distribution, is required, which enables us to compute a threshold for a given significance label α. Due to the nonlinearity of (5.4), at the first glance, it seems difficult to obtain an asymptotic null distribution of $\mathcal{L}_S$. The parametric cepstrum provides a useful framework to determine the asymptotic distribution of $\mathcal{L}_S$ under H_{S0}. In this section, the hypotheses in (5.4) are reformulated and a method to control FPR is discussed. To make the discussion complete, we briefly review noise modeling by the parametric cepstrum now.

5.3.1 Parametric Cepstrum

We consider a Fourier series expansion of the logarithm of PSD. Truncating the array of Fourier coefficients yields modeling by the parametric cepstrum [55]. For the index of temporal frequency k and the index of spatial wave-number l, a cepstral coefficient $\theta_{t,v}$ is defined by

$$\widetilde{F}_{k,l} \triangleq \log F_{k,l} = \sum_{t=-n}^{n} \sum_{v=-p}^{p} \theta_{t,v} e^{-j(\omega_k t + \lambda_l v)}, \tag{5.6}$$

where $\omega_k \triangleq \frac{2\pi k}{T}$ and $\lambda_l \triangleq \frac{2\pi l}{M}$. n and p denote a temporal order and a spatial order of the model, respectively. Notice that the cepstrum has a symmetry, namely $\theta_{t,v} = \theta_{(-t,-v)}$ for $\forall(t,v)$ due to real-valued PSD. For non-separable fields, thus there are $R_{all}(\triangleq 2np + n + p + 1)$ cepstral coefficients to be estimated.

For the current spatiotemporal setup of FMRI, noise modeling by the parametric cepstrum has several advantages over conventional AR-based methods. A linear description of the space and time separability in (5.3) is an example to show that modeling by the parametric cepstrum is very useful. For $(t,v) \neq (0,0)$, space-time separability can be expressed in the cepstrum domain as

$$\theta_{t,v} = \theta_{t,0}\delta_{0,v} + \theta_{0,v}\delta_{t,0}, \tag{5.7}$$

where $\theta_{t,0}$ denotes a cepstral coefficient along the temporal axis located at $v = 0$ and $\theta_{0,v}$ is a cepstral coefficient on the spatial plane located at $t = 0$. $\delta_{t,v}$ is a Kronecker delta function, which has 1 only when $(t,v) = (0,0)$ and 0 otherwise. Therefore, for separable fields, there are $R_s(\triangleq n+p+1)$ cepstral coefficients to be estimated. As an example of (5.7), Fig.5.2 shows cepstral coefficients of a two-dimensional (2D) random

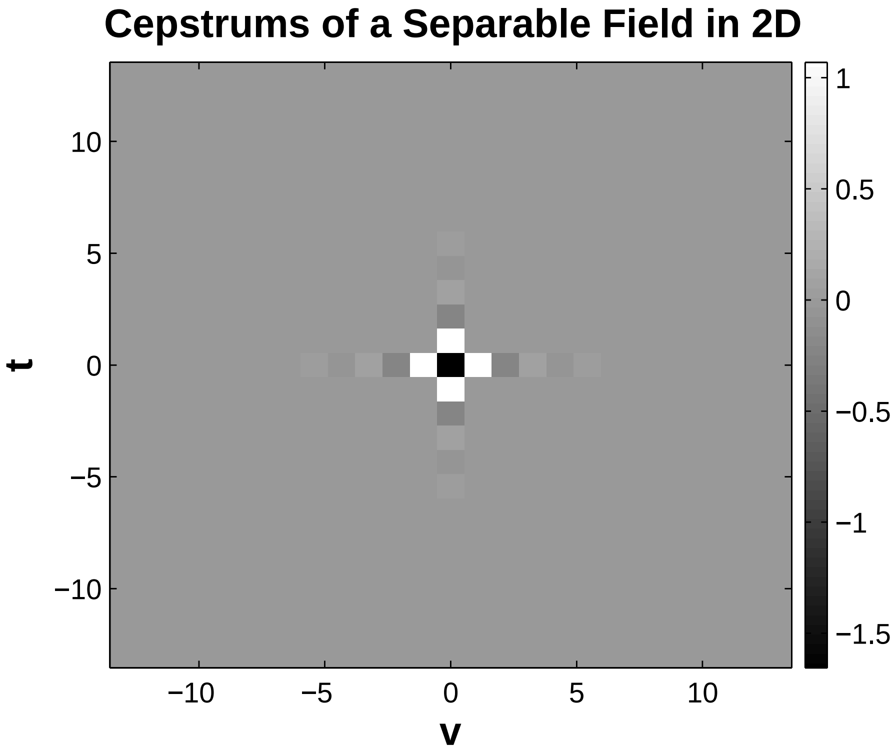


Figure 5.2: Cepstral coefficients of a 2D random field under space-time separability. The horizontal axis is associated with space and the vertical axis is associated with time. Notice that cepstral coefficients in the off-axis region have zero values.

field under space-time separability, in which we can clearly see that cepstrums only on two axes have non-zero values.

5.3.2 Controlling False Positive (False Alarm) Rate

Therefore, we have an equivalent hypothesis testing to (5.4) in the cepstral domain, for $(t, v) \in \Theta_{ns}$,

$$\begin{aligned} H_{S0} &: \theta_{t,v} = 0 \qquad for\ all\ \ (t,v), \\ H_{S1} &: \theta_{t,v} \neq 0 \qquad for\ some\ \ (t,v), \end{aligned} \tag{5.8}$$

where we call Θ_{ns} the non-separable region defined by

$$\begin{aligned}
\Theta_a &\triangleq \{|t| \leq n, |v| \leq p\} = \Theta_s \cup \Theta_{ns}, \quad \Theta_{ns} = \Theta_s^c, \qquad (5.9)\\
\Theta_s &\triangleq \{t = 0, |v| \leq p\} \cup \{|t| \leq n, v = 0\},
\end{aligned}$$

where the whole set of indices Θ_a is partitioned into two parts, the non-separable region Θ_{ns} and separable region Θ_s. From an asymptotic null distribution of LRT, which can be thought of as a special case of [58] when the non-centrality parameter is zero, the reformulation of the hypotheses in (5.8) allows a framework for thresholding $\mathcal{L}_S$. It can be easily checked that a threshold for a significance level α is given by

$$\gamma(\alpha) = \Psi_{R_{ns}}^{-1}(1 - \alpha), \qquad (5.10)$$

where $\Psi_{R_{ns}}(t)$ denotes the cumulative density function (CDF) of a chi-square distribution whose degrees of freedom is $R_{ns}(\triangleq R_{all} - R_s = 2np)$.

Details of signal and noise model fitting by the parametric cepstrum, and model order selection were discussed for three-dimensional (3D) FMRI measurements in Chapter 4. They can be easily extended to four-dimensional (4D) cases. Therefore, we do not provide the discussions on those topics here. To review those topics, the reader is referred to section 4.3 (signal and noise modeling) and section 4.4 (model selection and comparison).

5.4 Asymptotic Power Analysis

In this section, the asymptotic power of the proposed test for space-time separability is analyzed, requiring an asymptotic distribution of LRT under alternative hypothesis. Since the proposed test statistic involves non-identically distributed samples in the presence of nuisance parameters, e.g., cepstral coefficients in the separable region Θ_s or activation amplitudes, an asymptotic expansion of $\mathcal{L}_S$ under H_{S1} is a non-trivial task. While this kind of problem has long been discussed in the statistics literature, the conventional work deals only with independently and identically distributed samples and is not applicable to our situation, e.g., [28]. In a very recent work in time series, an asymptotic expansion of a class of test statistics including LRT is derived for a local alternative when serial correlation and nuisance parameters exist [58]. Here, the results in [58] are applied to derive the asymptotic distribution of $\mathcal{L}_S$ for non-separable random fields, i.e., under H_{S1}. As a matter of fact, [58] has some regularity conditions involving the differentiability of log-likelihood function and validity of asymptotic expansions of cumulants, which can be straightforwardly checked in our current setups for $\mathcal{L}_S$ but it is tedious.

5.4.1 Asymptotic Expansion of LRT under Local Alternatives

Suppose that η denotes a vector containing lexicographically ordered parameters of interest and μ is a vector containing ordered nuisance parameters. η_0 denotes η which is specified by a given null hypothesis. The standard approach to asymptotic power analysis is to consider asymptotic behavior for a sequence of local alternatives, defined as $\eta = \eta_0 + \varepsilon/\sqrt{n'}$, where n' is the number of samples and ε describes the deviation of η from η_0 as a function of n'.

According to [58, Theorem 1], a test statistic T belonging to a class which includes LRT has the following asymptotic expansion for a sequence of local alternatives: for $-\infty < t < \infty$,

$$\Pr\left(T < t\right) = \Phi_{d,\Delta}(t) + \frac{1}{\sqrt{n'}}\sum_{q=0}^{3} m_q \Phi_{d+2q,\Delta}(t) + o\left(\frac{1}{\sqrt{n'}}\right), \tag{5.11}$$

where $\Phi_{d,\Delta}(t)$ denotes the CDF of a non-central chi-square distribution with d degrees of freedom and Δ non-centrality parameter. m_qs can be computed from asymptotic expansions of the moments of the first and second order derivatives of a given log-likelihood function. Remarkably, the degrees of freedom d is the same as the length of the vector containing parameters of interest η and the non-centrality parameter Δ is given by

$$\Delta = \varepsilon^T \left(\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21}\right)\Big|_{\eta=\eta_0} \varepsilon, \tag{5.12}$$

where, following the partition of the parameter space into two pieces, η (parameter of interest) and μ (nuisance parameter), the Fisher information matrix (FIM) is partitioned into

$$\mathcal{I}(\eta,\mu) \triangleq \begin{bmatrix} \mathcal{I}_{11}(\eta) & \mathcal{I}_{12}(\eta,\mu) \\ \mathcal{I}_{21}(\mu,\eta) & \mathcal{I}_{22}(\mu) \end{bmatrix}. \tag{5.13}$$

Here, $\mathcal{I}_{11}(\eta)$ is associated with the parameter of interest and $\mathcal{I}_{22}(\mu)$ involves only nuisance parameter. If $\mathcal{I}_{12}(\eta_0,\mu)$ in (5.12) is a zero matrix, then the non-centrality parameter simplifies into

$$\Delta = \varepsilon^T \mathcal{I}_{11}(\eta_0)\varepsilon, \tag{5.14}$$

where note that the non-centrality parameter Δ is independent of nuisance parameter μ, so that asymptotic power does not depend on nuisance parameter, which is the case in the proposed test procedure for space-time separability we now discuss.

5.4.2 Asymptotic Power Function of the Space-Time Separability Test

From (5.6), the model by the parametric cepstrum allows the following expression for the logarithm of PSD:

$$\log F_{k,l} = x_{k,l}^T \theta^{ns} + z_{k,l}^T \theta^{s}, \tag{5.15}$$

where θ^{ns} denotes a column vector containing lexicographically ordered cepstral coefficients in the non-separable region Θ_{ns} and $x_{k,l}$ is a vector containing associated cosine terms. Note that sine terms in (5.6) are canceled due to the symmetry of cepstrums. θ^s and $z_{k,l}$ are similarly defined for cepstral coefficients in the separable region Θ_s. Therefore, θ^{ns} defines a vector η for parameters of interest and θ^s is contained in μ with other nuisance parameters.

Then, from the asymptotic complex-Gaussian distribution in (3.4), (5.15) yields a fully parameterized negative log-likelihood function whose form is given by

$$\begin{aligned} 2\ell(\tilde{\beta}, \tilde{f}, \theta^s, \theta^{ns}) \; &= \sum_{k=0}^{T-1}\sum_{l=0}^{M-1} x_{k,l}^T\theta^{ns} + z_{k,l}^T\theta^s \\ &\quad + \frac{exp(-x_{k,l}^T\theta^{ns} - z_{k,l}^T\theta^s)}{TM}\left|\tilde{y}_{k,l} - \tilde{X}_k^T\tilde{\beta}_l - \tilde{\xi}_k^T\tilde{f}_l\right|^2, \end{aligned} \tag{5.16}$$

where $\tilde{\beta}^T \triangleq [\tilde{\beta}_0, \ldots, \tilde{\beta}_{M-1}]$ and $\tilde{f}^T \triangleq [\tilde{f}_0, \ldots, \tilde{f}_{M-1}]$. Now we have $\eta \triangleq \theta^{ns}$ and $\mu^T \triangleq [\theta^{sT}, \tilde{\beta}^T, \tilde{f}^T]$. To simplify the discussion, we first derive the non-centrality parameter Δ for a simplified case without signal components, $\tilde{\beta}$ (baseline and linear drift) and $\tilde{f}$ (activation amplitude). Then, we reconsider the signal components in the general development. In fact, it turns out that results from the simplified case are useful in the general development. Using a threshold $\gamma(\alpha)$ determined to control FPR, the asymptotic power function can be computed from the non-central chi-square distribution, $\Phi_{d,\Delta}(t)$ given in (5.11).

Case I : development without signal components

To compute the non-centrality parameter Δ in (5.12), the FIM is required. Assuming that there are no drift and activation amplitudes, i.e., $\tilde{\beta} = 0$ and $\tilde{f} = 0$, it is easily checked that second order derivatives of $2\ell(\theta^s, \theta^{ns})$ with respect to θ_i^{ns}(the i-th element of θ^{ns}) and θ_j^s(the j-th element of θ^s) are given by

$$\frac{\partial^2 2\ell}{\partial\theta_i^{ns}\partial\theta_m^{ns}} = \sum_{k=0}^{T-1}\sum_{l=0}^{M-1}\frac{|\tilde{y}_{k,l}|^2}{TM}exp(-x_{k,l}^T\theta^{ns} - z_{k,l}^T\theta^s)\cdot x_{k,l,i}x_{k,l,m}, \tag{5.17}$$

$$\frac{\partial^2 2\ell}{\partial\theta_j^{s}\partial\theta_n^{s}} = \sum_{k=0}^{T-1}\sum_{l=0}^{M-1}\frac{|\tilde{y}_{k,l}|^2}{TM}exp(-x_{k,l}^T\theta^{ns} - z_{k,l}^T\theta^s)\cdot z_{k,l,j}z_{k,l,n}, \tag{5.18}$$

$$\frac{\partial^2 2\ell}{\partial\theta_i^{ns}\partial\theta_n^{s}} = \sum_{k=0}^{T-1}\sum_{l=0}^{M-1}\frac{|\tilde{y}_{k,l}|^2}{TM}exp(-x_{k,l}^T\theta^{ns} - z_{k,l}^T\theta^s)\cdot x_{k,l,i}z_{k,l,n}, \tag{5.19}$$

where $x_{k,l,i}$ denotes the i-th entry of $x_{k,l}$ and $z_{k,l,j}$ is the j-th entry of $z_{k,l}$. By CLT, we have $\mathbf{E}[|\tilde{y}_{k,l}|^2] = TM \cdot F_{k,l}$ for $\forall(k,l)$. By the cepstrum modeling, one arrives at

$$\mathbf{E}\left[\frac{\partial^2 2\ell}{\partial\theta_i^{ns}\partial\theta_m^{ns}}\right] = \sum_{k=0}^{T-1}\sum_{m=0}^{M-1} x_{k,l,i}x_{k,l,m} = 2TM \cdot \delta_{i-m}, \tag{5.20}$$

$$\mathbf{E}\left[\frac{\partial^2 2\ell}{\partial\theta_i^{ns}\partial\theta_n^{s}}\right] = \sum_{k=0}^{T-1}\sum_{m=0}^{M-1} x_{k,l,i}z_{k,l,n} = 0, \tag{5.21}$$

$$\begin{aligned}\mathbf{E}\left[\frac{\partial^2 2\ell}{\partial\theta_j^{s}\partial\theta_n^{s}}\right] &= \sum_{k=0}^{T-1}\sum_{m=0}^{M-1} z_{k,l,j}z_{k,l,n} \\ &= \begin{cases} TM & \text{if } j = n = 1 \\ 2TM \cdot \delta_{j-n} & \text{otherwise} \end{cases}, \end{aligned} \tag{5.22}$$

where δ_i denotes a Kronecker delta function. Since $x_{k,l}$ and $z_{k,l}$ involve only harmonic cosine functions, the last identities in (5.20)-(5.22) can be verified under the assumption that $0 < n \ll T$ and $0 < p \ll M$, which is usually the case in cepstrum modeling. For the completeness of this chapter, we provide the proofs of (5.20), (5.21), and (5.22) in Appendix 5.8. From (5.20)-(5.22), we have the following FIM:

$$\mathcal{I}(\theta^{ns}, \theta^{s}) \triangleq \begin{bmatrix} \mathbf{I}_{R_{ns}\times R_{ns}} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \mathbf{I}_{(R_s-1)\times(R_s-1)} \end{bmatrix}, \tag{5.23}$$

where the identity matrix whose size $R_{ns} \times R_{ns}$ in the upper block is associated with cepstral coefficients in Θ_{ns}, defining $\mathcal{I}_{11}(\theta^{ns})$. The entry in the center and the identity matrix in the lower block are associated with cepstral coefficients in Θ_s, giving $\mathcal{I}_{22}(\theta^s)$. From (5.12), due to $\theta^{ns} = 0$ under H_{S0}, we have the non-centrality parameter of

$$\Delta = TM \cdot (\theta^{ns})^T \mathbf{I}_{R_{ns}\times R_{ns}}(\theta^{ns}), \tag{5.24}$$

indicating that the non-centrality parameter involves parameters in Θ^{ns} in a quadratic way.

Case II : general development with signal components

We reconsider signal components. For convenience, $\tilde{\beta}$ and $\tilde{f}$ are decomposed into their real part and imaginary part, namely $\tilde{\beta} \triangleq \tilde{\beta}^R + j\tilde{\beta}^I$ and $\tilde{f} \triangleq \tilde{f}^R + j\tilde{f}^I$. Re-parameterization of the asymptotic negative log-likelihood function is made, so that we have $2\ell(\tilde{\beta}^R, \tilde{\beta}^I, \tilde{f}^R, \tilde{f}^I, \theta^s, \theta^{ns})$. Differentiating this function $2\ell(\,\cdot\,)$ twice with respect to each part of $\tilde{f}_l$ and θ_i^{ns}, and then taking the expectation of it yield the following equations:

$$\mathbf{E}\left[\frac{\partial^2 2\ell}{\partial \tilde{f}_l^R \partial \theta_i^{ns}}\right] = 0, \qquad \mathbf{E}\left[\frac{\partial^2 2\ell}{\partial \tilde{f}_l^I \partial \theta_i^{ns}}\right] = 0, \tag{5.25}$$

where $\tilde{f}_l^R$ is the l-th element of $\tilde{f}^R$ and $\tilde{f}_l^I$ is the l-th entry of $\tilde{f}^I$. Similarly, we obtain the following results for $\tilde{\beta}_l^R$ and $\tilde{\beta}_l^I$:

$$\mathbf{E}\left[\frac{\partial^2 2\ell}{\partial \tilde{\beta}_l^R \partial \theta_i^{ns}}\right] = 0, \qquad \mathbf{E}\left[\frac{\partial^2 2\ell}{\partial \tilde{\beta}_l^I \partial \theta_i^{ns}}\right] = 0. \tag{5.26}$$

Proofs of (5.25) and (5.26) are given in Appendix 5.9. For the general development with signal components, combining (5.23), (5.25), and (5.26) yields the following FIM:

$$\mathcal{I}(\theta^{ns}, \theta^s, \tilde{\beta}, \tilde{f}) \;\triangleq\; \begin{bmatrix} \mathbf{I}_{R_{ns}\times R_{ns}} & 0 & 0 \\ 0 & \mathcal{I}(\theta^s) & \mathcal{I}(\theta^s, \tilde{\beta}, \tilde{f}) \\ 0 & \mathcal{I}(\tilde{\beta}, \tilde{f}, \theta^s) & \mathcal{I}(\tilde{\beta}, \tilde{f}) \end{bmatrix}, \tag{5.27}$$

where $\mathbf{I}_{R_{ns}\times R_{ns}}$ defines $\mathcal{I}_{11}(\theta^{ns})$ and the remaining partition in the lower block specifies $\mathcal{I}_{22}(\theta^s, \tilde{\beta}, \tilde{f})$. Note that the other partitions are all zero matrices, yielding $\mathcal{I}_{12}(\theta^{ns}, \theta^s, \tilde{\beta}, \tilde{f}) = 0$. From (5.12), then, the non-centrality parameter is given by

$$\Delta \;=\; TM \cdot \sum_{i=1}^{R_{ns}} (\theta_i^{ns})^2, \tag{5.28}$$

which is, in fact, the same as (5.24), indicating that the non-centrality parameter Δ is independent of signal components of β_v (baseline and temporal linear drift) and

f_v (activation amplitude). It can be shown that Δ leads to

$$2\Delta = \left(\frac{1}{2\pi}\right)^2 \iint (\log F(\omega,\lambda) - \log F_S(\omega,\lambda))^2 \mathrm{d}\omega \mathrm{d}\lambda. \tag{5.29}$$

Thus, 2Δ is an Euclidean metric between the logarithms of a non-separable PSD, $F(\omega,\lambda)$ and a separable PSD, $F_S(\omega,\lambda)$ [35]. Since $\theta_{0,0}$ corresponds to the amplitude of PSD and other $\theta_{t,v}$'s are associated with the shape of PSD, the independence of Δ and $\theta_{0,0}$ can be recognized from (5.29) as well.

Therefore, for large T and M, we have the asymptotic distribution of $\mathcal{L}_S$ under H_{S1} as follows:

$$\mathcal{L}_S \sim \chi^2_{R_{ns},\Delta}, \tag{5.30}$$

where $\chi^2_{R_{ns},\Delta}$ denotes a non-central chi-square distribution with R_{ns} degrees of freedom and Δ non-centrality parameter. R_{ns} is the number of cepstral coefficients in the non-separable region Θ_{ns} and Δ is provided in (5.28). We can compute the power, probability that the proposed separability test detects non-separability when H_{S1} is true. By combining (5.10) and (5.30), for a significance level α, the power function is given by,

$$\mathrm{P}_{sep}\left(\theta^{ns}\right) \triangleq \Pr\left(\mathcal{L}_S > \gamma(\alpha) \;\middle|\; H_{S1}\right) = 1 - \Phi_{R_{ns},\Delta}\left(\Psi^{-1}_{R_{ns}}(1-\alpha)\right). \tag{5.31}$$

Assuming only one non-zero cepstral coefficient in Θ_{ns}, an example plot of $\mathrm{P}_{sep}\left(\theta^{ns}\right)$ is shown in Fig.5.3 for $\alpha = 0.05$, $R_{ns} = 2540$, $T = 99$, and $M = 1435$, the same setup for the AFNI dataset used in Chapter 4 and section 5.6.

5.4.3 Discussions

From the derived asymptotic power function of the proposed test procedure for space and time separability, some important remarks can be made for given cepstrum model orders, (n,p). Firstly, the asymptotic power of the test procedure is

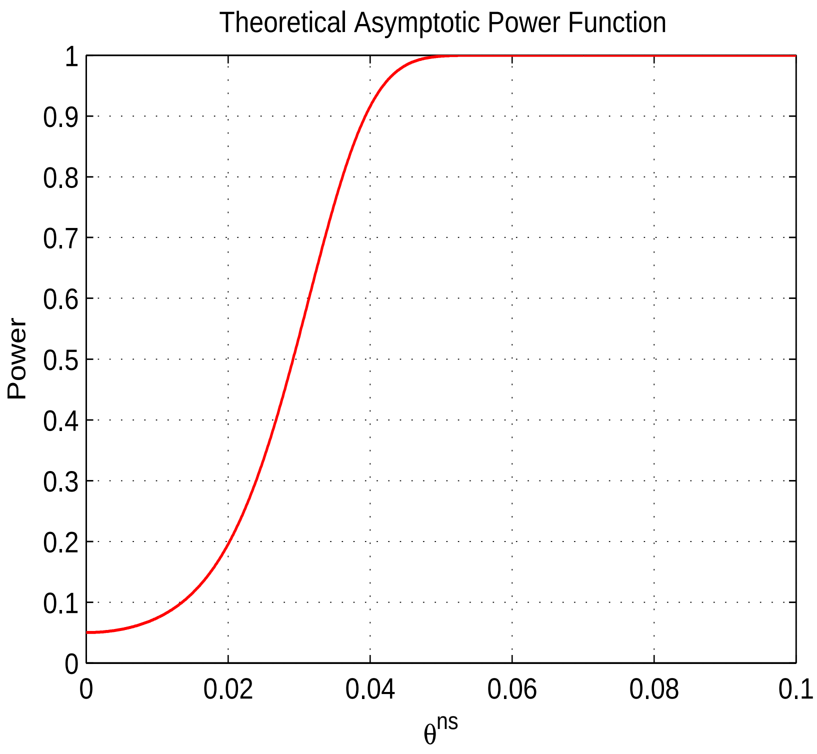


Figure 5.3: An example of the theoretical asymptotic power function $\mathrm{P}_{sep}(\theta^{ns})$ for $\alpha = 0.05$, $R_{ns} = 2540$, $T = 99$, and $M = 1435$, the same setup for the AFNI dataset except non-centrality parameter used in Chapter 4 and section 5.6. Only one non-zero cepstral coefficient in Θ_{ns} is assumed for this example plot.

not dependent on nuisance parameters, baseline, temporal linear drift, activation amplitudes, and cepstral coefficients in the separable region. Recall that the proposed test is derived from hypotheses involving cepstral coefficients in the non-separable region. Secondly, the asymptotic power function is independent of the locations of cepstral coefficients in the cepstral domains. It is affected by values of cepstrums in the non-separable region. Finally, cepstral coefficients in the non-separable region contribute to the non-centrality parameter Δ in a quadratic way.

5.5 A Simulation Study

We perform a simulation to test the proposed test statistic for space-time separability, $\mathcal{L}_S$. A truly space-time separable 3D random field with arbitrarily chosen spatial and temporal PSDs is generated by convolving a 3D Gaussian white noise with a temporal filter and a spatial kernel, sequentially. Using (5.1), a dataset consists of signal and noise is synthesized. To simplify the simulation, the simplest parametric modeling of the BOLD response in (1.7) is used, thus the number of basis functions for the BOLD response is one, i.e., $L = 1$. The number of time points, T is set to 99 and the number of voxels in a ROI, M is assigned to 63×63, which are the same setup as those of a human dataset used in the following section. Then, PSDs are estimated under H_{S0} and H_{S1} through the parametric cepstrum method with a spatiotemporal tapering to reduce edge effects and estimates of other nuisance parameters ($\tilde{\beta}_l$ and $\tilde{f}_l$) are also computed. Then, the test statistic $\mathcal{L}_S$ is calculated. The obtained $\mathcal{L}_S$ is 155 and a determined threshold, $\gamma(0.05)$ for a significance level $\alpha_S = 0.05$ is 173. These two quantities yield that p-value is $0.25 (\gg 0.05)$. Therefore, there is no evidence against space-time separability and we fail to reject the null hypothesis H_{S0}. This indicates that, for the synthesize dataset in this simulation, $\mathcal{L}_S$ works properly as desired.

5.6 Application to a Human Dataset

We apply the developed test procedure for space-time separability to the AFNI dataset used for the activation study in Chapter 4. For the completeness of this chapter, some information of the conducted experiment are provided here. The dataset is collected while a human subject is performing right-hand sequential finger-thumb opposition in the presence of a given stimulus signal. At a given time t, a

2D axial slice is analyzed instead of a 3D volume. The time point T is assigned to 99 and each axial slice consists of $M(= 63 \times 63)$ voxels. Each voxel has dimensions of $3.125 \times 3.125 \times 5$ (mm^3). From the original 21 slices, a slice expected to contain motor reactions to the motor stimulus is selected for analysis. To reduce a bias from edge effects, we perform a spatiotemporal tapering using a 3D Tukey-Hanning window. For the modeling of the BOLD response, a set of Laguerre functions with a time constant $a = 0.42$ is used. Since we have a 3D dataset, one dimension (1D) for time and 2D for space, we need to redefine the coordinate for space as $v = (v_1, v_2)$ and the model order for spatial cepstrum as $p = (p, q)$. These yield $R_{all} = (2p+1)(2q+1)n + 2pq + p + q + 1$ and $R_s = 2pq + p + q + n + 1$. Model orders are determined by AIC, producing $(L, n, p, q) = (3, 10, 7, 8)$. For more details about the observed data, the reader is referred to section 4.5.

5.6.1 Testing Space-Time Separability

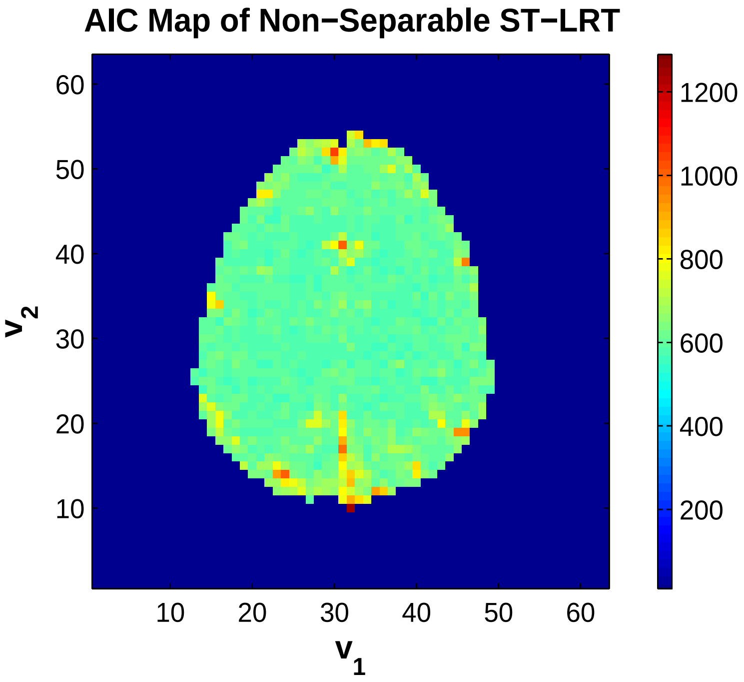
Since the determined model order is $(L, n, p, q) = (3, 10, 7, 8)$, $\mathcal{L}_S$ asymptotically follows a chi-square distribution with $R_{ns}(\triangleq R_{all} - R_s = 2540)$ degrees of freedom, χ^2_{2540} under the hypothesis for space-time separability, that is H_{S0}. The obtained value of $\mathcal{L}_S$ is 2591.8 and a determined threshold for a significance level $\alpha_S = 0.05$ is $\gamma(0.05) = 2658.4$, producing p-value is $0.23(\gg 0.05)$. Therefore, according to the developed test for space-time separability, there is no evidence that the tested AFNI dataset associated with motor responses is not under space-time separability. We fail to reject the null hypothesis, H_{S0}.

5.6.2 Model Comparison

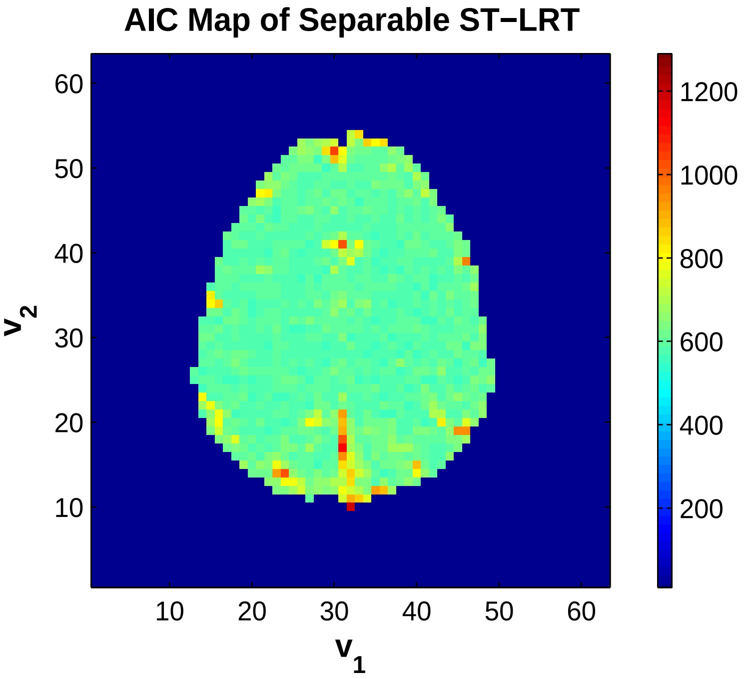
The result obtained from the proposed test for space-time separability, i.e., the claim that the tested AFNI dataset is under space-time separable, can be verified by the developed model comparison technique in Chapter 4, AIC map. To do that, two AIC maps are computed, one of which is from a ST-LRT model involving the non-separable random field. The other is from a ST-LRT model involving the separable random field. The AIC maps from ST-LRT models with and without space-time separability are shown on Fig.5.4(a) and Fig.5.4(b), respectively. Comparing Fig.5.4(a) and Fig.5.4(b) shows that there are no significant difference in AIC values of two ST-LRT models, indicating both ST-LRT models provide similar distances to the underlying unknown truth in the sense of average. Therefore, we have another evidence to support the claim made in the above section, saying that picked axial slice related to the motor responses from the AFNI dataset is under space-time separability.

5.7 Conclusions

We developed, for the first time in FMRI, a test procedure for space-time separability, which has been implicitly assumed without proper justifications, in the framework of the parametric cepstrum. The asymptotic power function of the proposed space-time separability test was also analyzed, being independent of nuisance parameters such as activation amplitudes. The developed procedure was tested by a simulation and a human dataset with model comparison.


AIC Map of Non−Separable ST−LRT

(a) AIC map of a non-separable ST-LRT model


AIC Map of Separable ST−LRT

(b) AIC map of a separable ST-LRT model

Figure 5.4: AIC maps from two different ST-LRT models : (a) ST-LRT model involving non-separable noise and (b) ST-LRT model involving separable noise.

5.8 Appendix I

In this section, we provide the detailed proofs for (5.20), (5.21), and (5.22). To do that, we consider the following sample covariance matrix from the regressors used for the modeling by the parametric cepstrum in (5.15):

$$\begin{aligned}\Sigma(x_{k,l}, z_{k,l}) \;&\triangleq\; \frac{1}{TM} \cdot \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} \begin{bmatrix} x_{k,l} \\ \\ z_{k,l} \end{bmatrix} [x_{k,l}^T \quad z_{k,l}^T] \\ &= \frac{1}{TM} \cdot \begin{bmatrix} \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} x_{k,l} x_{k,l}^T & \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} x_{k,l} z_{k,l}^T \\ \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} z_{k,l} x_{k,l}^T & \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} z_{k,l} z_{k,l}^T \end{bmatrix}, \end{aligned} \tag{5.32}$$

where $x_{k,l}$ is associated with cepstral coefficients in the non-separable region and $z_{k,l}$ is with cepstral coefficients in the separable region. According to the appendix of [55], the sample covariance matrix is expressed in a simple matrix form, implicitly assuming $0 < n \ll T$ and $0 < p \ll M$,

$$\Sigma(x_{k,l}, z_{k,l}) \;=\; \left[\begin{array}{c|cc} 2\mathbf{I}_{R_{ns}\times R_{ns}} & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 0 & 2\mathbf{I}_{(R_s-1)\times(R_s-1)} \end{array}\right], \tag{5.33}$$

where the identity matrix in the top and left corner defines $\sum_{k=0}^{T-1} \sum_{l=0}^{M-1} x_{k,l} x_{k,l}^T / TM$. Now, comparing (5.32) and (5.33) proves (5.20), (5.21), and (5.22).

5.9 Appendix II

In this section, proofs of (5.25) and (5.26) are provided. From (5.16), we have the following re-parameterization of the asymptotic negative log-likelihood function:

$$\begin{aligned} 2\ell(\tilde{\beta}^R, \tilde{\beta}^I, \tilde{f}^R, \tilde{f}^I, \theta^s, \theta^{ns}) \;&=\; \sum_{k=0}^{T-1} \sum_{l=0}^{M-1} x_{k,l}^T \theta^{ns} + z_{k,l}^T \theta^s + \frac{exp(-x_{k,l}^T \theta^{ns} - z_{k,l}^T \theta^s)}{TM} \\ &\times \left[\left\{ \tilde{y}_{k,l}^R - \left((\tilde{X}_k^R)^T \tilde{\beta}_l^R - (\tilde{X}_k^I)^T \tilde{\beta}_l^I \right) - \left((\tilde{\xi}_k^R)^T \tilde{f}_l^R - (\tilde{\xi}_k^I)^T \tilde{f}_l^I \right) \right\}^2 \right. \\ &+ \left. \left\{ \tilde{y}_{k,l}^I - \left((\tilde{X}_k^I)^T \tilde{\beta}_l^R + (\tilde{X}_k^R)^T \tilde{\beta}_l^I \right) - \left((\tilde{\xi}_k^I)^T \tilde{f}_l^R + (\tilde{\xi}_k^R)^T \tilde{f}_l^I \right) \right\}^2 \right], \end{aligned} \tag{5.34}$$

where, e.g., $\tilde{y}_{k,l} \triangleq \tilde{y}^R_{k,l} + j\tilde{y}^I_{k,l}$. It can be easily checked that the second order derivative of $2\ell(\,\cdot\,)$ with respect to $\tilde{f}^R_l$ (the l-th element of $\tilde{f}^R$) and θ^{ns}_i (the i-th element of θ^{ns}) is given by

$$\begin{aligned}\frac{\partial^2 2\ell}{\partial \tilde{f}^R_l \partial \theta^{ns}_i} &= \sum_{k=0}^{T-1} \frac{exp(-x^T_{k,l}\theta^{ns} - z^T_{k,l}\theta^s)}{TM} \\ &\times \left[2\left\{\tilde{y}^R_{k,l} - \left((\tilde{X}^R_k)^T\tilde{\beta}^R_l - (\tilde{X}^I_k)^T\tilde{\beta}^I_l\right) - \left((\tilde{\xi}^R_k)^T\tilde{f}^R_l - (\tilde{\xi}^I_k)^T\tilde{f}^I_l\right)\right\}\tilde{\xi}^R_k\right. \\ &+ \left.2\left\{\tilde{y}^I_{k,l} - \left((\tilde{X}^I_k)^T\tilde{\beta}^R_l + (\tilde{X}^R_k)^T\tilde{\beta}^I_l\right) - \left((\tilde{\xi}^I_k)^T\tilde{f}^R_l + (\tilde{\xi}^R_k)^T\tilde{f}^I_l\right)\right\}\tilde{\xi}^I_k\right] x_{k,l,i}.\end{aligned} \tag{5.35}$$

Since $\tilde{y}_{k,l}$ obeys a complex-valued Gaussian distribution given by

$$\tilde{y}_{k,l} \;\sim\; \mathcal{N}_c\left(\tilde{X}^T_k\tilde{\beta}_l + \tilde{\xi}^T_k\tilde{f}_l,\; TM\cdot F_{k,l}\right), \tag{5.36}$$

$\tilde{y}^R_{k,l}$ and $\tilde{y}^I_{k,l}$ are independent real-valued Gaussian distributions whose densities are given by

$$\tilde{y}^R_{k,l} \;\sim\; \mathcal{N}\left(\left((\tilde{X}^R_k)^T\tilde{\beta}^R_l - (\tilde{X}^I_k)^T\tilde{\beta}^I_l\right) + \left((\tilde{\xi}^R_k)^T\tilde{f}^R_l - (\tilde{\xi}^I_k)^T\tilde{f}^I_l\right),\; \frac{TM\cdot F_{k,l}}{2}\right), \tag{5.37}$$

$$\tilde{y}^I_{k,l} \;\sim\; \mathcal{N}\left(\left((\tilde{X}^I_k)^T\tilde{\beta}^R_l + (\tilde{X}^R_k)^T\tilde{\beta}^I_l\right) + \left((\tilde{\xi}^I_k)^T\tilde{f}^R_l + (\tilde{\xi}^R_k)^T\tilde{f}^I_l\right),\; \frac{TM\cdot F_{k,l}}{2}\right). \tag{5.38}$$

Taking the expectation of (5.35) yields the first equality in (5.25). Then, we have the second order derivative of $2\ell(\,\cdot\,)$ with respect to $\tilde{f}^I_l$ (the l-th element of $\tilde{f}^I$) and θ^{ns}_i (the i-th element of θ^{ns}) given by

$$\begin{aligned}\frac{\partial^2 2\ell}{\partial \tilde{f}^I_l \partial \theta^{ns}_i} &= \sum_{k=0}^{T-1} \frac{exp(-x^T_{k,l}\theta^{ns} - z^T_{k,l}\theta^s)}{TM} \\ &\times \left[2\left\{\tilde{y}^R_{k,l} - \left((\tilde{X}^R_k)^T\tilde{\beta}^R_l - (\tilde{X}^I_k)^T\tilde{\beta}^I_l\right) - \left((\tilde{\xi}^R_k)^T\tilde{f}^R_l - (\tilde{\xi}^I_k)^T\tilde{f}^I_l\right)\right\}\tilde{\xi}^I_k\right. \\ &- \left.2\left\{\tilde{y}^I_{k,l} - \left((\tilde{X}^I_k)^T\tilde{\beta}^R_l + (\tilde{X}^R_k)^T\tilde{\beta}^I_l\right) - \left((\tilde{\xi}^I_k)^T\tilde{f}^R_l + (\tilde{\xi}^R_k)^T\tilde{f}^I_l\right)\right\}\tilde{\xi}^R_k\right] (-x_{k,l,i}).\end{aligned} \tag{5.39}$$

Again, taking the expectation of (5.39) gives the second equality in (5.25). By the same way, for the second order derivatives of $2\ell(\,\cdot\,)$ with respect to $\tilde{\beta}^R_l$ and θ^{ns}_i, and with respect to $\tilde{\beta}^I_l$ and θ^{ns}_i, using the distributions of $\tilde{y}^R_{k,l}$ and $\tilde{y}^I_{k,l}$, (5.26) can be easily shown.

CHAPTER 6

Asymptotic Efficiency of Competing Detection Statistics

In functional Magnetic Resonance Imaging (FMRI), an important task in data analysis is to build up a detection statistic for activation to show regional responses to given temporal stimuli. In Chapter 2, we reviewed a method to construct an activation map from an existing SSK-DLM, which focused on only temporal correlation of observed data. In Chapter 3, we developed a new detection statistic, ST-LRT jointly considering spatial and temporal correlations. In this chapter, we discuss another important issue, performance comparison of ST-LRT and a detection statistic from SSK-DLM. The comparison is based on well established theoretical works on Asymptotic Relative Efficiency (ARE) and has different interpretations from comparisons by ROC curves in section 3.4 and AIC maps in section 4.4. The ARE measures an asymptotic ratio of sample sizes of two tests to achieve a pre-specified detection power for a given significance level.

6.1 Introduction

6.1.1 Background

Suppose that we have two consistent tests of a null hypothesis H_0 versus an alternative hypothesis H_1, where both hypotheses are simple. For a given power η, if n_A and n_B are the minimum sample sizes for test A and test B of size α to achieve the power, then relative efficiency of test A compared with test B is defined as n_B/n_A. If test B is the best test, e.g., from the Neyman-Pearson Lemma, then the ratio is called efficiency of test A. This concept of relative efficiency can be useful to assess a performance of a given test procedure or to compare two competing tests. Then Asymptotic Relative Efficiency (ARE) is defined as the limit of n_B/n_A as $n_A, n_B \to \infty$. For simple H_0 and H_1, an ARE is based on two quantities, a power η and a significance level α.

The idea of ARE can be extended to composite alternative hypothesis, thus a simple H_0 and a composite H_1. We considered this testing problem to build up an activation statistic in Chapter 2 and 3. In ARE, adopting the composite alternative hypothesis causes some complications as we now discuss details.

6.1.2 Preliminaries

The null hypothesis H_0 is typically represented as a specified family of distribution $\mathfrak{F}_0$ for the data. For samples of size n, the power function is defined as

$$\eta_n(T_n, \mathbf{F}) \triangleq P_{\mathbf{F}}(T_n \text{ rejects } H_0), \tag{6.1}$$

where T_n is a test procedure and $\mathbf{F}$ is a distribution function. For samples of size n and $\mathbf{F} \in \mathfrak{F}_0$, a size of the test is given by

$$\alpha_n(T_n, \mathfrak{F}_0) \triangleq \sup_{\mathbf{F} \in \mathfrak{F}_0} \eta_n(T_n, \mathbf{F}). \tag{6.2}$$

For samples of size n and $\mathbf{F} \notin \mathfrak{F}_0$, the probability of missing, as known as Type II error, is defined as

$$\beta_n(T_n, \mathbf{F}) \triangleq 1 - \eta_n(T_n, \mathbf{F}). \tag{6.3}$$

Usually, we are interested in the comparison of two tests which are consistent and unbiased. The consistency means that, for a fixed $\mathbf{F} \notin \mathfrak{F}_0$, $\beta_n(T_n, \mathbf{F}) \to 0$ as $n \to \infty$.

The unbiasedness means that, for a $\mathbf{F} \notin \mathfrak{F}_0$, $\eta_n(T_n, \mathbf{F}) \geq \alpha_n(T_n, \mathfrak{F}_0)$.

We consider the following quantities.

$$\alpha = \lim_{n\to\infty} \alpha_n(T_n, \mathfrak{F}_0), \qquad \beta = \lim_{n\to\infty} \beta_n(T_n, \mathbf{F}^n) \tag{6.4}$$

where $\mathbf{F}^n$ is a distribution function under the alternative hypothesis H_1 for samples of size n. According to values of α and β, and a behavior of $\mathbf{F}^n$, an ARE is specified. In details, with respect to α, the cases of $\alpha = 0$ and $\alpha > 0$ are distinguished. With respect to β, the cases of $\beta = 0$ and $\beta > 0$ are distinguished. With respect to $\mathbf{F}^n$, the cases of a fixed $\mathbf{F}^n$ regardless of the number of samples n, and $\mathbf{F}^n \to \mathfrak{F}_0$ are distinguished. In this chapter, we consider Pitman's ARE which is commonly used and under the following specifications. As the number of samples n approaches to infinity,

$$\alpha_n \to \alpha > 0, \qquad \beta_n \to \beta > 0, \qquad \mathbf{F}^n \to \mathfrak{F}_0. \tag{6.5}$$

In the Pitman approach, the key tool is Central Limit Theory (CLT). Other AREs can be defined under other specifications of α, β, and $\mathbf{F}^n$. For example, Bahadur's ARE, another commonly used asymptotic measure, is under the following conditions.

$$\alpha_n \to 0, \qquad \beta_n \to \beta > 0, \qquad \mathbf{F}^n = \mathbf{F} \text{ fixed}, \tag{6.6}$$

as n approaches to infinity. For the Bahadur approach, Large Deviation Theory (LDT) is employed. For more technical stuffs, the reader is referred to [50].

6.1.3 Pitman's Asymptotic Relative Efficiency (ARE)

Suppose that the distributions $\mathbf{F}$ under consideration is indexed by a parameter set Θ and consider the following simple null hypothesis H_0 and composite alternative hypothesis H_1,

$$H_0 \;:\; \theta = \theta_0, \qquad H_1 \;:\; \theta > \theta_0. \tag{6.7}$$

We consider two test sequences $T_A = \{T_{An}\}$ and $T_B = \{T_{Bn}\}$ satisfying the following Pitman's conditions. The parameter θ we are interested in is assumed close to θ_0, thus any θ such that $\theta_0 \le \theta \le \theta_0 + \delta$ is considered.

Pitman's conditions

(P1) For some continuous and strictly increasing distribution function $\mathbf{G}$, and functions $\mu_n(\theta)$ and $\sigma_n(\theta)$, the distribution of $(T_n - \mu_n(\theta))/\sigma_n(\theta)$ converges to $\mathbf{G}$ uniformly in $[\theta_0, \theta_0 + \delta]$, i.e.,

$$\sup_{\theta_0 \le \theta \le \theta_0 + \delta} \sup_{-\infty < t < \infty} \left| P\left(\frac{T_n - \mu_n(\theta)}{\sigma_n(\theta)} \le t \right) - \mathbf{G}(t) \right| \to 0, \tag{6.8}$$

as $n \to \infty$.

(P2) For $\theta \in [\theta_0, \theta_0 + \delta]$, $\mu_n(\theta)$ is k times differentiable, with $\mu_n^{(1)}(\theta_0) = \cdots = \mu_n^{(k-1)}(\theta_0) = 0 < \mu_n^{(k)}(\theta_0)$.

(P3) For some function $d(n) \to \infty$ and some constant $c > 0$,

$$c = \lim_{n \to \infty} \frac{d(n)\sigma_n(\theta_0)}{\mu_n^{(k)}(\theta_0)}, \tag{6.9}$$

where c is called the **efficacy** of T_n.

(P4) For $\theta_n = \theta_0 + O([d(n)]^{-\frac{1}{k}})$, as $n \to \infty$,

$$\mu_n^{(k)}(\theta_n) \sim \mu_n^{(k)}(\theta_0), \qquad \sigma_n(\theta_n) \sim \sigma_n(\theta_0). \tag{6.10}$$

■

We have two main theorems by Pitman and Noether.

Pitman-Noether Theorem I

Let a test sequence T satisfy above four conditions, (P1)-(P4) and consider testing H_0 by critical region $\{T_n > \lambda_{\alpha_n}\}$ with

$$\alpha_n = P_{\theta_0}(T_n > \lambda_{\alpha_n}) \to \alpha,$$

where $\alpha \in (0,1)$. For $0 < \beta < 1 - \alpha$ and $\theta_n = \theta_0 + O([d(n)]^{-\frac{1}{k}})$, we have

$$\beta_n(\theta_n) = P_{\theta_n}(T_n \leq \lambda_{\alpha_n}) \rightarrow \beta$$

if and only if

$$\frac{(\theta_n - \theta)^k}{k!} \cdot \frac{d(n)}{c} \rightarrow \mathbf{G}^{-1}(1-\alpha) - \mathbf{G}^{-1}(\beta). \tag{6.11}$$

■

Pitman-Noether Theorem II

Let two test sequences T_A and T_B satisfy above conditions, (P1)-(P4) with common $\mathbf{G}$, k, and $d(n)$ in (P1)-(P3). Let $d(n) = n^q,\ q > 0$. Then, Pitman's ARE of T_A compared with T_B is given by

$$e_P(T_A, T_B) = \left(\frac{c_B}{c_A}\right)^{\frac{1}{q}}. \tag{6.12}$$

■

Detailed proofs of Pitman-Noether theorems are provided in [50].

6.2 Pitman's ARE of ST-LRT and t-statistic

We have two competing detection statistics, ST-LRT and t-statistic from SSK-DLM. To make our discussion simple, we have the following assumptions in the observed signal model in (4.2).

(A1) The baseline and temporal linear drift are assumed known and dropped.

(A2) The simplest approach, the parametric approach in (1.7) is used for the BOLD response modeling, thus $L = 1$.

(A3) The PSD $F_{k,l}$ is assumed known and under space-time separability.

Thus, (4.2) reduces to

$$y_{t,v} = \xi_t f_v + w_{t,v}, \tag{6.13}$$

where ξ_t and f_v are both scalar, and $w_{t,v}$ a stationary Gaussian noise which is spatiotemporally correlated and space-time separable.

ST-LRT

Based on the assumptions (A1)-(A3), by plugging (6.13) into (3.32), one arrives at a ST-LRT whose form is given by

$$LRT_v \triangleq \left(\frac{\sum_{t=0}^{T-1} \xi_t^F \cdot \varepsilon_{t,v}}{\sqrt{\sum_{t=0}^{T-1} \left(\xi_t^F\right)^2}} \right)^2, \tag{6.14}$$

where ξ_t^F denotes a temporally whitened ξ_t and $\varepsilon_{t,v}$ is a spatiotemporally whitened $y_{t,v}$, namely,

$$\xi_t^F \triangleq q_t \circledast \xi_t, \qquad \varepsilon_{t,v} \triangleq K_v \circledast_s (q_t \circledast y_{t,v}). \tag{6.15}$$

Here, $\circledast$ and $\circledast_s$ are temporal and spatial circular convolution, respectively. Under space-time separability, a temporal whitening filter q_t and a spatial whitening kernel K_v are defined as

$$q_t \overset{DFT}{\longleftrightarrow} \tilde{q}_k, \qquad |\tilde{q}_k|^2 = \frac{1}{F_k}, \qquad K_v \overset{DFT}{\longleftrightarrow} \frac{1}{\sqrt{G_l}}, \tag{6.16}$$

where F_k denotes pure temporal PSD and G_l is pure spatial PSD. For convenience, we define the following detection statistic from (6.14),

$$L_v \triangleq \sqrt{LRT_v} = \left(\frac{\sum_{t=0}^{T-1} \xi_t^F \cdot \varepsilon_{t,v}}{\sqrt{\sum_{t=0}^{T-1} \left(\xi_t^F\right)^2}} \right). \tag{6.17}$$

Then, instead of calculating Pitman's ARE of LRT_v and F-statistic, $\mathcal{F}_v$, we compare L_v and t-statistic, $\mathcal{T}_v$. By plugging (6.13) into (6.17), we obtain

$$L_v = (K \circledast_s f)_v \cdot \sqrt{T\sigma_\xi^2} \; + \; \mathcal{N}(0,1), \tag{6.18}$$

where $\sigma_\xi^2 \triangleq \frac{1}{T}\sum_{t=0}^{T-1}\left(\xi_t^F\right)^2$ and the first term means a non-centrality parameter. Therefore, under the null hypothesis H_0, i.e., $f_v = 0$, L_v obeys $\mathcal{N}(0,1)$. Under the alternative hypothesis H_1, L_v obeys a normal distribution with mean $\left(K \circledast_s f\right)_v \cdot \sqrt{T\sigma_\xi^2}$.

t-statistic from SSK-DLM

After a SSK is performed, if a temporal whitening filter is adopted in the DLM step, from (2.10), t-statistic has a form of

$$\mathcal{T}_v \triangleq \frac{\sum_{t=0}^{T-1}(\xi_t^F)(K_v^G *_s y_{t,v})^F}{\sqrt{\sum_{t=0}^{T-1}\left(\xi_t^F\right)^2}}, \tag{6.19}$$

where K_v^G denotes a Gaussian amplitude kernel and $*_s$ is a spatial linear convolution. Notice that the temporal whitening is performed for a spatially smoothed dataset. Plugging (6.13) into (6.19) allows

$$\mathcal{T}_v = \left(K^G *_s f\right)_v \cdot \sqrt{T\sigma_\xi^2} + \mathcal{N}(0,1). \tag{6.20}$$

Thus, under H_0, i.e., $f_v = 0$, $\mathcal{T}_v$ obeys $\mathcal{N}(0,1)$. Under H_1, $\mathcal{T}_v$ obeys a normal distribution with mean $\left(K^G *_s f\right)_v \cdot \sqrt{T\sigma_\xi^2}$. Note that the only difference between (6.18) and (6.20) is spatial operator in non-centrality parameter.

6.2.1 Efficacy of L_v

We check if L_v in (6.18) satisfies Pitman's conditions. To do that, we focus on a L_v at a particular voxel position v and recall

$$\mathbf{E}(L_v) = \left(K \circledast_s f\right)_v \cdot \sqrt{T\sigma_\xi^2}, \qquad var(L_v) = 1. \tag{6.21}$$

In Pitman's conditions, we have $\theta_0 = 0$, $\theta = f_v$, and $n = T$, the number of time points. We define $\mu_T \triangleq \mathbf{E}(L_v)$ and $\sigma_T \triangleq \sqrt{var(L_v)}$.

Firstly, we have

$$\frac{L_v - \mu_T}{\sigma_T} \;\sim\; \mathcal{N}(0,1), \qquad \mathbf{G}(t) = \mathbf{\Phi}(t), \tag{6.22}$$

where $\mathbf{\Phi}(t)$ denotes a CDF of $\mathcal{N}(0,1)$. Thus, (P1) is satisfied.

Secondly, we have

$$\frac{\partial \mu_T}{\partial f_v} = K_0 \cdot \sqrt{T\sigma_\xi^2} \;>\; 0, \tag{6.23}$$

thus, (P2) is satisfied and $k = 1$.

Thirdly, for a function $d(T) \triangleq \sqrt{T}$, we have

$$c_L = \lim_{T\to\infty} \frac{\sqrt{T}\cdot 1}{K_0 \cdot \sqrt{T\sigma_\xi^2}} = \frac{1}{K_0\sigma_\xi} \;>\; 0, \tag{6.24}$$

thus, (P3) is satisfied and the efficacy of L_v is c_L.

Finally, for $f_v = O(\frac{1}{\sqrt{T}})$, since $\mu_T^{(1)}$ and $\sigma_T(f_v)$ are independent of f_v, (P4) is satisfied for all T. Therefore, all Pitman's conditions are satisfied.

We check the Pitman-Noether I theorem now. Let f_v has the following form, with a fixed constant c_f,

$$f_v = \frac{c_f}{\sqrt{T}}. \tag{6.25}$$

For a given α and β, we can choose a specific value of c_f satisfying

$$f_v \cdot \frac{\sqrt{T}}{c_L} = \mathbf{\Phi}^{-1}(1-\alpha) - \mathbf{\Phi}^{-1}(\beta). \tag{6.26}$$

Therefore, Pitman-Noether's first theorem holds and L_v has a size α and a power $1-\beta$ asymptotically.

6.2.2 Efficacy of $\mathcal{T}_v$

We check if $\mathcal{T}_v$ in (6.20) satisfies Pitman's conditions. Since (6.18) and (6.20) have the same form except the spatial kernel, we repeat the same procedure to obtain c_L

in (6.24). We define $\mu_T \triangleq \mathbf{E}(\mathcal{T}_v)$ and $\sigma_T \triangleq \sqrt{var(\mathcal{T}_v)} = 1$.

Firstly, we have

$$\frac{\mathcal{T}_v - \mu_T}{\sigma_T} \;\sim\; \mathcal{N}(0,1), \qquad \mathbf{G}(t) = \mathbf{\Phi}(t), \tag{6.27}$$

satisfying (P1).

Secondly, we have

$$\frac{\partial \mu_T}{\partial f_v} = K_0^G \cdot \sqrt{T\sigma_\xi^2} \;>\; 0, \tag{6.28}$$

satisfying (P2) with $k = 1$.

Thirdly, for a function $d(T) = \sqrt{T}$, we have

$$c_{\mathcal{T}} = \lim_{T\to\infty} \frac{\sqrt{T}\cdot 1}{K_0^G \cdot \sqrt{T\sigma_\xi^2}} = \frac{1}{K_0^G \sigma_\xi} \;>\; 0, \tag{6.29}$$

satisfying (P3) and defining the efficacy of $\mathcal{T}_v$, $c_{\mathcal{T}}$.

Finally, for $f_v = O(\frac{1}{\sqrt{T}})$, since $\mu_T^{(1)}$ and $\sigma_T(f_v)$ are independent of f_v, (P4) is satisfied for all T. Thus, all Pitman's conditions are satisfied.

We check the Pitman-Noether I theorem now. For a given α and β, by defining $f_v = c_f/\sqrt{T}$, we can choose a specific value of c_f satisfying

$$f_v \cdot \frac{\sqrt{T}}{c_{\mathcal{T}}} = \mathbf{\Phi}^{-1}(1-\alpha) - \mathbf{\Phi}^{-1}(\beta). \tag{6.30}$$

Therefore, Pitman-Noether's first theorem holds and $\mathcal{T}_v$ has a size α and a power $1-\beta$ asymptotically.

Pitman's ARE of L_v and $\mathcal{T}_v$

As we showed above, L_v and $\mathcal{T}_v$ satisfied all Pitman's conditions with common $\mathbf{G}(=\mathbf{\Phi})$, $k(=1)$, and $d(T)(=\sqrt{T})$. Therefore, according to Pitman-Noether Theorem II, Pitman's ARE of $\mathcal{T}_v$ compared with L_v is given by

$$e_P(\mathcal{T}_v, L_v) = \left(\frac{c_L}{c_{\mathcal{T}}}\right)^2 = \left(\frac{K_0^G}{K_0}\right)^2. \tag{6.31}$$

Note that the value of K_0 depends on a given dataset. We apply the developed Pitman's ARE, $e_P(\mathcal{T}_v, L_v)$ to the simulated dataset discussed in section 3.4. Based on the setup of the simulations, the following evaluation can be shown. In the case of spatial white noise with $\sigma = 1$, we obtain $e_P(\mathcal{T}_v, L_v) = 0.2824$. In the case of spatial colored noise with a known Gaussian ACF, namely $\gamma_v \triangleq exp(-v^2/2.254)$, we can obtain $e_P(\mathcal{T}_v, L_v) = 0.0181$. Therefore, we argue that $\mathcal{T}_v$ requires about $3.5(= 1/0.2824)$ times as many samples as does the L_v in the case of spatial white noise. In the case of spatial colored noise, $\mathcal{T}_v$ requires about $55(= 1/0.0181)$ times as many samples as does L_v. These values of $e_P(\mathcal{T}_v, L_v)$ support the results of simulations performed on Fig.3.2(a) and Fig.3.3(a).

6.3 Conclusions

We developed a method to compare the asymptotic efficiency of two competing detection statistics, ST-LRT and a statistic from SSK-DLM. For a simple signal model with spatially and temporally correlated noise, a formula for Pitman's ARE was derived. The developed ARE was computed for the synthesized dataset used in simulations of Chapter 3. The Pitman's ARE showed that ST-LRT was a much more efficient test procedure than a detection statistic from SSK-DLM for both spatial white noise and spatial colored noise. Therefore, ST-LRT can reduce experiment durations in FMRI, allowing subjects' less exposure to strong magnetic fields.

CHAPTER 7

Rician Modeling and Activation Detection

In FMRI data analysis, most statistical methods are based on magnitude voxel time courses, which are modeled by a Gaussian distribution. Since the magnitude images are produced from complex valued data, in fact, they obey a Rician distribution, which can only be approximated as a Gaussian distribution under the assumption that signal to noise ratio (SNR) is high. Therefore, statistical methods based on Gaussian modeling may perform poorly when the SNR is low. In this chapter, we develop an exact Rician maximum likelihood with an expectation maximization (EM) algorithm. The resulting procedure is remarkably simple and can be easily extended from solutions based on Gaussian modeling. Using estimated parameters, we build up a detection statistic for activation and analyze its asymptotic power. We perform simulations to compare the proposed Rician-EM and conventional Gaussian modeling.

7.1 Introduction

After an image reconstruction of observed FMRI data in k-space is performed, available images for statistical data analysis are complex valued. By taking the magnitude of complex value at each time point and at each voxel location, magnitude images are obtained. From the beginning of FMRI data analysis, most attention has focused on the analysis of the magnitude voxel time courses with a Gaussian distribution, which in fact obey a Rician distribution. It is known that a Rician probability density function (PDF) is well approximated by a Gaussian PDF when SNR is high [40]. Based on the Gaussian modeling, dynamic linear model (DLM) has been a dominant framework to model the BOLD response and to build up a detection statistic for activation such as t-statistic or F-statistic.

The conventional Gaussian modeling has been justified as follows. In practice, it is known that a typical FMRI dataset has sufficiently high SNRs to make the Gaussian approximation appropriate. However, since we have a fundamental tradeoff between spatial resolution and SNR, as demands for high resolution FMRI rise, the necessity of statistical modeling valid for the low range of SNRs increases. And it is reported that there may exist some regions of the human brain which have significant signal dropouts, decreasing SNRs in those regions, even in images with a moderate spatial resolution. In addition, methods based on non-BOLD mechanisms such as arterial spin labeling (ASL) is known not to have sufficiently high SNRs to apply Gaussian approximation to a Rician distribution.

In this chapter, therefore, we develop a robust method to construct activation detection for varying SNRs, which is based on Rician signal and noise modeling. Since the Rician PDF involves the zeroth order modified Bessel function of the first kind, the direct maximization of Rician log-likelihood function is a non-trivial statistical estimation. We tackle this problem with EM framework. From the proposed parameter estimation, then we construct a detection statistic for activation. We start discussions by reviewing some previous works on the Rician modeling.

7.1.1 Literature Review

For more exact modeling of the magnitude voxel time course which holds for all ranges of SNRs, a couple of methods based on Rician distributed models have been suggested in the FMRI literature. [54] developed an activation statistic from a Rician distribution model through likelihood ratio test (LRT). For a simple signal model involving one baseline and one activation amplitude, [54] used non-standard numerical optimizations to estimate parameters under null hypothesis (no activation exists) and alternative hypothesis (activation exists). However, the model used in [54] is highly constrained and the non-standard maximization is computationally expensive. The computational cost becomes severely demanding when the signal model is extended to a complicated one, e.g., a signal model involving temporal drift and multiple activation amplitudes or when the number of voxels in a region of interest (ROI) increases. Similar approaches based LRTs are found in [16, 17]. In [46], to estimate parameters for signal and noise model, an iterative method based on a Taylor series expansion of a Rician PDF was proposed instead of maximizing Rician log-likelihood function. Some approaches using complex valued data were suggested to build up activation maps in FMRI data analysis. [37, 47, 45] created detection statistics through LRTs with complex Gaussian distributions, which have different philosophies from approaches based on Rician modeling including our method.

In this chapter, we make exact parameter estimation from a Rician distributed model through an EM algorithm [15], which enables us to avoid computationally expensive numerical optimization, leading to an activation statistic through a LRT. The proposed Rician-EM approach allows a very simple iteration and a nice interpretation of the iteration in a very natural way from existing Gaussian modeling. In addition, we analyze the asymptotic power function of the proposed detection statistic, showing the effects of a nuisance signal component (baseline) and signal components of interest (activation amplitudes) on the detection power. For simplicity, as in previous works without exception [46, 37, 47, 45], noise in voxel time courses is assumed spatially and temporally independent. Although assuming white noise along time is an oversimplification of temporal characteristics of voxel time courses in FMRI data analysis, the work in this chapter can be a good starting point for future discussions to develop activation detection from a generalized model with a temporally colored Rician noise.

The remainder of this chapter is organized as follows. In section 7.2, we introduce signal and noise model formulations. In section 7.3, parameter estimations via the EM algorithm from a Rician distributed model is proposed. Then, section 7.4 describes a method to construct a detection statistic for activation based on the proposed Rician-EM approach. In section 7.5, the asymptotic power of the proposed detection statistic is analyzed. Then, in section 7.6, we perform simulations to compare performances of the Rician-EM method and existing Gaussian modeling. Finally, conclusions will be drawn in section 7.7. Mathematical details are provided in Appendix 7.8-7.10.

7.1.2 Acronyms and Notations

We collect acronyms and notations frequently used for the rest of this chapter in this section. ANR means activation to noise ratio; BNR is baseline to noise ratio; EM means expectation maximization; LRT is likelihood ratio test; GOA means goodness of approximation; LS is least square; MLE means maximum likelihood estimate; MSE is mean squared error; PDF means probability density function; SNR is signal to noise ratio.

In this chapter, spatial and temporal independence is assumed for Rician modeling and activation detection. Therefore, on the contrary to Chapter 3 and 4, we can model each voxel time course separately and then combine resultant detection statistics to make a spatial plot of them, namely an activation map. Since we concentrate on each voxel time series, we drop the index v for a voxel location and use different notations from those of Chapter 3, 4, and 5 dealing with a Gaussian distributed noise model with spatial and temporal correlations.

For a particular voxel position which we are interested in, $y_{c,t}$ denotes a complex valued FMRI measurement at a time t. We assume $y_{c,t}$ is observed from an one-dimensional (1D) temporal line indexed by $\{1, \ldots, n\}$. The magnitude measurement of $y_{c,t}$ is denoted as y_t and the phase measurement of $y_{c,t}$ is denoted as ϕ_t. We define vectors $y = [y_1, \ldots, y_n]^T$, denoting a collected magnitude time series, and $\phi = [\phi_1, \ldots, \phi_n]^T$, denoting a phase time series. A column vector β contains signal components such as baseline and activation amplitudes. x_t is an associated regressor at a time t, defining $\mathbf{X} = [x_1, \ldots, x_n]^T$, and σ^2 is the variance of noise. The null hypothesis H_N means that a particular voxel we are interested in is not activated by a given temporal stimulus. The alternative hypothesis H_A means that that voxel is activated. Under H_A, from Gaussian modeling, $\widetilde{\beta}_A$ and $\widetilde{\sigma}^2_A$ denote the MLEs of β and σ^2, respectively. From Rician-EM, $\beta_{A,1}$ and $\sigma^2_{A,1}$ are estimates of β and σ^2 for given estimates, $\beta_{A,0}$ and $\sigma_{A,0}$ at each iteration. Under H_N, parameter estimates are similarly defined. $\mathbf{Q}_A$ denotes a surrogate function obtained from an E-step and $\mathbf{W}_{A,0}$ is a weight matrix under H_A. $\mathbf{Q}_N$ and $\mathbf{W}_{N,0}$ are under H_N. $\mathcal{N}(\mu, \Sigma)$ denotes a real-valued Gaussian distribution with mean μ and covariance matrix Σ.

7.2 Signal and Noise Model Formulations

7.2.1 Complex Signal and Noise Model

Dropping the voxel index v for simplicity, we first consider a complex valued signal and noise model and then discuss its magnitude model. After image reconstruction is performed, at a time point t, a complex valued measurement model is defined as in [47],

$$y_{c,t} \;=\; (x_t^T\beta\cos\theta + \eta_{r,t}) \;+\; j(x_t^T\beta\sin\theta + \eta_{i,t}), \tag{7.1}$$

$$x_t^T\beta \;\triangleq\; m + bt + s_t, \quad [\eta_{r,t}, \eta_{i,t}]^T \sim \mathcal{N}(0, \sigma^2\mathbf{I}), \tag{7.2}$$

where $x_t^T\beta$ contains two types of signals, one of which involves nuisance components such as baseline m and temporal linear drift bt. The other involves activation related component s_t, e.g., the BOLD response, describing reactions of a subject's brain to a given temporal stimulus. θ denotes a phase imperfection due to magnetic field inhomogeneity in a MRI scanner and is assumed constant over time. Two noise terms, $\eta_{r,t}$ and $\eta_{i,t}$ are assumed stationary and independent Gaussian, where σ^2 denotes their common noise variance.

7.2.2 Hemodynamic Response Model

Several approaches have been suggested for the modeling of the BOLD response, s_t in (7.2) until now. The simplest parametric approach with a canonical HRF [11], a method using a FIR filter [14, 26], and Laguerre modeling [56] were reviewed in Chapter 1. To cover these three approaches and to provide a unified framework for the modeling of the BOLD response, we consider the following generalized representation:

$$s_t = \left(\sum_{i=1}^{L} h_{i,t} f_i\right) * c_t \;\triangleq\; \sum_{i=1}^{L} \xi_{i,t} f_i = \xi_t^T f, \tag{7.3}$$

where L is the number of basis functions to model the BOLD response, $h_{i,t}$ is the i-th temporal basis function, and f_i is the associated activation amplitude. c_t denotes given temporal stimuli in the experiment. By plugging (7.3) into (7.1) and (7.2), one arrive at a compact and parameterized representation which we use in this chapter,

$$y_{c,t} \;=\; (x_t^T\beta\cos\theta + \eta_{r,t}) \;+\; j(x_t^T\beta\sin\theta + \eta_{i,t}), \tag{7.4}$$

where

$$x_t^T = [1, t, \mid \xi_t^T], \qquad \beta^T = [m, b, \mid f^T], \tag{7.5}$$

$$\xi_t = [\xi_{1,t}, \ldots, \xi_{L,t}]^T, \quad f = [f_1, \ldots, f_L]^T. \tag{7.6}$$

7.2.3 Magnitude Signal and Noise Model

Now we consider a typical measurement model in FMRI data analysis. For a given time t, the magnitude of (7.4) is defined by

$$y_t \;\triangleq\; \sqrt{(x_t^T\beta\cos\theta + \eta_{r,t})^2 + (x_t^T\beta\sin\theta + \eta_{i,t})^2}, \tag{7.7}$$

resulting in a collected time course at a particular voxel location of interest. It is well known that y_t obeys a Rician distribution whose PDF is defined as, for $y_t > 0$,

$$p(y_t) \;=\; \frac{y_t}{\sigma^2} e^{-\frac{y_t^2 + (x_t^T\beta)^2}{2\sigma^2}} I_0\left(\frac{y_t x_t^T\beta}{\sigma^2}\right), \tag{7.8}$$

where $I_0(z)$ is the zeroth order modified Bessel function of the first kind. Some mathematical materials related to $I_0(z)$ are given in Appendix 7.8. For a time course $y \triangleq [y_1, \ldots, y_n]^T$, this leads to a Rician log-likelihood function under the assumption of temporal independence, that is,

$$\log\mathscr{L} \triangleq -n\log\sigma^2 + \sum_{t=1}^{n}\log y_t - \sum_{t=1}^{n}\left\{\frac{y_t^2 + (x_t^T\beta)^2}{2\sigma^2} - \log I_0\left(\frac{y_t x_t^T\beta}{\sigma^2}\right)\right\}. \tag{7.9}$$

7.3 Parameter Estimation

In this section, we discuss methods to estimate parameters of Rician PDF in (7.8), β (signal parameter) and σ^2 (noise parameter). First, we review a conventional statistical inference based on the Gaussian approximation of Rician PDF and then provide a new method to estimate the necessary parameters through EM algorithm. To build up a detection statistic for activation to test if a particular voxel is activated by given temporal stimuli, we consider the following two hypotheses and perform parameter estimation under each hypothesis.

$$H_N \ : \ \mathbf{C}\beta = 0, \qquad vs \qquad H_A \ : \ \mathbf{C}\beta \neq 0, \tag{7.10}$$

where the matrix $\mathbf{C}$ is of full rank and has the size of $r \times (L+2)$. By selecting $\mathbf{C}$, we can specify the hypothesis testing. For example, if we have a signal model with $L = 2$, that is $\beta = [m, b, f_1, f_2]^T$, one possibility of choosing the constraint matrix is

$$\mathbf{C} = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right], \tag{7.11}$$

where the first row in $\mathbf{C}$ is associated with f_1 (the first activation amplitude) and the second row is associated with f_2 (the second activation amplitude). The first two columns correspond to nuisance signal components, baseline m and temporal drift bt. Then, H_N states the voxel we are interested in is not activated by given temporal stimuli and the alternative hypothesis, H_A says the voxel is activated.

7.3.1 Gaussian Modeling

It is known that a Rician PDF in (7.8) can be approximated by a Gaussian PDF for a large value of SNR($\triangleq x_t^T\beta/\sigma$). According to [47], a simple rationale behind this approximation can be described as follows. The magnitude signal and noise model in (7.7) is rearranged as

$$\begin{aligned} y_t &= \sqrt{\left(x_t^T\beta\right)^2 + 2x_t^T\beta(\eta_{r,t}\cos\theta + \eta_{i,t}\sin\theta) + \eta_{r,t}^2 + \eta_{i,t}^2}, \\ &= x_t^T\beta \cdot \sqrt{1 + \frac{2(\eta_{r,t}\cos\theta + \eta_{i,t}\sin\theta)}{x_t^T\beta} + \frac{\eta_{r,t}^2 + \eta_{i,t}^2}{\left(x_t^T\beta\right)^2}}. \end{aligned} \tag{7.12}$$

If the signal component $x_t^T\beta$ dominates two noise terms $\eta_{r,t}$ and $\eta_{i,t}$, i.e., SNR is high, by a first order Taylor series expansion of the term inside the root square, (7.12) reduces to a classical linear regression equation,

$$y_t = x_t^T\beta + \eta_t, \tag{7.13}$$

where $\eta_t(\triangleq \eta_{r,t}\cos\theta + \eta_{i,t}\sin\theta)$ obeys a Gaussian distribution with zero mean and σ^2 variance or in matrix form,

$$y = \mathbf{X}\beta + \eta, \tag{7.14}$$

where $y = [y_1, \ldots, y_n]^T$, $\mathbf{X} \triangleq [x_1, \ldots, x_n]^T$, and $\eta \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.

It is well known that least square (LS) allows maximum likelihood estimates (MLEs) of β and σ^2 under H_A given as follows [49].

$$\widetilde{\beta}_A = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y, \tag{7.15}$$

$$\widetilde{\sigma}_A^2 = \frac{1}{n} \cdot \left(y - \mathbf{X}\widetilde{\beta}_A\right)^T (y - \mathbf{X}\widetilde{\beta}_A). \tag{7.16}$$

Under H_N, with the linear constraint specified by $\mathbf{C}$, MLEs are given by

$$\widetilde{\beta}_N = \mathbf{\Phi} \cdot \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y, \tag{7.17}$$

$$\widetilde{\sigma}_N^2 = \frac{1}{n} \cdot \left(y - \mathbf{X}\widetilde{\beta}_N\right)^T (y - \mathbf{X}\widetilde{\beta}_N), \tag{7.18}$$

where a matrix $\mathbf{\Phi}$ is defined as

$$\mathbf{\Phi} = \mathbf{I} - \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{P}, \quad \mathbf{P} = \mathbf{C}^T\left[\mathbf{C}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{C}^T\right]^{-1}\mathbf{C}.$$

These MLEs are used to build up a detection statistic for activation based on the Gaussian modeling in section 7.4.

7.3.2 Rician Modeling with EM Algorithm

Since a direct maximization of the Rician log-likelihood function in (7.9) to obtain MLEs of β and σ^2 requires a non-standard numerical optimization, we consider an indirect method for parameter estimation from the Rician PDF, an EM algorithm. The EM algorithm has guaranteed convergence properties and in the present case yields remarkably simple iterations. The main idea of the EM algorithm is to replace a difficult direct maximization of the Rician likelihood function with a sequence of easier maximizations of surrogate functions. Starting from deriving an iterative equation without a constraint under H_A (activation exists), we move the discussion into a constrained case under H_N (no activation exists) with a Lagrange multiplier later.

EM Algorithm under H_A

For a given magnitude time course y and a phase time course ϕ, from the complex valued time series $y_{c,t}$ in (7.4), it can be shown that so called complete likelihood $\mathscr{L}_c$, namely joint PDF of y and ϕ is given by

$$\log \mathscr{L}_c = -n \log(2\pi\sigma^2) + \sum_{t=1}^{n} \log y_t - \frac{1}{2\sigma^2} \sum_{t=1}^{n} \left\{ y_t^2 + (x_t^T\beta)^2 - 2y_t x_t^T \beta \cos(\phi_t - \theta) \right\}. \tag{7.19}$$

In the EM framework, since only the magnitude voxel time series y is available for data analysis, the phase time course ϕ is treated as unobserved data or missing data. The EM algorithm is composed of two steps, one of which is expectation step (E-step), defining a surrogate criterion based on a conditional expectation of $\log \mathscr{L}_c$. The other is maximization step (M-step), finding a point to give a maximum of the surrogate function defined in the E-step [15]. We first consider E-step and then discuss M-step.

Following E-step, taking a conditional expectation of $\log \mathscr{L}_c$ in (7.19) yields the following surrogate function:

$$\mathbf{Q}_A = -n\log(2\pi\sigma^2)+\sum_{t=1}^{n}\log y_t-\frac{1}{2\sigma^2}\left\{\sum_{t=1}^{n}y_t^2+\sum_{t=1}^{n}\left(x_t^T\beta\right)^2-2y_t x_t^T\beta\cos(\theta_{A,0}-\theta)\cdot w_{A,0,t}\right\}, \tag{7.20}$$

where a weight is defined as

$$w_{A,0,t} \triangleq \frac{I_1(y_t x_t^T\beta_{A,0}/\sigma_{A,0}^2)}{I_0(y_t x_t^T\beta_{A,0}/\sigma_{A,0}^2)} \tag{7.21}$$

and $\beta_{A,0}$, $\theta_{A,0}$, and $\sigma^2_{A,0}$ are given parameter estimates in each E-step. $I_1(z)$ denotes the first order modified Bessel function of the first kind. Here, it can be shown that the ratio, $I_1(z)/I_0(z)$ is a strictly increasing function and approaches to 1 as z goes to infinity. The numerical evaluation of $I_1(z)/I_0(z)$ is given on Fig.7.1. The details of derivation of the surrogate function $\mathbf{Q}_A$ and some properties of the ratio $I_1(z)/I_0(z)$ are given in Appendix 7.8.

M-step involves the maximization of $\mathbf{Q}_A$ with respect to parameters, yielding the following equations for updates:

$$\beta_{A,1} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}_{A,0}y, \tag{7.22}$$

$$\sigma^2_{A,1} = \frac{1}{2n}\cdot y^T\left(\mathbf{I}-\mathbf{W}_{A,0}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}_{A,0}\right)y, \tag{7.23}$$

where a weight matrix $\mathbf{W}_{A,0}$ under H_A is defined as

$$\mathbf{W}_{A,0} \triangleq \operatorname{diag}_{t=1}^{n}\{w_{A,0,t}\}. \tag{7.24}$$

Note that, except the weight matrix $\mathbf{W}_{A,0}$, (7.22) and (7.23) have very similar forms to the conventional LS solutions from Gaussian modeling in (7.15) and (7.16), which are easily implemented. Therefore, by simply weighting an observed magnitude time course, we can easily update parameter estimates in each step. From the M-step, we also produce an equation for update of θ, which is not informative, because

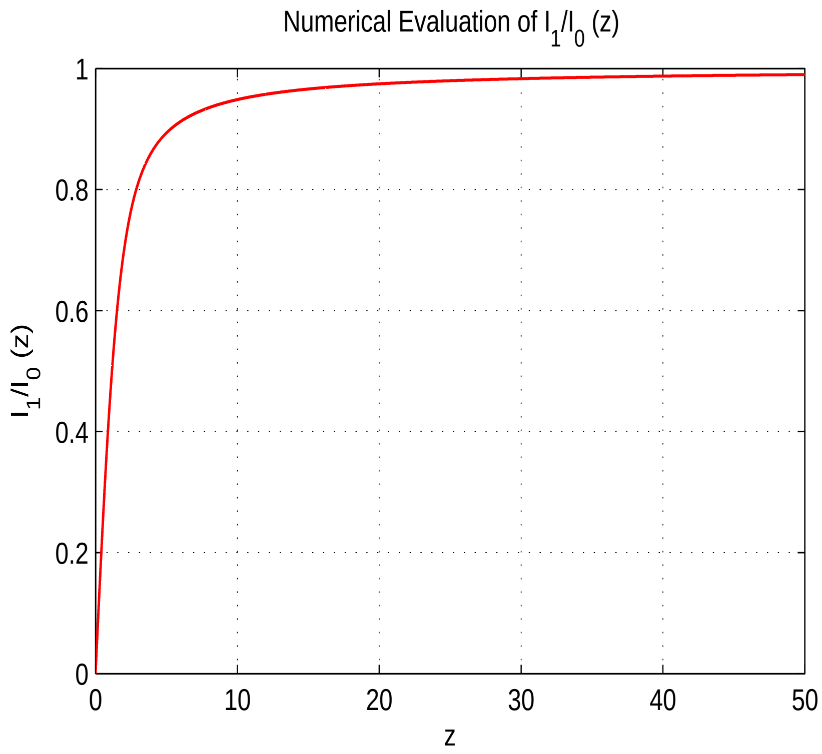


Figure 7.1: Numerical evaluation of $\frac{I_1(z)}{I_0(z)}$

$\theta_{A,1} = \theta_{A,0}$. Thus, we can not extract any information about the phase imperfection from a magnitude voxel time course at the voxel we are interested in. The detailed derivations of (7.22) and (7.23) from $\mathbf{Q}_A$ are provided in Appendix 7.9.

Since the ratio of $I_1(z)$ to $I_0(z)$ is an increasing function for z and is upper-bounded by 1, if SNR($\triangleq x_t^T\beta/\sigma$) is sufficiently high for all t, then the weight matrix $\mathbf{W}_{A,0}$ is close to an identity matrix $\mathbf{I}$ after the algorithm converges, resulting in that the iterative equation in (7.22) reduces to $\widetilde{\beta}_A$ in (7.15). In addition, when SNR is high for all t, it can be shown that the MLE of σ^2 based on the Rician PDF in (7.9) also boils down to $\widetilde{\sigma}_A^2$ in (7.16) by using a well known approximation of $I_0(z)$. The rigorous proof of this will be shown in Appendix 7.10. Conversely speaking, after the iterations converge, $\mathbf{W}_{A,0}$ can be used to measure how well conventional Gaussian modeling approximates Rician modeling. Details of this idea are discussed after deriving updates for parameters through an EM algorithm under H_N.

EM Algorithm under H_N

In order to obtain estimates of β and σ^2 under H_N, we follow the same E-step and M-step as above with a Lagrange multiplier based on the linear constraint imposed by H_N. We arrive at the following equations for updates:

$$\beta_{N,1} = \mathbf{\Phi} \cdot \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}_{N,0}y, \tag{7.25}$$

$$\sigma^2_{N,1} = \frac{1}{2n} \cdot y^T \left(\mathbf{I} - \mathbf{W}_{N,0}\mathbf{X} \cdot \mathbf{\Phi} \cdot \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}_{N,0}\right) y, \tag{7.26}$$

where a weight matrix $\mathbf{W}_{N,0}$ under H_N is defined similarly to (7.24) with $\beta_{N,0}$ and $\sigma^2_{N,0}$,

$$\mathbf{W}_{N,0} \triangleq \operatorname{diag}_{t=1}^{n}\{w_{N,0,t}\}. \tag{7.27}$$

Again, except the weight matrix $\mathbf{W}_{N,0}$, it can be easily shown that (7.25) and (7.26) have the same forms as $\widetilde{\beta}_N$ and $\widetilde{\sigma}^2_N$ in (7.17) and (7.18), respectively. From M-step under H_N, an iteration for θ is obtained as well, that is $\theta_{N,1} = \theta_{N,0}$, which is not useful. Details of derivation of a surrogate function in E-step under H_N, $\mathbf{Q}_N$ will be given in Appendix 7.8. Derivations of (7.25) and (7.26) from $\mathbf{Q}_N$ are provided in Appendix 7.9.

Measures of GOA : Weight Maps

From the discussion about weight matrix $\mathbf{W}_{A,0}$ and SNR under H_A, we define two useful sample quantities, which are sample mean and sample variance of converged weights,

$$\bar{w} \triangleq \frac{1}{n}\operatorname{tr}(\mathbf{W}_A), \qquad S^2_w \triangleq \frac{1}{n-1}\operatorname{tr}\left((\mathbf{W}_A - \bar{w}\mathbf{I})^2\right), \tag{7.28}$$

where $\mathbf{W}_A$ denotes a weight matrix after convergence of the algorithm and $\operatorname{tr}(\mathbf{W}_A)$ means its trace. For a given magnitude time series from a voxel position we are interested in, if SNR is sufficiently high to make the Gaussian approximation of a Rician PDF work well at all time points, then $\bar{w}$ is close to 1 with a small S_w. Conversely, if $\bar{w}$ is close to 1 and S_w is sufficiently small, it can be argued that the SNR at the voxel we are interested in is high enough to apply the Gaussian approximation to a Rician PDF. However, if $\bar{w}$ is less than 1 or S_w does not have a sufficiently small value, then the Gaussian modeling is not recommended. Therefore, spatial plots of $\bar{w}$ and S_w can be used as measures of goodness of approximation (GOA) to Rician modeling by Gaussian modeling at each voxel in a ROI. We call these spatial plots of $\bar{w}$ and S_w weight maps. Thus, using weight maps, voxels whose SNRs are not high enough can be identified and Rician modeling is recommended for those voxels.

7.4 Detection of Activations

One significant task in FMRI data analysis is to detect functional activations related to a pre-specified stimulus in a ROI. Based on the proposed Rician-EM approach, in this section, we build up an activation map using a LRT. An activation map is a just spatial plot of a detection statistic for activation. First, we review an existing detection statistic based on Gaussian modeling. Then, a detection statistic derived from Rician modeling is provided.

7.4.1 Detection Statistic from Gaussian Modeling

For independently and identically distributed Gaussian samples, it is well known that a LRT based on the hypotheses in (7.10) reduces to,

$$\lambda_G \triangleq 2\log\Lambda_G = n\log\left(\frac{\widetilde{\sigma}_N^2}{\widetilde{\sigma}_A^2}\right), \tag{7.29}$$

where Λ_G is the ratio of likelihood functions, one of which is maximized under H_A and the other is maximized under H_N [49]. Two MLEs, $\widetilde{\sigma}_A^2$ and $\widetilde{\sigma}_N^2$ are given in (7.16) and (7.18).

7.4.2 Detection Statistic from Rician-EM

With parameter estimates from the proposed Rician-EM approach, we construct a LRT based on the hypotheses in (7.10), leading to

$$\begin{aligned}\lambda_R \triangleq\ 2\log\Lambda_R = 2n\log\left(\frac{\widehat{\sigma}_N^2}{\widehat{\sigma}_A^2}\right) &+ \sum_{t=1}^{n}\frac{y_t^2 + (x_t^T\widehat{\beta}_N)^2}{\widehat{\sigma}_N^2} \\ &- \sum_{t=1}^{n}\left\{\frac{y_t^2 + (x_t^T\widehat{\beta}_A)^2}{\widehat{\sigma}_A^2} - 2\log\left(\frac{I_0(y_t X_t^T\widehat{\beta}_A/\widehat{\sigma}_A^2)}{I_0(y_t X_t^T\widehat{\beta}_N/\widehat{\sigma}_N^2)}\right)\right\},\end{aligned} \tag{7.30}$$

where, e.g., $\widehat{\beta}_A$ and $\widehat{\beta}_N$ are MLEs obtained from iterative equations in (7.22) and (7.25) after EM algorithm converges, respectively. Λ_R denotes the ratio of likelihood functions based on Rician distributions under H_A and H_N.

7.4.3 Family-Wise Error Rate Control

For a given significance level α_R, a threshold can be determined by asymptotics of λ_R. To be specific, when H_N is true (i.e., the tested voxel is not activated by the given stimuli), λ_R asymptotically obeys χ_r^2, a chi-square distribution with r degrees of freedom, where r is the full row rank of the constraint matrix $\mathbf{C}$ [63]. For the example in (7.11), we have $r = 2$. Since there are lots of voxels in a ROI, controlling overall rate is a multiple comparison problem (MCP). A measure to deal with the MCP is family-wise error (FWE) rate whose definition and an equivalent expression are given by, under H_N,

$$\text{FWE} \triangleq \Pr\left(\bigcup_{v=0}^{M-1} \{\lambda_R(v) > \gamma_R(\alpha_R)\} \;\middle|\; H_N \right) = \Pr\left(\max_v \lambda_R(v) > \gamma_R(\alpha_R) \;\middle|\; H_N \right), \tag{7.31}$$

where $\lambda_R(v)$ denotes the proposed detection statistic from the Rician-EM approach at the voxel location v, $\gamma_R(\alpha_R)$ is a threshold for a given α_R, and M denotes the number of voxels in a ROI. Due to the assumption of spatial independence, for a significance level α_R, we analytically determine a threshold,

$$\gamma_R(\alpha_R) = \Psi_r^{-1}(\sqrt[M]{1-\alpha_R}), \tag{7.32}$$

where $\Psi_r(t)$ is the cumulative density function (CDF) of χ_r^2. In FMRI, the significance level α_R is typically set to 0.05. Since the detection statistic from Gaussian modeling in (7.29) has the same asymptotic behavior for large n, $\gamma_R(\alpha_R)$ can be also applied to λ_G to control FWE rate.

7.5 Asymptotic Power Analysis of Rician Detection

The asymptotic power of the Rician detection statistic proposed in the previous section is derived, which requires an asymptotic distribution of LRT under alternative hypothesis H_A. The derived power function indicates which parameters have what impacts on the asymptotic power, thus providing a nice criterion to compare two competing detection statistics built up from a Rician distributed model. For Rician detection, we are especially interested in some interesting features of the effects caused by nuisance parameters, e.g., baseline m and noise variance σ^2, as we did analysis for the proposed space-time separability test in Chapter 5. We have discussed an asymptotic expansion of LRT under alternative hypothesis when serial correlation and nuisance parameters exist in section 5.4.1. Therefore, in this section, we apply that result to Rician likelihood function to obtain the asymptotic power function.

7.5.1 Asymptotic Expansion under Local Alternatives

For simplicity, we assume that there is no temporal linear drift in a given voxel time course, thus $b = 0$ in (7.2) and the constraint matrix $\mathbf{C}$ in (7.10) has a form like (7.11). Since motion artifacts can be corrected to some extents, the assumption on drift might be reasonable. Thus, a parameter of interest is f corresponding to activation amplitudes and nuisance parameters are m and σ^2 corresponding to baseline and noise variance. From (5.11), for a sequence of local alternatives defined as $f = f_0 + \varepsilon/\sqrt{n}$, we have the following asymptotic distribution of λ_R under H_A:

$$\lambda_R \sim \chi^2_{L,\Delta}, \tag{7.33}$$

where $\chi^2_{L,\Delta}$ denotes a non-central chi-square distribution with L degrees of freedom and Δ non-centrality parameter. The degree of freedom is the same as the dimension of f. From (5.12), the expression for the non-centrality parameter is given by

$$\Delta \;=\; \varepsilon^T \left(\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21}\right)\Big|_{f=f_0} \varepsilon, \tag{7.34}$$

where, since we are interested in testing whether the given voxel is activated by stimuli, we have $f_0 = 0$ and $\varepsilon = \sqrt{n}f$. The Fisher information matrix (FIM) is partitioned according to the partition of parameter space, namely,

$$\mathcal{I}(f, m, \sigma^2) \;\triangleq\; \begin{bmatrix} \mathcal{I}_{11}(f) & \mathcal{I}_{12}(f, m, \sigma^2) \\ \mathcal{I}_{21}(m, \sigma^2, f) & \mathcal{I}_{22}(m, \sigma^2) \end{bmatrix}, \tag{7.35}$$

where $\mathcal{I}_{11}(f)$ is associated with the parameter of interest and $\mathcal{I}_{22}(m, \sigma^2)$ involves nuisance parameters. On the contrary to the FIM corresponding to space-time separability test in Chapter 5, it turns out that $\mathcal{I}_{12}(f, m, \sigma^2)$ is a non-zero matrix, indicating that the asymptotic power function for Rician detection depends on the interesting and nuisance parameters. For details of an asymptotic expansion of LRT in the presence of nuisance parameters, the reader is referred to [28, 58].

7.5.2 Asymptotic Power Function of Rician Detection

We consider the FIM to compute the non-centrality parameter Δ in (7.34). Assuming that there is no temporal linear drift, we have the following negative Rician log-likelihood function:

$$-\log \mathscr{L} \triangleq n \log \sigma^2 - \sum_{t=1}^{n} \log y_t + \sum_{t=1}^{n} \left\{ \frac{y_t^2 + (m + \xi_t^T f)^2}{2\sigma^2} - \log I_0 \left(\frac{y_t(m + \xi_t^T f)}{\sigma^2} \right) \right\}. \tag{7.36}$$

From this negative Rician log-likelihood function, by computing the expectation of the second order derivatives, one arrives at the FIM evaluated at $f = 0$ that is given by

$$\mathcal{I}(0, m, \sigma^2) = \begin{bmatrix} \frac{(\sigma^2 - A^0)}{n\sigma^4} \cdot \sum_{t=1}^{n} \xi_t \ \xi_t^T & \frac{\sigma^2 - A^0}{n\sigma^4} \cdot \sum_{t=1}^{n} \xi_t & \frac{(mA^0 + \sigma^2 B^0 - m\sigma^2)}{n\sigma^6} \cdot \sum_{t=1}^{n} \xi_t \\ \frac{\sigma^2 - A^0}{n\sigma^4} \cdot \sum_{t=1}^{n} \xi_t & \frac{\sigma^2 - A^0}{\sigma^4} & \frac{(mA^0 + \sigma^2 B^0 - m\sigma^2)}{\sigma^6} \\ \frac{(mA^0 + \sigma^2 B^0 - m\sigma^2)}{n\sigma^6} \cdot \sum_{t=1}^{n} \xi_t & \frac{(mA^0 + \sigma^2 B^0 - m\sigma^2)}{\sigma^6} & \frac{\sigma^4 - m^2 A^0 - 2m\sigma^2 (B^0 - m)}{\sigma^8} \end{bmatrix}, \tag{7.37}$$

where the matrix in the top and left corner is associated with $\mathcal{I}_{11}(0)$ and has the size of $L \times L$. The entry at the center corresponds to baseline m and the entry at the bottom and right corner corresponds to noise variance σ^2. $A^0(m, \sigma^2)$ and $B^0(m, \sigma^2)$ are defined as

$$A^0(m, \sigma^2) \triangleq \mathbf{E} \left[\frac{I_1' \left(\frac{y_t (m + \xi_t^T f)}{\sigma^2} \right) I_0 \left(\frac{y_t (m + \xi_t^T f)}{\sigma^2} \right) - I_1^2 \left(\frac{y_t (m + \xi_t^T f)}{\sigma^2} \right)}{I_0^2 \left(\frac{y_t (m + \xi_t^T f)}{\sigma^2} \right)} \ y_t^2 \right] \Bigg|_{f=0}, \tag{7.38}$$

$$B^0(m, \sigma^2) \triangleq \mathbf{E} \left[\frac{I_1 \left(\frac{y_t (m + \xi_t^T f)}{\sigma^2} \right)}{I_0 \left(\frac{y_t (m + \xi_t^T f)}{\sigma^2} \right)} \ y_t \right] \Bigg|_{f=0}, \tag{7.39}$$

where $I_1'(z)$ denotes the first order derivative of $I_1(z)$ with respect to z. We define a useful function $\bar{A}^0(\bar{m})$ whose definition is given by

$$\bar{A}^0(\bar{m}) \triangleq \frac{A^0(m, \sigma^2)}{\sigma^2}, \qquad \bar{m} \triangleq \frac{m}{\sigma}, \tag{7.40}$$

where we call $\bar{m}$ baseline to noise ratio (BNR). Note that $\bar{A}^0(\bar{m})$ involves only BNR, which is easily shown by noting that

$$\begin{aligned} \bar{A}^0(\bar{m}) &= \int_0^\infty \left(I_1'(x\bar{m}) - \frac{I_1^2(x\bar{m})}{I_0(x\bar{m})} \right) \cdot x^3 exp\left(-\frac{x^2 + \bar{m}^2}{2} \right) \mathrm{d}x, \qquad (7.41) \\ &= exp\left(-\frac{\bar{m}^2}{2} \right) \cdot \int_0^\infty \left(\frac{I_2(x\bar{m}) + I_0(x\bar{m})}{2} - \frac{I_1^2(x\bar{m})}{I_0(x\bar{m})} \right) x^3 exp\left(-\frac{x^2}{2} \right) \mathrm{d}x, \end{aligned}$$

where $x \triangleq y_t/\sigma$. Therefore, we conclude that $\bar{A}^0(0) = 1$ and $\bar{A}^0(\bar{m})$ goes to 0 as $\bar{m}$ approaches to infinity. The numerical evaluation of (7.41) on Fig.7.2 shows that

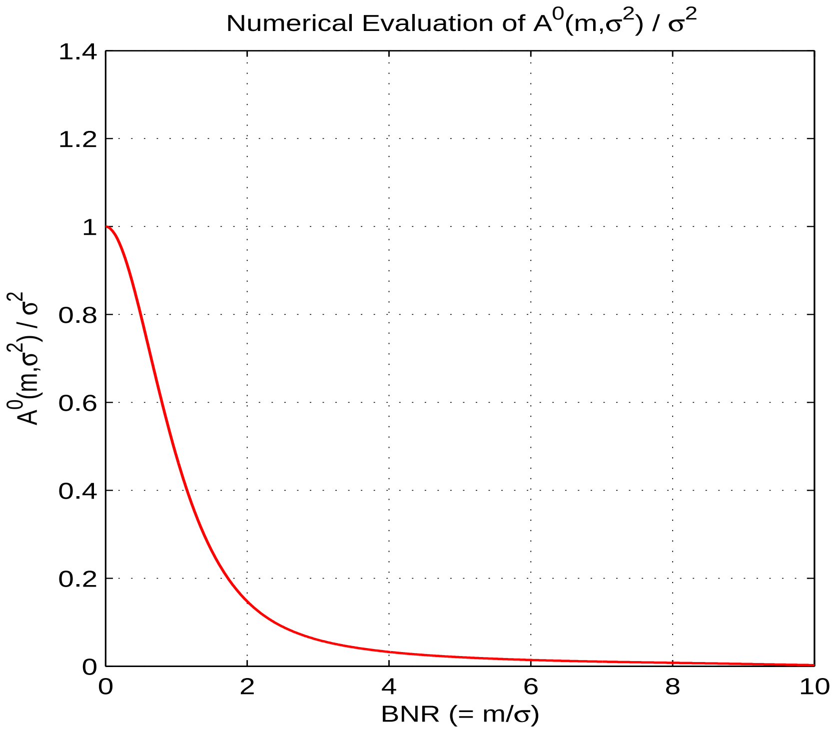


Figure 7.2: Numerical evaluation of $\bar{A}^0(\bar{m})$

$\bar{A}^0(\bar{m})$ is a monotonically decreasing function with respect to $\bar{m}$. Similarly, we define and can show that

$$\bar{B}^0(\bar{m}) \triangleq \frac{B^0(m,\sigma^2)}{\sigma} = \int_0^\infty I_1(x\bar{m}) \cdot x^2 exp\left(-\frac{x^2+\bar{m}^2}{2}\right) \mathrm{d}x. \tag{7.42}$$

It turns out that $\bar{A}^0(\bar{m})$ plays an important role in the expression for the non-centrality parameter but $\bar{B}^0(\bar{m})$ is canceled out as shown in (7.43). The derivation of (7.37) from (7.36) is provided in Appendix 7.11.

By plugging (7.37) into (7.34), we now have the non-centrality parameter given by

$$\Delta_R = n\left(1-\bar{A}^0(\bar{m})\right) \cdot \bar{f}^T \mathbf{\Sigma}_\xi \bar{f}, \tag{7.43}$$

where the covariance matrix and the activation to noise ratio (ANR) are defined as

$$\mathbf{\Sigma}_\xi \triangleq \frac{1}{n}\sum_{t=1}^{n} \xi_t \xi_t^T - \left(\frac{1}{n}\sum_{t=1}^{n}\xi_t\right)\left(\frac{1}{n}\sum_{t=1}^{n}\xi_t^T\right), \qquad \bar{f} \triangleq \frac{f}{\sigma}, \tag{7.44}$$

respectively. Therefore, (7.43) shows that the non-centrality parameter in the asymptotic expansion of λ_R under a sequence of local alternatives depends on not only parameter of interest f (activation amplitude) but also on nuisance parameters m (baseline) and σ^2 (noise variance). Remarkably, it depends on ANR($\triangleq f/\sigma$) and BNR($\triangleq m/\sigma$) in a multiplicative and separable way. In addition, (7.43) indicates that, for a fixed BNR, ANR contributes to the non-centrality parameter Δ_R in a quadratic way. For a high BNR, since $\bar{A}^0(\bar{m})$ is close to 0, the non-centrality parameter is given by

$$\Delta_R \;=\; n\bar{f}^T \mathbf{\Sigma}_\xi \bar{f}, \tag{7.45}$$

being independent of baseline m. This is well matched to conventional results on power analysis in classical linear regression [4].

We can compute the power, probability that the proposed statistic detects the activation when H_A is true. By combining (7.32) and (7.33), for a significance level α_R, the power function has a form of

$$\mathrm{P}_{Rician}\left(\bar{m}, \bar{f}\right) \;\triangleq\; \Pr\left(\lambda_R > \gamma(\alpha_R) \;\middle|\; H_A\right) \;=\; 1 - \Phi_{L,\Delta_R}\left(\Psi_L^{-1}(\sqrt[M]{1-\alpha_R})\right), \tag{7.46}$$

where $\Phi_{L,\Delta_R}(t)$ denotes the CDF of a non-central chi-square distribution with L degrees of freedom and Δ_R non-centrality parameter, and M is the number of voxels in a ROI. By the numerical evaluation of (7.46), an example of the theoretical asymptotic power function, $\mathrm{P}_{Rician}\left(\bar{m}, \bar{f}\right)$ is given on Fig.7.3 for $L = 1$, $\alpha_R = 0.05$, $M = 100$, $n = 100$, and $\Sigma_\xi = 1$.

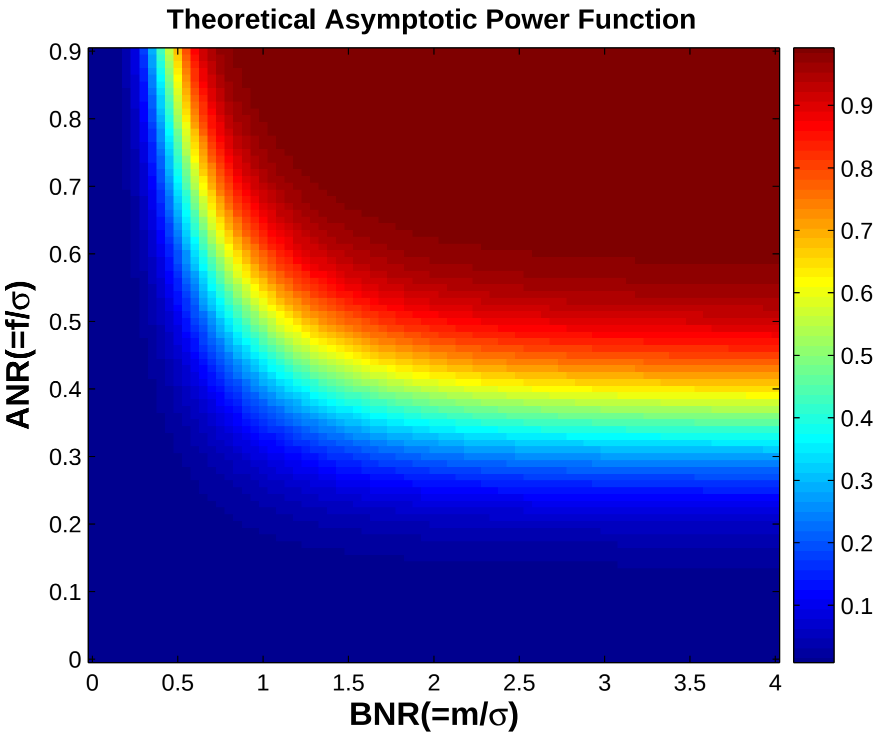


Figure 7.3: An example of the theoretical asymptotic power function $\mathrm{P}_{Rician}\left(\bar{m}, \bar{f}\right)$ for $L = 1$, $\alpha_R = 0.05$, $M = 100$, $n = 100$, and $\Sigma_\xi = 1$

7.6 Simulation Study

7.6.1 Discussions on SNR

Before we start simulations, we need to mention an important point to make discussions clear. Recall that, according to the conventional definition of SNR($\triangleq x_t^T\beta/\sigma$), SNR varies along time [47]. Since SNR is time-variant, it is not appropriate to summarize differences between multivariate Rician PDF and an approximation by multivariate Gaussian PDF for a given voxel time course. For example, we may have a voxel time course in which a SNR at one time point is not sufficiently high whereas a SNR at another time point is high enough to apply the Gaussian approximation. Therefore, in simulations, we use two newly defined terminologies in this chapter, activation to noise ratio (ANR) and baseline to noise ratio (BNR), which are constants along time, instead of time-variant SNR. In addition, as shown in (7.46), these two quantities, i.e., $\bar{f}$ and $\bar{m}$, are natural choices from the discussions on asymptotic power analysis for Rician detection.

Using the newly defined ANR and BNR, we perform simulations to compare the developed Rician-EM approach and the existing Gaussian modeling. We investigate unbiasedness and mean squared error (MSE) of parameter estimates from each modeling for various BNRs. In addition, we check the usefulness of newly developed two weight maps, spatial plots of $\bar{w}$ and S_w, for various BNRs as well. Data based on the complex signal and noise model in (7.4) are synthesized for simulations. The number of time points, n is set to 100 and a parameter for phase imperfection is assigned to $\theta = \pi/4$. The noise variance is set to $\sigma^2 = 1$ for simplicity and the simplest parametric modeling of the BOLD response in (1.7) is used.

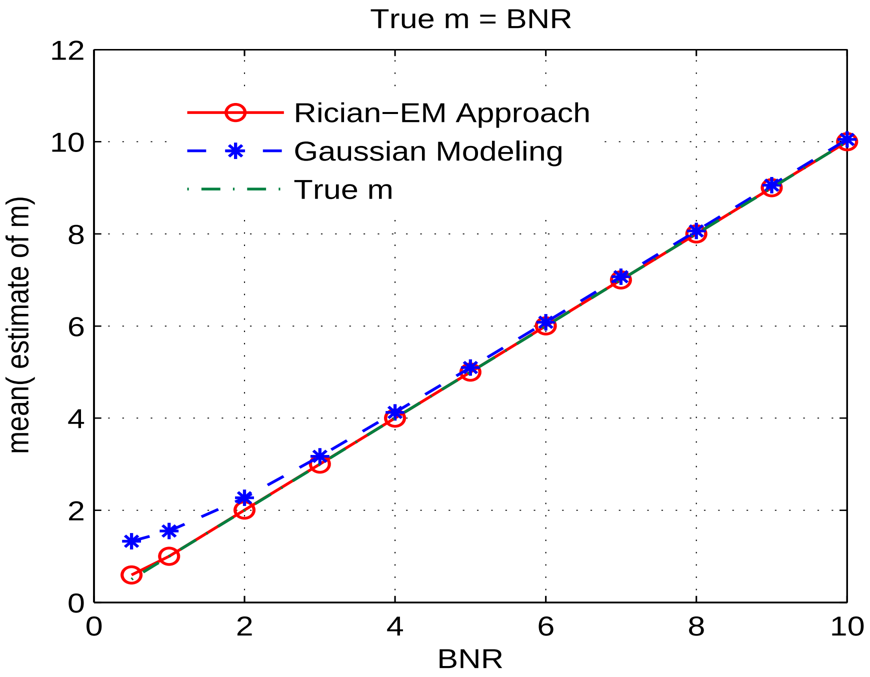
7.6.2 Parameter Estimation

We fix an activation amplitude at $f = 0.5$, thus ANR is fixed and examine the bias and MSEs of $\widehat{m}$, $\widehat{f}$, and $\widehat{\sigma}^2$ for various BNRs. As shown in the two figures on Fig.7.4 and the left figure on Fig.7.5, the Rician-EM approach provides less biased parameter estimates than Gaussian modeling does for all regions of BNRs. According to the right figure on Fig.7.5, when BNR is very low, e.g., when it is less than 2, the bias caused by the Gaussian modeling becomes severe and Gaussian modeling as an approximation to Rician modeling is not recommended. The tolerance of bias or MSE can be determined by researchers conducting FMRI experiments. Thus, by looking at spatial plots of $\bar{w}$ and S_w, voxels whose ANRs and BNRs are not sufficiently high for the Gaussian modeling can be identified. On Fig.7.6, MSEs of parameter estimates are shown for varying BNRs. For all values of BNRs, the

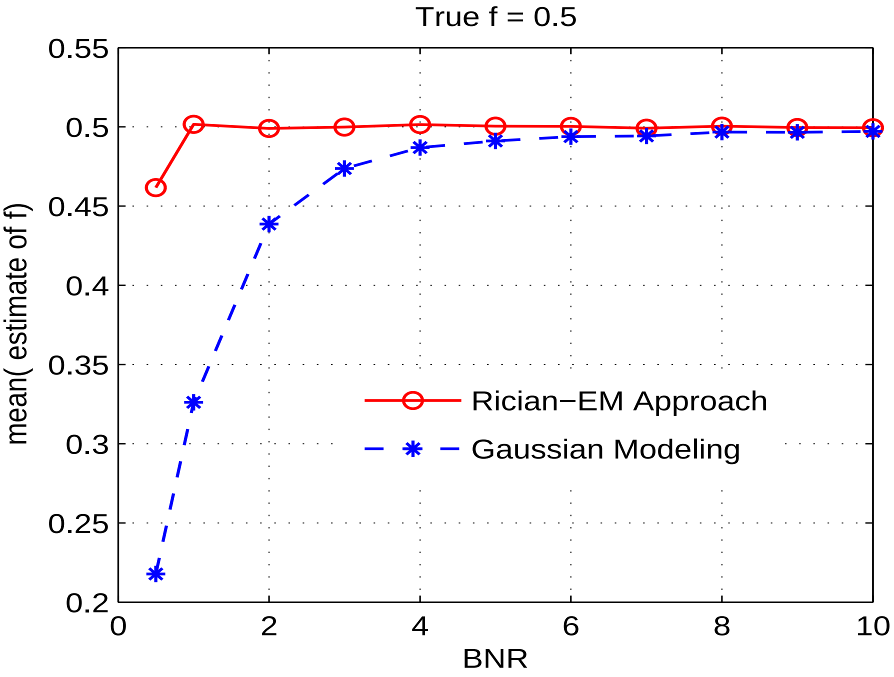
Rician-EM method provides smaller MSEs than the conventional Gaussian modeling. Especially, for very low BNR which is less than 2, Rician-EM is much better than Gaussian modeling in terms of MSE. For different values of activation amplitude ($f = 0.1$ and $f = 0.3$), although results are not shown here, simulations show the same behaviors of biases and MSEs for varying BNRs as shown in Fig.7.4-7.6.

7.7 Conclusions

We developed a method for parameter estimation from a Rician distributed model via EM algorithm, which allowed very simple forms of iterations, more reliability, and better interpretation compared to existing Rician modeling. The developed Rician-EM approach can be easily modified from conventional Gaussian modeling due to their similar forms. Using estimated parameters, we built up a detection statistic for activation by LRT. In addition, we analyzed the asymptotic power of the proposed activation statistic. According to simulations, the Rician-EM provided parameter estimates which were less biased and had smaller MSEs than the conventional Gaussian modeling did for all ranges of BNRs.

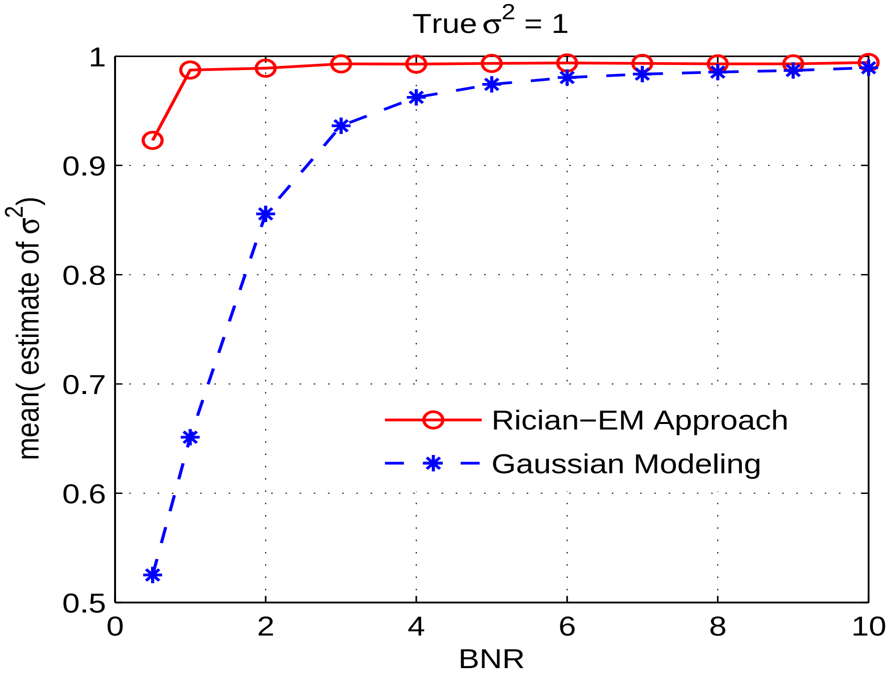


(a) Mean of estimates of baseline, $\widehat{m}$

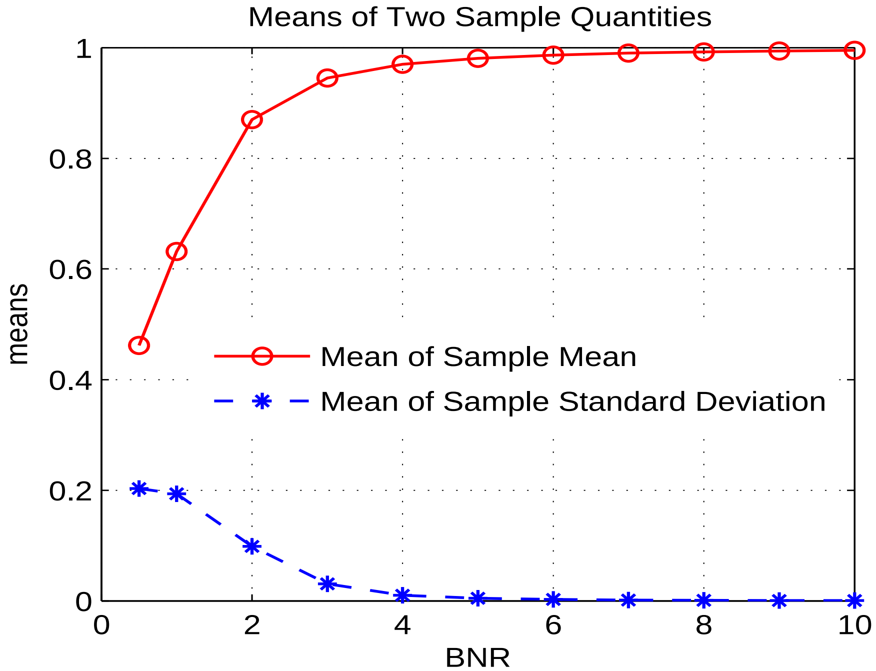


(b) Mean of estimates of activation amplitude, $\widehat{f}$

Figure 7.4: Means of estimates of baseline m and activation amplitude f for several BNRs. True values of activation amplitude and noise variance are set to $f = 0.5$ and $\sigma^2 = 1$, respectively.

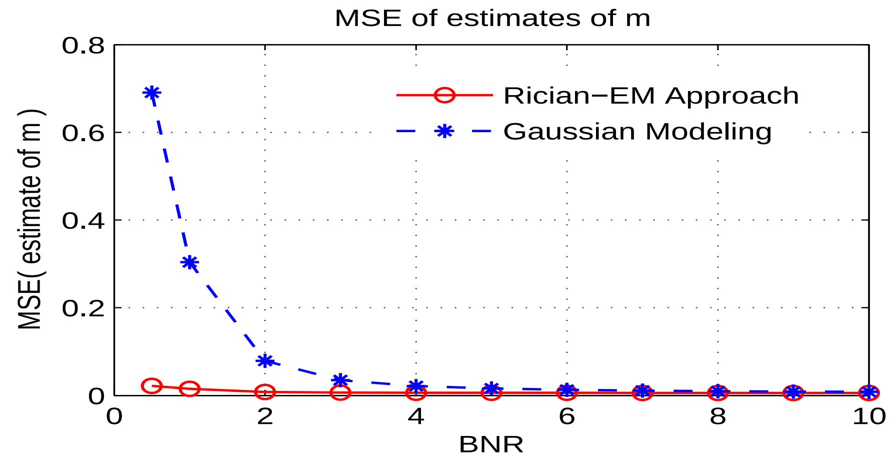


(a) Mean of estimates of noise variance, $\widehat{\sigma}^2$

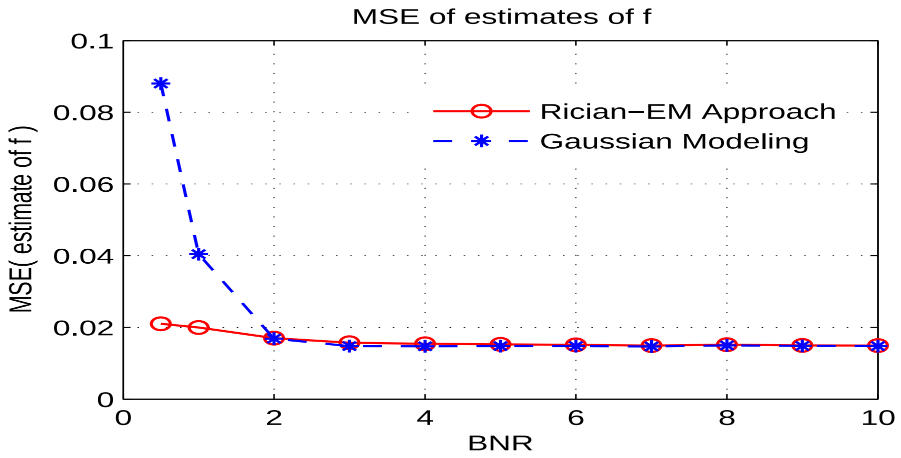


(b) Means of sample mean, $\bar{w}$ and sample standard deviation, S_w

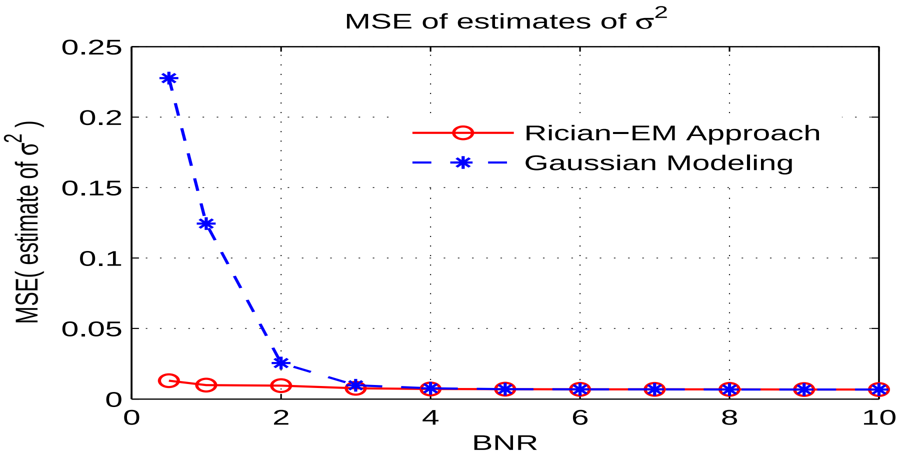
Figure 7.5: Means of estimates of noise variance σ^2 and two sample quantities ($\bar{w}$ and S_w) for several BNRs. True values of activation amplitude and noise variance are set to $f = 0.5$ and $\sigma^2 = 1$, respectively.



(a) MSE of estimates of baseline, $\widehat{m}$



(b) MSE of estimates of activation amplitude, $\widehat{f}$



(c) MSE of estimates of noise variance, $\widehat{\sigma}^2$

Figure 7.6: MSEs of estimates of baseline m, activation amplitude f, and noise variance σ^2 for several BNRs. True values of activation amplitude and noise variance are set to $f = 0.5$ and $\sigma^2 = 1$, respectively.

7.8 Appendix I : Derivations of Surrogate Functions

In this section, we perform a E-step under each hypothesis, i.e., derive surrogate functions, $\mathbf{Q}_A$ in (7.20) under H_A and $\mathbf{Q}_N$ under H_N. Since $\mathbf{Q}_N$ is obtained by a simple modification of $\mathbf{Q}_A$ with a Lagrange multiplier, we mainly focus on developing $\mathbf{Q}_A$. Before the discussion of these derivations, we make brief mathematical reviews of two Bessel functions, I_0 and I_1, and their ratio which are necessary for our surrogate derivations.

Modified Bessel Functions of the First Kind

The zeroth order and first order modified Bessel function of the first kind have the following definitions :

$$I_0(z) = \int_{-\pi}^{\pi} e^{z\cos\theta}\frac{d\theta}{2\pi}, \qquad I_1(z) = \int_{-\pi}^{\pi} e^{z\cos\theta}\cos\theta\frac{d\theta}{2\pi}, \tag{7.47}$$

where note that $I_1(z)$ is the first order derivative of $I_0(z)$ with respect to z. Here, we define a function, the ratio of $I_1(z)$ to $I_0(z)$, which is useful to understand behaviors of the weight matrix $\mathbf{W}_{A,0}$ in (7.24),

$$r(z) = \frac{I_1(z)}{I_0(z)}. \tag{7.48}$$

According to a proof in [3], the first derivative of $r(z)$ with respect to z is positive, thus $r(z)$ is strictly increasing. It can be easily checked that $0 \leq r(z) < 1$ for all $z \geq 0$ and $r(z)$ is an odd function, that is $r(-z) = -r(z)$. Therefore, $r(z)$ has a nice asymptotic property, $r(z)$ approaches 1 as z goes positive infinity.

Derivation of $\mathbf{Q}_A$

In E-step, a surrogate function is defined by a conditional expectation of $\log \mathscr{L}_c$ in (7.19) for a given magnitude observation y as follows.

$$\mathbf{Q}_A = \mathbf{E}\left[\log \mathscr{L}_c \mid y; \beta_{A,0}, \theta_{A,0}, \sigma^2_{A,0}\right], \tag{7.49}$$

where $\beta_{A,0}$, $\theta_{A,0}$, and $\sigma^2_{A,0}$ are given parameter estimates in each E-step. By plugging (7.19) into (7.49), $\mathbf{Q}_A$ has a form of

$$\mathbf{Q}_A = -n\log(2\pi\sigma^2) + \sum_{t=1}^{n}\log y_t - \frac{1}{2\sigma^2}\left\{\sum_{t=1}^{n} y_t^2 + \left(x_t^T\beta\right)^2 \right.$$
$$\left. -2\sum_{t=1}^{n} y_t x_t^T\beta \cdot \mathbf{E}\left[\cos(\phi_t-\theta) \;\middle|\; y_t;\beta_{A,0},\theta_{A,0},\sigma^2_{A,0}\right]\right\}.$$

Thus, to obtain (7.20), we need to show the following identity for a time point t,

$$\mathbf{E}\left[\cos(\phi_t-\theta) \;\middle|\; y_t;\beta_{A,0},\theta_{A,0},\sigma^2_{A,0}\right] = \cos(\theta_{A,0}-\theta)\cdot w_{A,0,t}. \tag{7.50}$$

Here, recall that the joint PDF of y_t and ϕ_t has the following form, for $y_t > 0$,

$$p_c(y_t,\phi_t) = \frac{y_t}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left(y_t^2+\left(x_t^T\beta\right)^2-2y_t x_t^T\beta\cos(\phi_t-\theta)\right)}. \tag{7.51}$$

By dividing (7.51) with a marginal PDF of y_t in (7.8), one obtains a conditional PDF of ϕ_t for a given y_t,

$$p(\phi_t \mid y_t) = \frac{1}{2\pi}\cdot\frac{exp\left(\frac{y_t x_t^T\beta}{\sigma^2}\cdot\cos(\phi_t-\theta)\right)}{I_0\left(\frac{y_t x_t^T\beta}{\sigma^2}\right)}. \tag{7.52}$$

Then, we consider the following conditional expectation,

$$\mathbf{E}\left[\cos(\phi_t-\theta) \;\middle|\; y_t;\beta_{A,0},\theta_{A,0},\sigma^2_{A,0}\right] = \int_{-\pi}^{\pi}\cos(\phi_t-\theta)\cdot p(\phi_t \mid y_t;\beta_{A,0},\theta_{A,0},\sigma^2_{A,0})\, d\phi_t. \tag{7.53}$$

Here, recall the following trigonometric identity,

$$\cos(\phi_t-\theta) = \cos(\theta_{A,0}-\theta)\cos(\phi_t-\theta_{A,0}) - \sin(\theta_{A,0}-\theta)\sin(\phi_t-\theta_{A,0}). \tag{7.54}$$

After plugging (7.52) and (7.54) into (7.53), since the integration term associated with $\sin(\theta_{A,0}-\theta)$ vanishes, we obtain (7.50) by recognizing $I_1(z)$ evaluated at $z = y_t x_t^T\beta_{A,0}/\sigma^2_{A,0}$ and complete the derivation of $\mathbf{Q}_A$.

Derivation of $\mathbf{Q}_N$

In E-step, a surrogate function with a Lagrange multiplier based on the linear constraint under H_N is simply defined as follows.

$$\begin{aligned}\mathbf{Q}_N &= \mathbf{E}\left[\log \mathscr{L}_c \mid y; \beta_{N,0}, \theta_{N,0}, \sigma^2_{N,0}\right] + \lambda^T \mathbf{C}\beta, \\ &= \mathbf{Q}^*_A + \lambda^T \mathbf{C}\beta, \qquad (7.55)\end{aligned}$$

where $\mathbf{Q}^*_A$ is similarly defined as $\mathbf{Q}_A$ in (7.49) with parameter estimates under H_N, $\beta_{N,0}$, $\theta_{N,0}$, and $\sigma^2_{N,0}$.

7.9 Appendix II : Derivations of Iterative Equations

In this section, we perform a M-step under each hypothesis. First, we derive iterative equations under H_A by finding a maximizer of $\mathbf{Q}_A$ and then develop iterative equations under H_N by maximizing $\mathbf{Q}_N$. To do that, for compact expressions, we redefine $\mathbf{Q}_A$ in a matrix form as follows. Neglecting terms which are independent of parameters, we have

$$\mathbf{Q}_A = -n \log \sigma^2 - \frac{1}{2\sigma^2}\left[y^T y + \beta^T\left(\mathbf{X}^T\mathbf{X}\right)\beta - 2\cos(\theta_{A,0} - \theta)\cdot \beta^T \mathbf{X}^T \mathbf{W}_{A,0}\, y\right], \quad (7.56)$$

where a weight matrix $\mathbf{W}_{A,0}$ has a definition of

$$\mathbf{W}_{A,0} = \mathrm{diag}_{t=1}^{n}\{w_{A,0,t}\}.$$

Developing Iterations under H_A

Firstly, taking a derivative of the newly defined $\mathbf{Q}_A$ in (7.56) with respect to θ and zeroing it give

$$\frac{\partial \mathbf{Q}_A}{\partial \theta} = \beta^T \mathbf{X}^T \mathbf{W}_{A,0}\, y \cdot \frac{\sin(\theta_{A,0} - \theta)}{\sigma^2} = 0, \quad (7.57)$$

leading to $\sin(\theta_{A,0} - \theta) = 0$, thus $\theta_{A,1} = \theta_{A,0}$.

Secondly, taking a derivative of $\mathbf{Q}_A$ with respect to β and zeroing it allow

$$\frac{\partial \mathbf{Q}_A}{\partial \beta} = \frac{1}{\sigma^2}\left[\cos(\theta_{A,0} - \theta) \cdot \mathbf{X}^T \mathbf{W}_{A,0} y - \left(\mathbf{X}^T \mathbf{X}\right)\beta\right] = 0. \tag{7.58}$$

Rearranging terms and plugging $\theta_{A,1}$ into θ leads to (7.22). Note that we also obtain an identity which is useful for the derivation of $\sigma^2_{A,1}$ below, that is,

$$\beta^T \left(\mathbf{X}^T \mathbf{X}\right)\beta = \beta^T \mathbf{X}^T \mathbf{W}_{A,0} y. \tag{7.59}$$

Finally, by taking a derivative of $\mathbf{Q}_A$ with respect to σ^2 and setting it zero, one arrives at

$$\frac{\partial \mathbf{Q}_A}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\left[y^T y + \beta^T \left(\mathbf{X}^T \mathbf{X}\right)\beta - 2\cos(\theta_{A,0} - \theta) \cdot \beta^T \mathbf{X}^T \mathbf{W}_{A,0} y\right] = 0. \tag{7.60}$$

Rearranging terms for σ^2 and plugging $\theta_{A,1}$ into θ, then (7.59) allows

$$\sigma^2_{A,1} = \frac{1}{2n}\left(y^T y - \beta^T \mathbf{X}^T \mathbf{W}_{A,0} y\right), \tag{7.61}$$

replacing β with $\beta_{A,1}$ in (7.22) gives the desired result in (7.23). Then, we obtain the necessary iterations under H_A.

Developing Iterations under H_N

We start the derivations from $\mathbf{Q}_N$ in (7.55). Firstly, since the Lagrange multiplier term in $\mathbf{Q}_N$ does not depend on θ, taking a derivative of $\mathbf{Q}_N$ with respect to θ and zeroing it give the same iterative equation as that under H_A, $\theta_{N,1} = \theta_{N,0}$.

Secondly, taking a derivative of $\mathbf{Q}_N$ with respect to β and plugging $\theta_{N,1}$ into θ lead to

$$\frac{\partial \mathbf{Q}_N}{\partial \beta} = \frac{1}{\sigma^2}\left[\mathbf{X}^T \mathbf{W}_{N,0} y - \left(\mathbf{X}^T \mathbf{X}\right)\beta\right] + \mathbf{C}^T \lambda = 0. \tag{7.62}$$

Arranging terms for β gives

$$\beta_{N,1} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\left(\sigma^2\mathbf{C}^T\lambda + \mathbf{X}^T\mathbf{W}_{N,0}y\right), \tag{7.63}$$

which should satisfy the constraint, $\mathbf{C}\beta = 0$. Applying the constraint gives an expression for λ, which is

$$\lambda = -\frac{1}{\sigma^2}\Big[\mathbf{C}\big(\mathbf{X}^T\mathbf{X}\big)^{-1}\mathbf{C}^T\Big]^{-1}\mathbf{C}\big(\mathbf{X}^T\mathbf{X}\big)^{-1}\mathbf{X}^T\mathbf{W}_{N,0}y. \tag{7.64}$$

Plugging the expression for λ into (7.63) allows the desired form of $\beta_{N,1}$ in (7.25). Again, as under H_A, we obtain an identity which is useful for the derivation of $\sigma^2_{N,1}$, that is,

$$\beta^T\left(\mathbf{X}^T\mathbf{X}\right)\beta = \left(\mathbf{X}^T\mathbf{W}_{N,0}y\right)^T\beta, \tag{7.65}$$

where because $\mathbf{C}\beta = 0$ implies $\mathbf{\Phi}\beta = \beta$.

Finally, since the Lagrange multiplier term in $\mathbf{Q}_N$ does not depend on σ^2, taking a derivative of $\mathbf{Q}_N$ with respect to σ^2 and zeroing it give the same expression as that under H_A, thus we have

$$\frac{\partial\mathbf{Q}_N}{\partial\sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\left[y^Ty + \beta^T\left(\mathbf{X}^T\mathbf{X}\right)\beta - 2\cos(\theta_{N,0}-\theta)\cdot\beta^T\mathbf{X}^T\mathbf{W}_{N,0}y\right] = 0. \tag{7.66}$$

Rearranging terms for σ^2 and plugging $\theta_{N,1}$ into θ, then (7.65) allows

$$\sigma^2_{N,1} = \frac{1}{2n}\left(y^Ty - \beta^T\mathbf{X}^T\mathbf{W}_{N,0}y\right), \tag{7.67}$$

replacing β with $\beta_{N,1}$ allows the desired iteration in (7.26). Then, we complete the derivations of necessary iterations under H_A and H_N.

7.10 Appendix III

In EM algorithm, it is guaranteed that $\sigma^2_{A,1}$ converges to the MLE of σ^2_A, say $\widehat{\sigma}^2_A$. Thus, we need to show that the MLE from Rician PDF, $\widehat{\sigma}^2_A$ has the same form as the

MLE from Gaussian PDF, $\widetilde{\sigma}_A^2$ when $x_t^T\beta/\sigma$ is high. From (7.9), we start the proof. Taking a derivative of (7.9) with respect to σ^2 gives

$$\frac{\partial \log \mathscr{L}}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\sigma^4}\left\{\sum_{t=1}^{n} y_t^2 + (x_t^T\beta)^2 - w_{A,t}\cdot 2y_t x_t^T\beta\right\}, \tag{7.68}$$

where a weight is defined as

$$w_{A,t} = \frac{I_1(y_t x_t^T\beta/\sigma^2)}{I_0(y_t x_t^T\beta/\sigma^2)}.$$

Then, by zeroing (7.68) and rearranging terms, we obtain

$$2n\sigma^2 = \sum_{t=1}^{n} y_t^2 + \left(x_t^T\beta\right)^2 - w_{A,t}\cdot 2y_t x_t^T\beta. \tag{7.69}$$

Here, when z is high, recall that the following approximation of $I_0(z)$ by [1, Sec.9.6],

$$I_0(z) = \frac{e^z}{\sqrt{2\pi z}}, \tag{7.70}$$

which gives an approximation of $I_1(z)$ by taking the first derivative with respect to z. Then, one arrives at

$$r(z) = \frac{I_1(z)}{I_0(z)} = \frac{2z-1}{2z},$$

leading to an approximation of $w_{A,t}$,

$$w_{A,t} = \frac{2y_t x_t^T\beta - \sigma^2}{2y_t x_t^T\beta}. \tag{7.71}$$

Plugging (7.71) into (7.69) allows

$$2n\sigma^2 = \left(\sum_{t=1}^{n} y_t^2 + \left(x_t^T\beta\right)^2 - 2y_t x_t^T\beta\right) + n\sigma^2, \tag{7.72}$$

reducing to, in a matrix form,

$$\widehat{\sigma}^2 = \frac{1}{n}(y-\mathbf{X}\beta)^T(y-\mathbf{X}\beta). \tag{7.73}$$

By replacing β with the MLE of β, $\widehat{\beta}$, we have an expression for $\widehat{\sigma}^2$ which has the same form as $\widetilde{\sigma}_A^2$, completing the proof.

7.11 Appendix IV : Derivation of the FIM

In this section, we derive the FIM given in (7.37) from the negative Rician log-likelihood function in (7.36). We consider second order derivatives of $-\log \mathscr{L}$ with respect to f_i (the i-th entry of f), m, and σ^2, which are given as follows:

$$\frac{\partial^2(-\log \mathscr{L})}{\partial m^2} = \sum_{t=1}^{n} \frac{1}{\sigma^2} - \frac{I_1^{'}(z_t)I_0(z_t) - I_1^2(z_t)}{I_0^2(z_t)} \left(\frac{y_t}{\sigma^2}\right)^2, \tag{7.74}$$

$$\frac{\partial^2(-\log \mathscr{L})}{\partial f_i \partial f_j} = \sum_{t=1}^{n} \frac{\xi_{i,t}\xi_{j,t}}{\sigma^2} - \frac{I_1^{'}(z_t)I_0(z_t) - I_1^2(z_t)}{I_0^2(z_t)} \left(\frac{y_t}{\sigma^2}\right)^2 \xi_{i,t}\xi_{j,t} \tag{7.75}$$

$$\frac{\partial^2(-\log \mathscr{L})}{\partial m \partial f_i} = \sum_{t=1}^{n} \frac{\xi_{i,t}}{\sigma^2} - \frac{I_1^{'}(z_t)I_0(z_t) - I_1^2(z_t)}{I_0^2(z_t)} \left(\frac{y_t}{\sigma^2}\right)^2 \xi_{i,t}, \tag{7.76}$$

$$\frac{\partial^2(-\log \mathscr{L})}{\partial m \partial \sigma^2} = \sum_{t=1}^{n} -\frac{m + \xi_t^T f}{\sigma^4} + \frac{I_1^{'}(z_t)I_0(z_t) - I_1^2(z_t)}{I_0^2(z_t)} \left(\frac{y_t^2(m + \xi_t^T f)}{\sigma^6}\right) + \frac{I_1(z_t)}{I_0(z_t)} \left(\frac{y_t}{\sigma^4}\right), \tag{7.77}$$

$$\begin{aligned} \frac{\partial^2(-\log \mathscr{L})}{\partial f_i \partial \sigma^2} &= \sum_{t=1}^{n} -\frac{m + \xi_t^T f}{\sigma^4}\xi_{i,t} + \frac{I_1^{'}(z_t)I_0(z_t) - I_1^2(z_t)}{I_0^2(z_t)} \left(\frac{y_t^2(m + \xi_t^T f)}{\sigma^6}\xi_{i,t}\right) \\ &+ \frac{I_1(z_t)}{I_0(z_t)} \left(\frac{y_t}{\sigma^4}\right) \xi_{i,t}, \end{aligned} \tag{7.78}$$

$$\begin{aligned} \frac{\partial^2(-\log \mathscr{L})}{\partial (\sigma^2)^2} &= \sum_{t=1}^{n} -\frac{1}{\sigma^4} + \frac{y_t^2 + (m + \xi_t^T f)^2}{\sigma^6} - \frac{I_1^{'}(z_t)I_0(z_t) - I_1^2(z_t)}{I_0^2(z_t)} \left(\frac{y_t^2(m + \xi_t^T f)^2}{\sigma^8}\right) \\ &- \frac{2I_1(z_t)}{I_0(z_t)} \left(\frac{y_t(m + \xi_t^T f)}{\sigma^6}\right), \end{aligned} \tag{7.79}$$

where $I_1^{'}(z)$ is the first order derivative of $I_1(z)$ with respect to z, $\xi_{i,t}$ is the i-th element of ξ_t, and z_t is defined as

$$z_t \triangleq \frac{y_t(m + \xi_t^T f)}{\sigma^2}. \tag{7.80}$$

Recall that the indices of activation amplitude f_i and f_j are $i = 1, \ldots, L$ and $j = 1, \ldots, L$, respectively. Taking the expectations of above obtained second derivatives and evaluating them at $f = 0$ yield partitions of the FIM. Firstly, we have

$$\left(\mathcal{I}_{11}(f)\,\Big|_{f=0}\right)_{ij} \triangleq \frac{1}{n}\cdot\mathbf{E}\left[\frac{\partial^2(-\log\mathscr{L})}{\partial f_i \partial f_j}\right]\,\Bigg|_{f=0} = \frac{(\sigma^2 - A^0(m,\sigma^2))}{n\sigma^4}\cdot\sum_{t=1}^{n}\xi_{i,t}\xi_{j,t}, \quad (7.81)$$

leading to

$$\mathcal{I}_{11}(0) = \frac{(\sigma^2 - A^0(m,\sigma^2))}{n\sigma^4}\cdot\sum_{t=1}^{n}\xi_t\;\xi_t^T. \quad (7.82)$$

Secondly, we have

$$\left(\mathcal{I}_{12}(f,m,\sigma^2)\,\Big|_{f=0}\right)_{i1} \triangleq \frac{1}{n}\cdot\mathbf{E}\left[\frac{\partial^2(-\log\mathscr{L})}{\partial f_i \partial m}\right]\,\Bigg|_{f=0} = \frac{(\sigma^2 - A^0(m,\sigma^2))}{n\sigma^4}\cdot\sum_{t=1}^{n}\xi_{i,t}, \quad (7.83)$$

and

$$\left(\mathcal{I}_{12}(f,m,\sigma^2)\,\Big|_{f=0}\right)_{i2} \triangleq \frac{1}{n}\cdot\mathbf{E}\left[\frac{\partial^2(-\log\mathscr{L})}{\partial f_i \partial \sigma^2}\right]\,\Bigg|_{f=0} = \frac{(mA^0 + \sigma^2 B^0(m,\sigma^2) - m\sigma^2)}{n\sigma^6}\cdot\sum_{t=1}^{n}\xi_{i,t}, \quad (7.84)$$

leading to

$$\mathcal{I}_{12}(0,m,\sigma^2) = \left(\frac{1}{n}\sum_{t=1}^{n}\xi_t\right)\cdot\left[\frac{\sigma^2 - A^0(m,\sigma^2)}{\sigma^4}, \quad \frac{(mA^0(m,\sigma^2) + \sigma^2 B^0(m,\sigma^2) - m\sigma^2)}{\sigma^6}\right], \quad (7.85)$$

where $\mathcal{I}_{12}(0,m,\sigma^2)$ has a size of $L \times 2$ and ξ_t has a size of $L \times 1$. Therefore, we have

$$\mathcal{I}_{21}(m,\sigma^2,f)\,\Big|_{f=0} = \mathcal{I}_{12}^T(0,m,\sigma^2), \quad (7.86)$$

whose size is $2 \times L$.

Finally, we have the partition of the FIM for m and σ^2 whose size is 2×2. The entries of $\mathcal{I}_{22}(m,\sigma^2)$ are given by

$$\left(\mathcal{I}_{22}(m,\sigma^2)\,\Big|_{f=0}\right)_{11} \triangleq \frac{1}{n}\cdot\mathbf{E}\left[\frac{\partial^2(-\log\mathscr{L})}{\partial m^2}\right]\,\Bigg|_{f=0} = \frac{\sigma^2 - A^0(m,\sigma^2)}{\sigma^4}, \quad (7.87)$$

$$\left(\mathcal{I}_{22}(m,\sigma^2)\,\Big|_{f=0}\right)_{12} \triangleq \frac{1}{n}\cdot\mathbf{E}\left[\frac{\partial^2(-\log\mathscr{L})}{\partial m \partial \sigma^2}\right]\,\Bigg|_{f=0} = \frac{(mA^0(m,\sigma^2) + \sigma^2 B^0(m,\sigma^2) - m\sigma^2)}{\sigma^6}, \quad (7.88)$$

$$\left(\mathcal{I}_{22}(m,\sigma^2)\,\Big|_{f=0}\right)_{22} \triangleq \frac{1}{n}\cdot\mathbf{E}\left[\frac{\partial^2(-\log\mathscr{L})}{\partial(\sigma^2)^2}\right]\,\Bigg|_{f=0}$$

$$= \frac{\sigma^4 - m^2A^0(m,\sigma^2) - 2m\sigma^2\left(B^0(m,\sigma^2)-m\right)}{\sigma^8}, \quad (7.89)$$

leading to

$$\mathcal{I}_{22}(m,\sigma^2) = \begin{bmatrix} \frac{\sigma^2-A^0}{\sigma^4} & \frac{mA^0+\sigma^2B^0-m\sigma^2}{\sigma^6} \\ \frac{mA^0+\sigma^2B^0-m\sigma^2}{\sigma^6} & \frac{\sigma^4-m^2A^0(m,\sigma^2)-2m\sigma^2\left(B^0(m,\sigma^2)-m\right)}{\sigma^8} \end{bmatrix}. \quad (7.90)$$

From (7.82), (7.85), and (7.90), we obtain the FIM in (7.37). ■

CHAPTER 8

Conclusions and Future Works

8.1 Summaries and Conclusions

In this thesis, we mainly pursued one important goal in FMRI data analysis and solved related problems toward that goal. The main question was how to create an activation map from collected measurements, showing regional reactions of the human brain to given stimuli in the experiment. The activation map is a spatial plot of detection statistics derived from hypothesis testings and building it has been one of the most significant tasks from the very beginning of FMRI. We developed a detection statistic for activation considering spatial and temporal correlations without space-time separability in Chapter 3. Related issues involved signal and noise modeling in multi dimensions to implement a detection statistic (Chapter 4), e.g., four dimensions (4D) for a volumetric dataset, building up a test procedure for space and time separability (Chapter 5), and constructing a proper criterion to rank or a performance measure to compare competing detection approaches (Chapter 6). Another important issue we dealt with in the thesis was recently introduced and is for detecting an activation from observed signals with low SNR (Chapter 7). Although high SNR is a typically valid assumption in FMRI until now, since we have a fundamental tradeoff between spatial resolution and SNR and we have a clear trend toward high spatial resolution, the demands for a SNR robust detection, i.e., Rician detection, are rapidly increasing.

Summaries

Because of the importance of "the activation study" in FMRI data analysis, there have been many approaches to build up a detection statistic for activation. However, with rare exceptions, most attention has focused on temporal characteristic of noise under the assumption of spatial independence. In addition, most data analysis methods have implicitly been under the assumption of space-time separability. For the first time in FMRI, we developed an activation statistic, spatio-temporal likelihood ratio test (ST-LRT), from a model jointly considering spatial and temporal correlation of background noise without space-time separability. This is one of the main contributions of Chapter 3, in which it was shown that a spatiotemporal whitening filter was necessary to implement ST-LRT. ST-LRT under space-time separability indicated that, to obtain a properly formulated activation statistic, we needed a spatial whitening kernel, not a conventional Gaussian amplitude kernel for spatial smoothing. Simulations showed ST-LRT performed better than the conventional F-statistic did in terms of detection power.

To implement the proposed ST-LRT, signal and noise modeling in three dimensions (3D) or 4D was necessary. Since a widely used AR-based method in spectral estimation required huge computations due to non-linearity of asymptotic likelihood function, we introduced a new method to the FMRI community. Our modeling approach was unusual, based on a truncated cepstrum expansion, but allowed linear model fitting, thus leading to a substantial amount of computational reductions. To compare ST-LRT and a detection statistic derived from spatial smoothing by Gaussian kernel (SSK) and dynamic linear model (DLM), a model comparison method based on Akaike information criterion (AIC) was developed as well. These are the main contributions of Chapter 4. The developed model comparison method based on AIC showed that the ST-LRT model was on average much closer to the underlying unknown truth than SSK-DLM for a tested human dataset involving motor responses.

The proposed modeling method, the parametric cepstrum, had an additional advantage which was useful for space-time separability. In the framework of the cepstrum, a testing procedure for space-time separability was developed and its asymptotic power was analyzed in Chapter 5. Remarkably, the asymptotic power function only involved cepstral coefficients in the non-separable region, parameters of interest, and depended on them in a quadratic way. In other words, the asymptotic power was independent of nuisance parameters, e.g., activation amplitudes and cepstral coefficients in the separable region. The simulations and application to a human dataset indicated that the developed test procedure worked well.

For the purpose of performance comparison, an asymptotic relative efficiency (ARE) of ST-LRT and a detection statistic from SSK-DLM was evaluated in Chapter 6. Pitman's ARE was derived for ST-LRT and t-statistic from SSK-DLM. It turned out that Pitman's ARE had a simple form and depended on spatial operators used to build up detection statistics, namely, spatial whitening filter and Gaussian amplitude kernel. For the dataset used in simulations of Chapter 3, it was shown that ST-LRT was much more efficient than t-statistic derived from SSK-DLM. In other words, ST-LRT required much less number of samples than a detection statistic from SSK-DLM to achieve the same detection power, reducing the FMRI experiment time.

For SNR robust activation detection, a method to build up an activation map from a Rician distributed model was suggested in Chapter 7. An iterative method for parameter estimation was developed via an EM algorithm, providing a simpler and more reliable updates compared to conventional Rician modelings. From estimated parameters, a detection statistic for activation was constructed and its asymptotic power was analyzed. On the contrary to Chapter 5, the asymptotic power of Rician detection depended on both of interesting and nuisance parameters such as baseline. In simulations, it was shown that Rician-EM approach provided parameter estimates which were less biased and had smaller MSEs than conventional Gaussian modeling for all ranges of baseline to noise ratios (BNRs).

Conclusions

By comparing ROC curves in Chapter 3, AIC maps in Chapter 4, and measuring Pitman's ARE in Chapter 6, we arrived at the conclusion that the proposed ST-LRT provided better detection statistics than did the conventional SSK-DLM in terms of detection power, average distance to a underlying unknown truth, and asymptotic relative efficiency. From the discussions comparing ST-LRT and SSK-DLM, we have an important conclusion that spatial whitening operation is necessary for proper activation detection rather than the spatial smoothing by Gaussian kernel, which has been widely used but under many ad hoc assumptions. Additionally, from simulations in Chapter 7, it was shown that the proposed Rician-EM approach was much more robust to activation to noise ratio (ANR) and BNR than conventional Gaussian modeling.

8.2 Future Works

From the discussions presented in the thesis, we provide interesting future research topics in this section.

Detection Statistic for Activation with Spatial Non-stationarity

In Chapter 3, we developed ST-LRT from a FMRI measurement model jointly considering spatial and temporal correlation of background noise without space-time separability. However, the spatial stationarity was assumed to develop ST-LRT. Therefore, a natural extension will be the development of a detection statistic for activation based on a measurement model considering spatial non-stationarity. The idea of the local spatial regularization of likelihood function in [39, 57] can be a good starting point. Based on this extension, we expect that a test procedure for space-time separability for spatially non-stationary random fields can be developed.

Asymptotic Relative Efficiency for the Whole ROI

In Chapter 6, we developed voxel-wise Pitman's ARE to compare the proposed ST-LRT and F-statistic derived from existing SSK-DLM. Since we usually have sparse activations in the brain, most voxels are not activated. Therefore, a next step will be the development of Pitman's ARE for the whole ROI. In addition, to compare two competing detection statistics in terms of other types of AREs is an interesting topic as well, for example, Bahadur's ARE [50].

Rician-EM Modeling with Serial Correlation

In Chapter 7, we derived EM algorithm from a Rician distributed model under the assumption of temporal independence. A natural extension from this thesis will be the development of an EM algorithm from a Rician distributed model with serial correlation to model FMRI measurements more exactly.

Asymptotic Power Analysis for Mismatched Modeling

In Chapter 7, we analyzed the asymptotic power of a detection statistic derived from a Rician distributed model. This analysis was under the assumption that the model and observed FMRI samples obey Rician distributions. Since Gaussian modeling has been widely used in FMRI data analysis, analyzing the asymptotic power of Gaussian modeling is very interesting when measured samples are Rician distributed.

BIBLIOGRAPHY

[1] M. Abramowitz and I.A. Stegun. Handbook of mathematical functions with formulas, graphs, and mathematical tables. 9th printing*, New York; Dover*, pages 374–377, 1972.

[2] R.J. Adler. The Geometry of Random Fields. *Wiely, New York*, 1981.

[3] D.E. Amos. Computation of modified Bessel functions and their ratios. *Mathematics of Computation*, 28, No. 125, Jan. 1974.

[4] P.J. Bickel and K.A. Doksum. Mathematical statistics : basic ideas and selected topics. *San Francisco, CA : Holdenday*, 1977.

[5] G.M. Boynton, S.A. Engel, G.G. Glover, and D.J. Heeger. Linear systems analysis of functional magnetic resonance imaging in Human V1. *Journal of Neuroscience*, 16, No.13:4207–4221, 1996.

[6] D.R. Brillinger. Fourier analysis of stationary processes. *Proc. IEEE*, 62:628–642, 1974.

[7] D.R. Brillinger. Time Series : Data Analysis and Theory. *San Francisco, CA : Holdenday*, 1981.

[8] E. Bullmore, M. Brammer, S. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham. Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35:261–277, 1996.

[9] M.A. Burock and A.M. Dale. Estimation and detection of event-related FMRI signals with temporally correlated noise : a statistically efficient and unbiased approach. *Human Brain Mapping*, 11:249–260, 2000.

[10] R.B. Buxton. Introduction to functional magnetic resonance imaging : principles and techniques. *Cambridge University Press*, 2001.

[11] M.S. Cohen. Parametric analysis of FMRI data using linear system methods. *NeuroImage*, 6:93–103, 1997.

[12] M.S. Cohen and S.Y. Bookheimer. Localization of brain function using magnetic resonance imaging. *Trends in Neurosciences*, 17:268–276, 1994.

[13] R. Dahlhaus and H. Künsch. Edge effects and efficient parameter estimation for stationary random fields. *Biometrika*, 74, No.4:877–882, Dec. 1987.

[14] A.M. Dale and R.L. Buckner. Selective averaging of rapidly presented individual trials using FMRI. *Human Brain Mapping*, 5:329–340, 1997.

[15] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Stat. Soc. Series B*, 23(1):1–38, 1977.

[16] A.J. den Dekker and J. Sijbers. Detection of brain activation from magnitude fmri data using a generalized likelihood ratio test. *Proc. EUROSIPCO, Vienna, Austria*, Sep. 2004.

[17] A.J. den Dekker and J. Sijbers. Implication of the Rician distribution for FMRI generalized likelihood ratio tests. *Magnetic Resonance Imaging*, 23:953–959, 2005.

[18] X. Descombes, F. Kruggel, and D.Y. von Cramon. Spatio-temporal FMRI analysis using Markov random fields. *IEEE Trans. on Medical Imaging*, 17, No.6:1028–1039, Dec. 1998.

[19] B. Dickinson. Two dimensional Markov spectrum estimates need not exist. *IEEE Trans. Information Theory*, IT-26, No.1:120–121, Jan. 1976.

[20] M.P. Ekstrom and J.W. Woods. Two-dimensional spectral factorization with application in recursive digital filtering. *IEEE Trans. on Acoust. Speech and Signal Processing*, ASSP-24, No.2:115–128, April 1976.

[21] R.S.J. Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, C.J. Price, S. Zeki, J. Ashburner, and W. Penny. Human Brain Function. 2nd edition*, Academic Press USA*, 2004.

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

[23] K.J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI time series. *Human Brain Mapping*, 1:153–171, 1994.

[24] M. Fuentes. Testing for separability of spatial-temporal covariance functions. *Journal of Statistical Planning and Inference*, 136:447–466, 2005.

[25] A. Geissler, R. Lanzenberger, M. Barth, A.R. Tahamtan, D. Milakara, A. Gartus, and R. Beisteiner. Influence of FMRI smoothing procedures on replicability of fine scale motor localization. *NeuroImage*, 24:323–331, 2005.

[26] C. Goutte, F. Nielson, and L.K. Hansen. Modeling the hemodynamic response in FMRI using smooth FIR filters. *IEEE Trans. on Medical Imaging*, 19(12), Dec. 2000.

[27] X. Guyon. Random Fields on a Network : modeling, statistics, and applications. *Springer-Verlag*, 1995.

[28] T. Hayakawa. The likelihood criterion for a composite hypothesis under a local alternative. *Biometrika*, 62, No.2:451–460, Aug. 1975.

[29] A.P. Holmes, R.C Blair, J.D.G. Watson, and I. Ford. Non-parametric analysis of statistic images from functional mapping experiments. *Journal of Cerebral Blood Flow and Metabolism*, 16, No.1:7–22, 1996.

[30] A.K. Jain. Fundamentals of digital image processing. *Prentice Hall*, pages 204–213, 1989.

[31] S.M. Kay. Fundamentals of statistical signal processing, Vol.2 : detection theory. *Prentice Hall : Upper Saddle River, NJ*, 1998.

[32] H. Linhart and W. Zucchini. Model Selection. *New York : Wiley*, 1986.

[33] J. Locascio, P. Jennings, C. Moore, and S. Corkin. Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance imaging. *Hum. Brain Mapp.*, 5:168–193, 1997.

[34] C. Long, E.N. Brown, D. Manoach, and V. Solo. Spatiotemporal wavelet analysis for functional MRI. *NeuroImage*, 23:500–516, 2004.

[35] R.J. Martin. A metric for ARMA processes. *IEEE Trans. on Signal Processing*, 48, No.4:1164–1170, 2000.

[36] M.W. Mitchell, M.G. Genton, and M.L. Gumpertz. Testing for separability of space-time covariances. *Environmetrics*, 16:819–831, 2005.

[37] F.Y. Nan and R.D. Nowak. Generalized likelihood ratio detection for FMRI using complex data. *IEEE Trans. on Medical Imaging*, 18, No.4:320–329, Apr. 1999.

[38] T. Nichols and S. Hayasaka. Controlling the familywise error rate in functional neuroimaging : a comparative review. *Stat. Meth. in Med. Research*, 12:419–446, 2003.

[39] P. Purdon, V. Solo, R. Weisskoff, and E. Brown. Locally regularized spatiotemporal modeling and model comparison for functional MRI. *NeuroImage*, 14:912–923, 2001.

[40] S.O. Rice. Mathematical analysis of random noise. *Bell Syst. Tech. J.*, 23:282, 1944.

[41] J. Riera, J. Bosch, O. Yamashita, R. Kawashima, N. Sadato, T. Okada, and T. Ozaki. FMRI activation maps based on the NN-ARx model. *NeuroImage*, 23:680–697, 2004.

[42] D.L. Robinson. Estimation and use of variances components. *The Statistician*, 36 3 14.

[43] B.R. Rosen, R.L. Buckner, and A.M. Dale. Event-related functional MRI : past, present, and future. *Proc. Natl. Acad. Sci.*, 95, Feb. 1998.

[44] A. Rosenfeld and A.C. Kak. Digital Picture Processing. Vol. 2*, Orlando FA : Academic Press*, 1982.

[45] D.B. Rowe. Modeling both the magnitude and phase of complex valued FMRI data. *NeuroImage*, 25, No.4:1310–1324, May 2005.

[46] D.B. Rowe. Parameter estimation in the magnitude only and complex valued FMRI data models. *NeuroImage*, 25, No.4:1124–1132, May 2005.

[47] D.B. Rowe and B.R. Logan. A complex way to compute FMRI activation. *NeuroImage*, 23, No.3:1078–1092, Nov. 2004.

[48] S. Saha, C.J. Long, E. Brown, E. Aminoff, M. Bar, and V. Solo. Hemodynamic transfer function estimation with Laguerre polynomials and confidence interval construction, from functional magnetic resonance imaging data. *In Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing, Montreal Canada*, May 2004.

[49] G.A. Seber and A.J. Lee. Linear Regression Analysis. 2nd edition*, Wiley*, 2003.

[50] R.J. Serfling. Approximation Theorems of Mathematical Statistics. *Wiley*, 1980.

[51] G. Sharma and R. Chellappa. Two-dimensional spectrum estimation using noncausal autoregressive models. *Trans. on Information Theory*, IT-32, No.2:268–275, Mar. 1986.

[52] R.H. Shumway and D.S. Stoffer. Time series analysis and its applications. *Springer*, 2000.

[53] D.O. Siegmund and K.J. Worsley. Testing for a signal with unknown location and scale in a stationary Gaussian random field. *Annals of statistics*, 23:608–639, 1995.

[54] J. Sijbers and A.J. den Dekker. Generalized likelihood ratio tests for complex FMRI data : a simulation study. *IEEE Trans. on Medical Imaging*, 24, No.5:604–611, May 2005.

[55] V. Solo. Modeling of two-dimensional random fields by parametric cepstrum. *IEEE Trans. on Information Theory*, IT-32, No.6:743–750, Nov. 1986.

[56] V. Solo, C.J. Long, E.N. Brown, E. Aminoff, M. Bar, and S. Saha. FMRI signal modeling using Laguerre polynomials. *In Proc. IEEE Intl. Conf. on Image Processing, Singapore*, Oct. 2004.

[57] V. Solo, P. Purdon, R. Weisskoff, and E. Brown. A signal estimation approach to functional MRI. *IEEE Trans. on Medical Imaging*, 20(1):26–35, 2001.

[58] K. Tamaki. Second order asymptotic properties of a class of test statistics under the existence of nuisance parameters. *Sci. Math. Jpn.*, 61, No.1:119–143, 2005.

[59] A.W. van der Vaart. Asymptotic Statistics. *Cambridge University Press*, 1998.

[60] D. Van De Ville, T. Blu, and M. Unser. Integrated wavelet processing and spatial statistical testing of FMRI data. *NeuroImage*, 23:1472–1485, 2004.

[61] B. Walberg. System Identification using Laguerre Models. *IEEE Trans. Automatic Control*, vol.36, No.5, May 1991.

[62] P. Whittle. On stationary processes in the plane. *Biometrika*, 41, No.3/4:434–449, Dec. 1954.

[63] S.S. Wilks. The large-sample distribution of the likelihood ratios for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9, No.1:60–62, Mar. 1938.

[64] M.W. Woolrich, M. Jenkinson, J.M. Brady, and S.M. Smith. Fully Bayesian spatio-temporal modeling of FMRI data. *IEEE Trans. on Medical Imaging*, 23, No.2:213–231, Feb. 2004.

[65] K. Worsley, A. Evans, S. Marett, and P. Neelin. A three dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, 12:900–918, 1992.

[66] K.J. Worsley and K.J. Friston. Analysis of FMRI time series revisited - again. *NeuroImage*, 2:173–181, 1995.

[67] K.J. Worsley, S. Marrett, P. Neelin, and A.C. Evans. Searching scale space for activation in PET images. *Human Brain Mapping*, 4:74–90, 1996.