



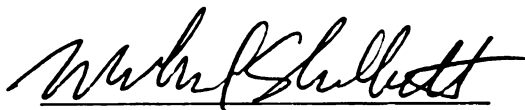
This is to certify that the  
dissertation entitled  
NOVEL DETECTION METHODS FOR  
FUNCTIONAL MAGNETIC RESONANCE IMAGING

presented by

Fangyuan Nan

has been accepted towards fulfillment  
of the requirements for

Ph.D degree in Electrical Eng

  
Major professor

Date 5/5/2000

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.  
**MAY BE RECALLED** with earlier due date if requested.

| DATE DUE | DATE DUE | DATE DUE |
|----------|----------|----------|
|          |          |          |
|          |          |          |
|          |          |          |
|          |          |          |
|          |          |          |
|          |          |          |
|          |          |          |
|          |          |          |
|          |          |          |
|          |          |          |

# Novel Detection Methods for Functional Magnetic Resonance Imaging

By

*Fangyuan Nan*

A Dissertation

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

2000

# ABSTRACT

## Novel Detection Methods for Functional Magnetic Resonance Imaging

By

*Fangyuan Nan*

This dissertation considers the detection problem in functional magnetic resonance imaging (fMRI), *i.e.*, to determine which parts of the brain show active response to some stimulus. Some basic magnetic resonance imaging and fMRI background is introduced in the beginning. The underlying principle is that subtle variations in the image intensity over time can be detected to reveal brain activity.

The detection problem is first attacked on a pixel by pixel basis. A new nonlinear detector for fMRI based on Generalized Likelihood Ratio Test (GLRT) is systematically studied. Theoretical analysis and Monte Carlo simulation are used to explore the performance of the new detector. At relatively low baseline signal intensities, the GLRT detector outperforms both the conventionally used magnitude correlation (MC) detector and the newly proposed complex correlation (CC) test. At high baseline signal intensities, the nonlinear GLRT performs as well as the standard MC test and significantly better than the CC test.

fMRI signals are actually both temporally and spatially dependent. Pixel-wise

detection, however, considers only temporal correlation information and ignores spatial correlation information. In order to remedy this deficiency, the dissertation then uses a multi-scale image segmentation algorithm to first segment an fMRI correlation image into several regions, each with homogeneous statistical behavior. A single pixel detection algorithm is then applied to each homogeneous region. Extensive simulations demonstrate the efficacy of the new method.

## ACKNOWLEDGMENTS

I'm most grateful to God, Jesus and Li Hongzhi for giving me a new birth and teaching me a lot of unearthly wisdom. *Everyone the Lord has rescued from trouble should praise him* [Bible – Psalm 107:1-2].

I'm very grateful to my grandmother and my parents, Mr. Shu-Guang Nan and Mrs. Benzhen Li, for nurturing me. The unfailing love from them as well as from my sisters, Mrs. Xiaoping Nan and Mrs. Xueqing Nan, is always a strong encouragement and support for my life. *Pay attention to your father, and don't neglect your mother when she grows old. Invest in truth and wisdom, discipline and good sense, and don't part with them. Make your father truly happy by living right and showing sound judgment. Make your parents proud, especially your mother* [Bible — Proverbs 23:22-25].

I'm grateful to my advisor Dr. Nowak, a brilliant young scholar and a very nice gentleman, who helped me greatly not only in my research, but also in my life. Without his help, this dissertation would have been impossible. I'd also like to express gratitude to Dr. Deller, Dr. Pierre and Dr. LePage for their willingness to serve on my committee and for their helpful suggestions for my research and this dissertation.

I'm very thankful to many friends' help at a very difficult time during my research. I am obligated to mention the following names: Dr. Robert Barger and Dr. Paul Liu, Tom and Ruth Shillair, Yi Wan, Dr. Yingcheng Dai and his wife Dr. Qing Li, Dr. Dimitri Mihalas, Dr. William Schmidt, Dr. Jen-Je Su and his wife Dr. Shu-Shyan Wong, Ms. Smith, Dr. Yingnian Wu and his wife Ms. Liping Li.

I'm thankful to many teachers. Among them, of particular importance and help to me are my previous advisors: Prof. Siyong Zhou at Beijing Institute of Technology, Prof. E. C. Young, Prof. Christopher Hunter at Florida State University (FSU), and Prof. Kun-Mu Chen and Prof. Edward Rothwell at Michigan State University (MSU).

I am also indebted to the Dept. of Mathematics at FSU and the Dept. of Electrical and Computer Engineering at MSU. Without the teaching assistantship provided by them, my graduate studies in the States would be impossible.

# TABLE OF CONTENTS

|                                                                       |             |
|-----------------------------------------------------------------------|-------------|
| <b>LIST OF TABLES</b>                                                 | <b>vii</b>  |
| <b>LIST OF FIGURES</b>                                                | <b>viii</b> |
| <b>1 Introduction</b>                                                 | <b>1</b>    |
| 1.1 How does fMRI work? . . . . .                                     | 4           |
| 1.2 Characteristics of the fMRI Signal . . . . .                      | 5           |
| 1.3 Outline for Detection Methods for fMRI . . . . .                  | 5           |
| 1.4 Organization of This Dissertation . . . . .                       | 6           |
| <b>2 Pixel-wise Detection for fMRI</b>                                | <b>9</b>    |
| 2.1 Overview . . . . .                                                | 9           |
| 2.2 fMRI Signal Model . . . . .                                       | 12          |
| 2.3 Generalized Likelihood Ratio Tests . . . . .                      | 17          |
| 2.4 Method 1: Magnitude Correlation Detection . . . . .               | 19          |
| 2.5 Method 2: Complex Correlation Detection . . . . .                 | 21          |
| 2.6 Method 3: Nonlinear GLR Detection . . . . .                       | 23          |
| 2.6.1 Derivation of the GLRT Test Statistic . . . . .                 | 24          |
| 2.6.2 Invariance of GLRT Test Statistics . . . . .                    | 28          |
| 2.6.3 Upperbound of GLRT Test Statistics . . . . .                    | 32          |
| 2.6.4 Asymptotic of GLRT Test Statistics . . . . .                    | 33          |
| 2.6.5 Threshold Selection Based on Numerical Simulation . . . . .     | 34          |
| 2.6.6 Distribution of GLRT Test Statistics . . . . .                  | 34          |
| 2.7 Comparisons of the Three Detectors . . . . .                      | 35          |
| 2.8 A Simulated fMRI Study . . . . .                                  | 38          |
| <b>3 Multi-scale Detection for fMRI</b>                               | <b>42</b>   |
| 3.1 Overview . . . . .                                                | 42          |
| 3.1.1 Spatial Modeling and Outline for the Method . . . . .           | 43          |
| 3.1.2 General Setting of Bayesian Image Segmentation . . . . .        | 44          |
| 3.1.3 Multi-scale Image Segmentation Methods and Advantages . . . . . | 45          |

|       |                                                                                                        |    |
|-------|--------------------------------------------------------------------------------------------------------|----|
| 3.2   | Multi-scale Analysis . . . . .                                                                         | 47 |
| 3.2.1 | 1-D Multi-resolution Analysis (MRA) . . . . .                                                          | 47 |
| 3.2.2 | Properties of the Discrete Wavelet Transform . . . . .                                                 | 54 |
| 3.3   | Image Segmentation by Multi-scale Markov Modeling Discrete Field .                                     | 55 |
| 3.3.1 | Main Points of Algorithm I . . . . .                                                                   | 55 |
| 3.3.2 | Simulation Result . . . . .                                                                            | 57 |
| 3.4   | Image Segmentation by Multi-scale Hidden Markov Model (MHMM)<br>of Wavelet Coefficients . . . . .      | 58 |
| 3.4.1 | Likelihood Function for fMRI Data . . . . .                                                            | 60 |
| 3.4.2 | Key Point for Edge Detection and <i>a Prior</i> Distribution for the<br>Wavelet Coefficients . . . . . | 64 |
| 3.4.3 | A Failure Modeling . . . . .                                                                           | 67 |
| 3.4.4 | Solution for Joint <i>a Posteriori</i> State Probability . . . . .                                     | 69 |
| 3.4.5 | Marginal <i>a Posteriori</i> State Probability Calculation . . . . .                                   | 70 |
| 3.4.6 | Image Label Estimation . . . . .                                                                       | 72 |
| 3.4.7 | Extension to 2 Dimensions . . . . .                                                                    | 73 |
| 3.4.8 | A Simulation of the New Segmentation Method . . . . .                                                  | 74 |
| 3.5   | Processing Results for fMRI Data by Two-Step Approach . . . . .                                        | 76 |
| 4     | <b>Discussions and Conclusions</b>                                                                     | 81 |
| 4.1   | For Pixelwise Detection . . . . .                                                                      | 81 |
| 4.2   | Consideration on Spatial Information . . . . .                                                         | 83 |
| 4.3   | Epilogue . . . . .                                                                                     | 84 |
| 5     | <b>Appendices</b>                                                                                      | 85 |
| 5.1   | Appendix I: Notations and Some Results on Projection Matrix . . . .                                    | 85 |
| 5.2   | Appendix II: Some Preliminary Results on $\chi^2$ and $F$ Distribution . .                             | 86 |
| 5.3   | Appendix III: Definition of Rician, Rayleigh and $t$ Distribution . . . .                              | 88 |
| 5.4   | Appendix IV: Analysis of $t$ Test Used in fMRI Detection . . . . .                                     | 89 |
| 5.5   | Appendix V: Principle of Invariant Test . . . . .                                                      | 91 |
| 5.6   | Appendix VI: SMAP Algorithm for Multi-scale Image Segmentation .                                       | 94 |
| 5.6.1 | <i>A Prior</i> Consideration . . . . .                                                                 | 95 |
| 5.6.2 | Likelihood Function . . . . .                                                                          | 96 |
| 5.6.3 | Criterion and Solution . . . . .                                                                       | 97 |

## BIBLIOGRAPHY

101

# **LIST OF TABLES**

|     |                                             |    |
|-----|---------------------------------------------|----|
| 2.1 | $P_d$ with $P_f = 0.01, N = 120$ . . . . .  | 36 |
| 2.2 | $P_d$ with $P_f = 0.025, N = 120$ . . . . . | 36 |
| 2.3 | $P_d$ with $P_f = 0.05, N = 120$ . . . . .  | 37 |

## LIST OF FIGURES

|     |                                                                                                                                                                                                                                                                             |    |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | A block diagram of an MRI system. . . . .                                                                                                                                                                                                                                   | 3  |
| 1.2 | A series of images are acquired in fMRI to detect neuronal activity. .                                                                                                                                                                                                      | 4  |
| 1.3 | Illustration of fMRI stimulus and responses. a) Controlled stimulus;<br>b) Response from an activated area; c) Response from an inactivated<br>area. . . . .                                                                                                                | 8  |
| 2.1 | (a) Activation-baseline pattern for an fMRI experiment; (b) Modeling<br>the activated pixel as a system. . . . .                                                                                                                                                            | 14 |
| 2.2 | One time series from a real fMRI experiment. (a) Real part; (b) Imag-<br>inary part; (c) Phase. . . . .                                                                                                                                                                     | 16 |
| 2.3 | Structure of matched subspace detector. . . . .                                                                                                                                                                                                                             | 23 |
| 2.4 | Three curves comparing the performance of three detectors. (a) $a/\sigma =$<br>1; (b) $a/\sigma = 3.162$ ; (c) $a/\sigma = 10$ . . . . .                                                                                                                                    | 39 |
| 2.5 | A simulated fMRI experiment illustrating pixel-wise detection by three<br>detectors. (a) Brain image with simulated activation region high-<br>lighted; (b) MC test results: $P_d = 0.77$ ; (c) CC test results: $P_d = 0.70$ ;<br>(d) GLRT results: $P_d = 0.79$ . . . . . | 41 |
| 3.1 | Illustration of image segmentation. . . . .                                                                                                                                                                                                                                 | 46 |
| 3.2 | Nested scale spaces and wavelet spaces . . . . .                                                                                                                                                                                                                            | 52 |
| 3.3 | Computation of DWT by filter bank . . . . .                                                                                                                                                                                                                                 | 53 |
| 3.4 | Synthesis by filter bank . . . . .                                                                                                                                                                                                                                          | 53 |
| 3.5 | Pyramid structure of the MSRF. (a) Continuous image $Y$ at different<br>scales. (b) The random field $X^j$ at each scale is causally dependent on<br>the coarser scale field $X^{j-1}$ above it. . . . .                                                                    | 56 |
| 3.6 | Neighborhood structure used in Algorithm I. (a) Quad-tree structure<br>used for 2-D case; (b) Binary tree structure used for 1-D case. . . . .                                                                                                                              | 58 |
| 3.7 | One simulated image segmentation using Algorithm I. (a) Original<br>noisy image; (b) Segmentation result. . . . .                                                                                                                                                           | 59 |
| 3.8 | Binary data tree structure for Harr wavelet analysis. . . . .                                                                                                                                                                                                               | 61 |
| 3.9 | Propagation of wavelet coefficients (three scales) . . . . .                                                                                                                                                                                                                | 64 |

|      |                                                                                                                                                                                                                                                                                                                           |     |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 3.10 | Wavelet-based HMM. . . . .                                                                                                                                                                                                                                                                                                | 66  |
| 3.11 | Conversion of a 2-D image to 1-D sequence . . . . .                                                                                                                                                                                                                                                                       | 74  |
| 3.12 | One simulated image segmentation by Algorithm II. a) Noisy image;<br>b) Detected edges; c) Segmented image. . . . .                                                                                                                                                                                                       | 75  |
| 3.13 | Processing results for fMRI data by two-step approach. a) One fMRI<br>correlation image; b) Segmented image of (a), also final detection re-<br>sults by combinational use of image segmentation and single pixel de-<br>tection; (c) Another correlation image; (d) Segmentation and detection<br>result of (c). . . . . | 80  |
| 5.1  | Experiment setup and two hypotheses for $t$ -test used in fMRI detection.                                                                                                                                                                                                                                                 | 90  |
| 5.2  | Invariant Test. . . . .                                                                                                                                                                                                                                                                                                   | 93  |
| 5.3  | Segmentation Result Using SMAP Algorithm. . . . .                                                                                                                                                                                                                                                                         | 100 |

# CHAPTER 1

## Introduction

*No matter how advanced the computer is, it is still no match for the human brain, which to this day remains an unfathomable enigma.*

Li Hongzhi, Zhuan Falun

*Magnetic resonance imaging* (MRI) is a powerful diagnostic imaging technique based on the principle of nuclear magnetic resonance (NMR), describing the interaction of nuclei and magnetic fields [31]. A block diagram of an MRI system is shown in Figure (1.1), adapted from [31] and [39]. It is a very complicated system, embracing many aspects of electrical engineering. The patient serves both as a transmitter and as a receiver.

There are three types of magnetic fields in an MRI system [39, 67]. A static and very strong magnetic field is generated by huge superconducting magnets. A much weaker pulsing radio frequency field is employed to generate MR signals. Three sets of orthogonal gradient fields are used for imaging purposes, *i.e.*, to spatially resolve the patient's small structures to form an image. Although other imaging methods such as projection methods do exist, the current trend for MR imaging is to use Fourier inversion. The first set of gradient pulses uses linear changes in field strength to localize a region of interest in the subject's body to be imaged — “slice selection”.

The second set of gradient pulses employs linear changes in frequency to distinguish columns in an image — “frequency encoding”. The third set of gradient pulses utilizes linear changes in phase to distinguish the rows — “phase encoding” [39, 12].

While traditional MRI provides only *static* images to analyze *anatomical* structure, *functional* MRI (fMRI), a newer imaging modality which is based on MRI and just comes to the stage during the past decade, acquires a series of images to detect *neural activity*, that is, to locate where— and how — the brain responds to certain stimuli [12, 37]. In other words, *the central task for fMRI is to obtain maps of active and non-active regions of the brain corresponding to specific stimuli*. Figure (1.2) illustrates a series of images acquired in an fMRI experiment. In this figure, the white small square indicates an activated region; the black small square indicates an inactive region.

Compared with other imaging techniques, such as X-ray computerized tomography (CT), MRI is considered safer since it does not require the subject to be exposed to ionizing radiation [31]. MR images are also of high contrast and resolution. MRI provides more information since MR signals depend on several tissue parameters [31, 37, 68]. In addition to spin density  $\rho(\mathbf{r})$ , the number of NMR visible spins in a given region, there are two principal relaxation times, each of which, in principle, can be used individually or combined [12]. fMRI does not use exogenous agents, an advantage over another popular brain mapping technique known as *positron \*emission tomography* (PET).

Research on fMRI involves substantial knowledge in physics, physiology, neurology, and psychology. This dissertation, however, concentrates on signal processing aspect. Note that the materials in this chapter are just for illustrative purposes; they are not meant to be comprehensive. For details, the reader is encouraged to refer to [12, 31, 37, 39, 68], which are the main sources for materials in this chapter.

---

*An elementary particle with the mass of an electron and a charge of the same amount as the electron's but positive.*

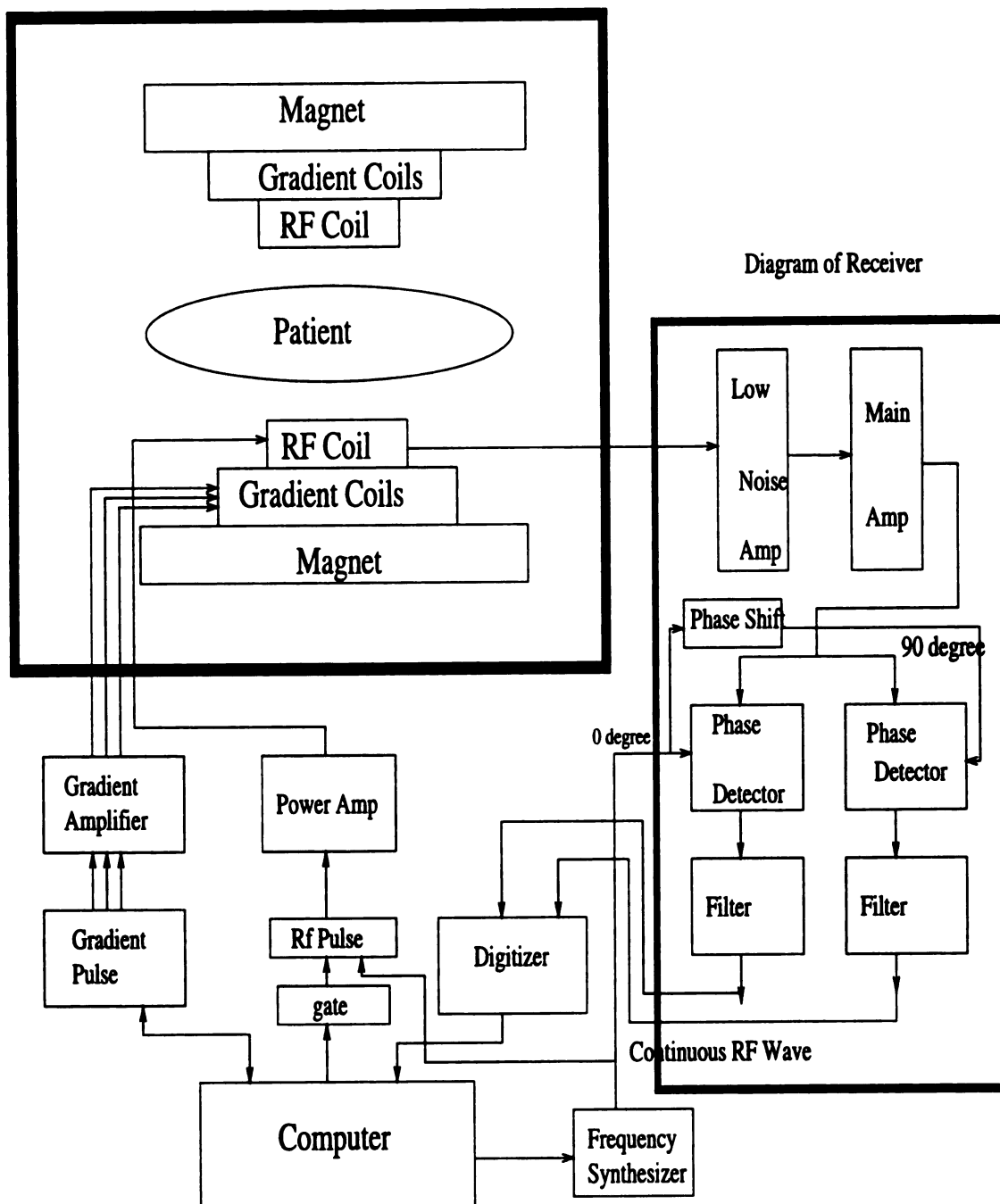


Figure 1.1. A block diagram of an MRI system.

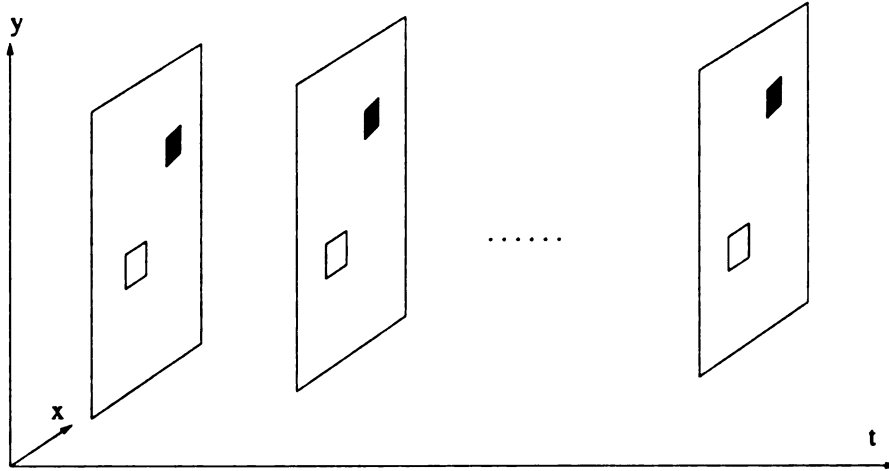


Figure 1.2. A series of images are acquired in fMRI to detect neuronal activity.

## 1.1 How does fMRI work?

In short, there are basically two foundations [12, 37]. The first is due to *physiological* reasons. The other foundation comes from *physics*, that is, *magnetic susceptibility*.<sup>†</sup> Blood contains iron (hemoglobin). Neural activity is linked to blood oxygenation levels in blood vessels close to active neurons. This relationship is called the Blood Oxygenation Level Dependent (BOLD) effect [66]. Specifically, neuronal activity causes an excess of oxygenation level in the blood nearby active brain tissue. De-oxygenated blood is more paramagnetic than oxygenated blood. Changes in the local concentration of de-oxyhemoglobin within the brain lead to alterations in the MR signal. Subtle variations in the magnetic susceptibility of oxygenated and de-oxygenated blood that are then detected in the MR signal indicate neural activity. For more details, refer to [12, 37].

---

Any object placed in a magnetic field will magnetize to a degree slightly more than (**paramagnetic**) or less than (**diamagnetic**) the applied field. The relationship between the field experienced within a sample and the applied field is known as the **magnetic susceptibility** calculated as the ratio of the internal field to the applied field.

## 1.2 Characteristics of the fMRI Signal

Currently, the procedure to do fMRI experiments is to instruct the subject to perform experimental (E) and control (C) tasks in an alternating sequence of some design [12]. Refer to Figure (1.3a). Specifically, the subject is instructed to remain relaxed in the controlled (resting) state, and to perform some kind of consecutive and repetitive task (for example, finger tapping) in the activation (experiment) state [36]. In addition to controlled stimulus, two responses are also made up in Figure (1.3), one of which may be from the white spot (active pixel) in Figure (1.2) and the other of which may be from the black spot (inactive pixel) in Figure (1.2).

It takes some time for fMRI signal to reach its peak after the onset of the stimulus, which is called *Response Latency* [12].

An fMRI time series has a *complicated structure*. It contains noise varying with anatomical location. There is random noise as well as structured noise due to instrumental (MR system characteristics), physiological (cardiac and respiratory pulsations), and experimental (such as patient motion) factors [ 2, 12, 29, 38].

In addition to this complicated structure, the temporal and spatial characteristics of the time series are also unknown [29]. All these render the analysis of fMRI time series very difficult.

The hemodynamic signal changes in fMRI during brain activation are extremely small, from 2 to 5% at moderate magnetic field strengths (1.5T) [12].

## 1.3 Outline for Detection Methods for fMRI

So far most fMRI detection methods are only pixel-wise. Generally, analysis of changes in neural activity is explored using *statistical parametric map* (SPM) [69], which is a two dimensional (2-D) image of a test statistic determined at each pixel by some operation between signal and reference. Originally,  $t$  statistics were usually

used [14]. In repetitive experiments involving a dynamic time sequence of images, a correlation method is now common, in which the correlation between each time series and a reference signal is used to decide whether or not activity is present [3]. It reduces to an  $F$  statistic test. Many generalizations and extensions of this simple idea have been proposed [24, 25, 36, 55, 60, 64]. Chapter 2 is also on pixel-wise detection, but contrary to most methods, complex-valued data are considered.

Detection methods exploiting spatial information (correlation) of fMRI signal have also been recently proposed for fMRI [21, 35, 58]. Most of them use Bayesian strategies. For example in [35, 17], Bayesian principles and Markov Random Field (MRF) models are employed to facilitate joint spatio-temporal analysis of fMRI data. Chapter 3 develops a new multi-scale framework for similar purposes.

## 1.4 Organization of This Dissertation

Chapter 2 stresses pixel-wise testing, which is most common in practice. First a model for the complex fMRI time series is proposed. The most distinct feature of the model is that the baseline *nuisance component* and reference *signal component* share a common phase. A nonlinear detection problem ensues. Based on the classical generalized likelihood ratio test (GLRT), three methods are investigated to attack that detection problem: the conventionally used magnitude correlation (MC) detector, the complex correlation (CC) detector newly proposed by Lai and Glover [36], and a new nonlinear GLRT detector, with emphasis on the last one. The properties of the nonlinear GLR detector are carefully studied and a method for threshold selection from numerical simulations is presented. The nonlinear GLR detector has the best performance among the three [48].

An multi-scale detection method exploiting spatial information is presented in Chapter 3. Specifically, the idea of multi-scale image segmentation is used to improve

the performance of fMRI detection. An multi-scale Bayesian framework for image segmentation is introduced in the beginning. Two image segmentation algorithms are then investigated, with emphasis on the second one whose application to fMRI detection is the second part of this research.

Some discussions and conclusions regarding work of this dissertation are gathered in Chapter 4.

Chapter 5 collects Appendices including some notations used throughout this dissertation, some results on  $\chi^2$ ,  $F$  and  $t$  distributions useful for the development in Chapter 2, the  $t$ -test used in fMRI detection [14], principle of and some definitions and theorems on invariant tests [7, 56] used in Chapter 2, and the sequential maximal *a posterior* (SMAP) algorithm for multi-scale image segmentation [10].

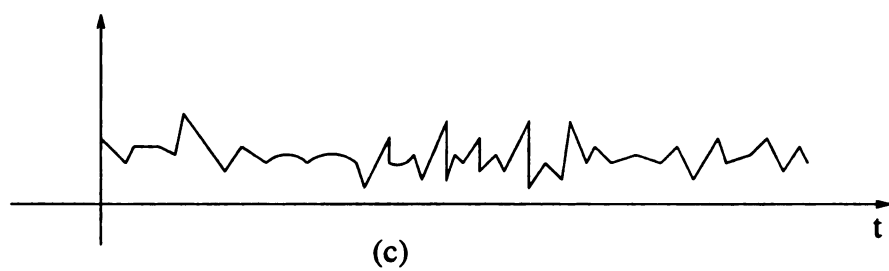
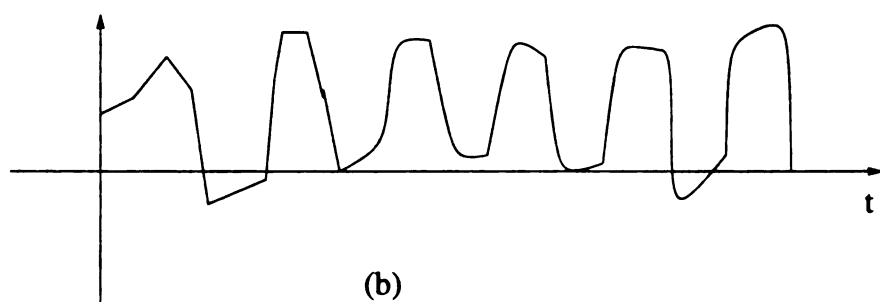
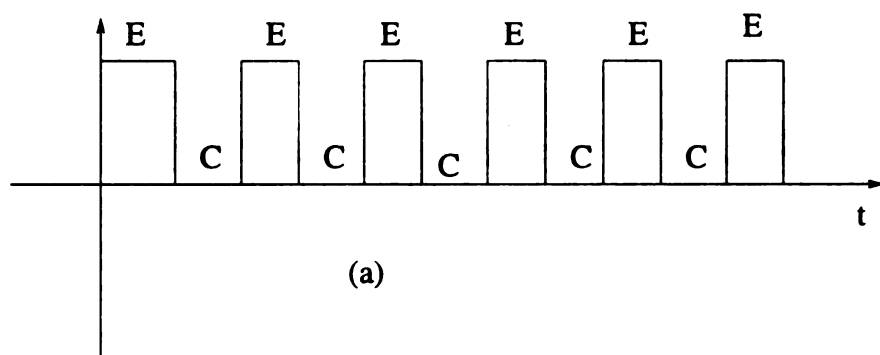


Figure 1.3. Illustration of fMRI stimulus and responses. a) Controlled stimulus; b) Response from an activated area; c) Response from an inactivated area.

# CHAPTER 2

## Pixel-wise Detection for fMRI

*A tree broader than a man can embrace is born of a tiny shoot;  
A dam greater than a river can overflow starts with a clod of earth;  
A journey of a thousand miles begins with a single step.*

Lao Tzu (500 BC, China), Tao Te Ching

### 2.1 Overview

In fMRI a series of MR images of the brain are acquired over time to detect neural activity. As explained in Section (1.1), the BOLD effect can be used to obtain maps of active and non-active regions of the brain. In order to achieve high signal noise ratio, the spatial *and* temporal imaging resolution must be limited [34]. Unfortunately, low resolution imaging may lead to a loss in signal information originating from microvasculature [60]. Hence, there is a fundamental trade-off between resolution and SNR in fMRI. It is therefore of great interest to develop reliable detection methods for fMRI in the presence of noise.

Almost all fMRI tests are based on the image magnitude data. In standard practice, the raw MRI data is reconstructed and the magnitude is taken to eliminate the (unknown) phase. I focus in this chapter on the repetitive experiments involving

a dynamic time sequence of images. Under various assumptions and experimental setups the fMRI detection problem reduces to well-known statistical tests including  $t$ -tests and  $F$ -tests [69].

When comparing two groups of images — “rest state” and “active state” images — a  $t$ -test is usually used [14], the derivation of which is at Appendix IV of Chapter 5. Another common approach to fMRI detection, called magnitude correlation (MC) detection in this dissertation, is based on a test statistic obtained at each pixel by correlating the magnitude time-series with a reference signal, which is assumed to be known and representative of the BOLD response [3]. Many generalizations and extensions of this simple idea have been proposed [24, 25, 36, 55, 58, 60, 64]. For example, recently Lai and Glover proposed a complex correlation (CC) test based on the complex data (*i.e.*, image data before taking the magnitude at each pixel), in order to take advantage of phase information in the data and improve the detectability of fMRI responses [36].

The CC test statistic is  $F$ -distributed and the detector has a *constant false-alarm rate* (CFAR) property, which means that a specified false-alarm rate, *i.e.*, the probability of deciding a pixel is active when in fact it is not, can be achieved irrespective of the unknown parameters. *Throughout this dissertation, the false-alarm rate is denoted by  $P_f$  and the probability of detection is denoted by  $P_d$ .* Despite the CFAR property, the CC test focuses only on the response component (called *signal component*) of the data and ignores the baseline component (*nuisance component* in a general setting) of the data. The baseline component does not contain information relevant to the response itself (and is hence called *nuisance component*), but it does contain important information about the phase. In this chapter, a new test is proposed based on the GLRT principle that allows us to incorporate the phase information contained in the baseline component.

In this chapter simple pixel-wise testing is stressed based on a Gaussian white

noise observation model. Pixel-wise testing ignores spatial relationships in fMRI data, which is to be considered in Chapter 3. However, since the focus of this chapter is to assess the potential benefits of fMRI detection using complex data, a simple data model and testing procedure is employed to explore this basic issue. The assumptions are perhaps too simplistic in many practical cases. However, it is possible to extend the results and conclusions to more elaborate approaches based on more realistic data and/or correlated noise models, potentially accounting for uncertainties in the BOLD response and/or different nuisance component. Such extensions are briefly discussed in the conclusions of Chapter 4.

This chapter is organized as follows. In Section (2.2), a basic model for fMRI data is reviewed. Three tests will be studied, all of which may be interpreted as GLRTs under different data model assumptions. Therefore, before looking at each method, the GLRT principle is briefly reviewed in Section (2.3). In Section (2.4) and Section (2.5), the standard MC and recently-proposed CC tests are examined and the statistical properties of each are studied. The new GLRT for fMRI is derived in Section (2.6). Its properties are explored in detail. In Section (2.7), the performance of the MC test, CC test, and GLRT in various regimes of baseline signal intensity to noise ratio are compared. Extensive Monte Carlo simulation is used to assess the performance of the detectors. The results show that *the GLRT does have a CFAR property and a simple rule for choosing the threshold to achieve a desired  $P_f$  is observed*. The distribution of the GLRT statistic at high ratio of baseline signal intensity to noise is approximated from observation on threshold selection. Numerical studies show that the GLRT outperforms the CC test. Specifically, for a fixed false-alarm rate  $P_f$ , the GLRT's detection rate  $P_d$  is higher than that of the CC test. Furthermore, the GLRT also performs significantly better than the MC test at low baseline signal intensity. The performances of the GLRT and MC detectors are roughly the same at high baseline signal intensity, and in such situations both

perform better than the CC detector. In Section (2.8), the performance of all three detectors is demonstrated in a simulated fMRI experiment.

Section (2.9) contains the three detectors for the same model but with known noise variance, which is a by-product of my research.

## 2.2 fMRI Signal Model

Due to phase errors which are difficult to control, the signal component of the measurements occurs in both real and imaginary channels [41, 6]. This suggests the following simple model for an fMRI pixel time-series. Suppose there are  $N$  images acquired in the experiment. Let  $\mathbf{x}$  denote an  $N \times 1$  vector containing the time-series data from one pixel:

$$\mathbf{x} = (a\mathbf{1} + b\mathbf{r})e^{i\theta} + \sigma\mathbf{n}_c. \quad (2.1)$$

The data vector  $\mathbf{x}$  consists of three complex-valued components. The first component  $a\mathbf{1}$  is a constant (DC) *baseline component*, where  $\mathbf{1}$  denotes an  $N \times 1$  vector of ones and  $a > 0$  is the amplitude of the DC component. This vector represents the average value of the time-series. The baseline component model proposed here is the simplest version of the *nuisance component*. The second component  $b\mathbf{r}$  is the *oscillatory response (signal) component*. The vector  $\mathbf{r}$  is a reference function that models the expected response characteristic. The amplitude  $b$  characterizes the strength of the response. In the absence of activity,  $b = 0$ .

Typically, in fMRI studies, while the subject is under some baseline condition (for example, at rest), a number of frames  $N_b$  is acquired; then the subject is asked to perform some task (for example, finger tapping) and a number of activation frames  $N_a$  is acquired, or vice versa. These constitute one cycle. During each cycle, the total

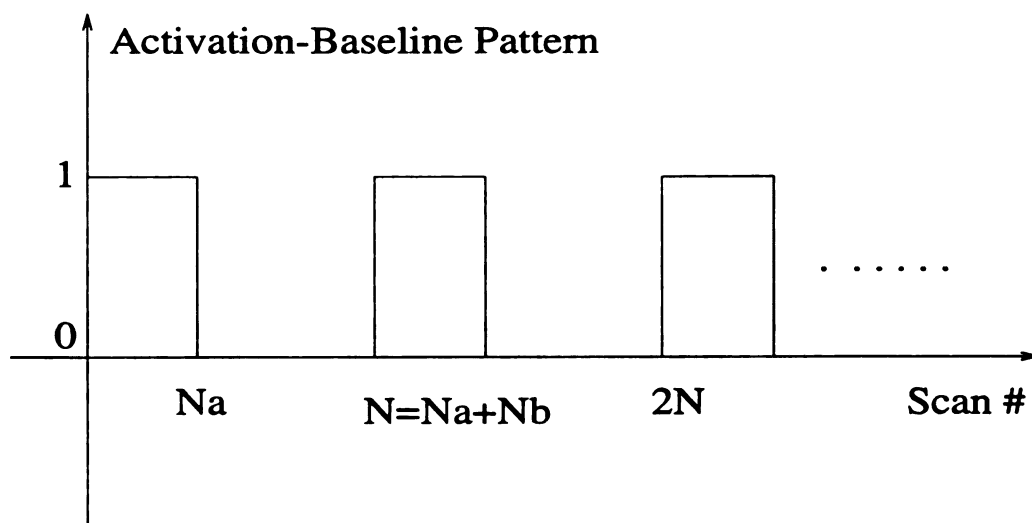
---

This model is attributed to Dr. Nowak.

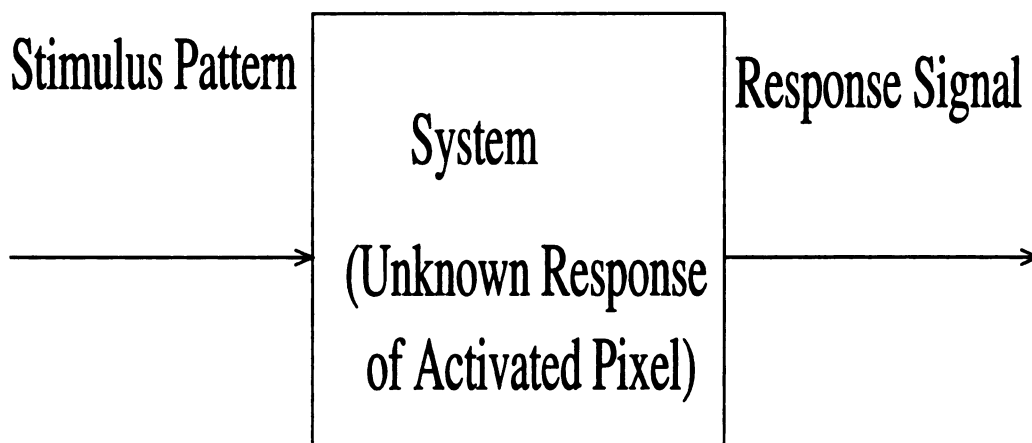
number of frames is thus  $N = N_b + N_a$ . This pattern is repeated for a number of cycles. In Figure (2.1) adapted from [2], the activation-baseline pattern is represented by a periodic rectangular waveform of period  $N$ , with 1 and 0 representing activation and baseline, respectively [2].

One can think of the signal component as the response of a system whose input is the activation-baseline pattern. See Figure (2.1b). In real problems, the response signal  $\mathbf{r}$  (the output of the system in Figure (2.1b)) is unknown. Friston *et al.* modeled the system as a linear time invariant (LTI) system [22], which is questioned by [2, 3]. Several possible estimates for the reference signal are suggested in [2, 3]. The first method is to use a delayed version of the activation-baseline pattern. It is most easily implemented. The disadvantage of this method is that the delay is not known *a priori*, and it may vary from pixel to pixel. The second suggestion is to select the response of one or more activated pixels as the reference signal. The third option is to average the response of some activated pixels across cycles. The reference signal would then be formed by periodically replicating this time-averaged cycle throughout the time course. None of these approaches is perfect. Throughout this dissertation, the reference signal  $\mathbf{r}$  is assumed to be known.

The baseline component and signal component share a common phase  $\vartheta$ . Hence, we model this *phase-coupling* by multiplying both components by the complex number  $e^{i\vartheta}$ , where  $i = \sqrt{-1}$ . In addition to these two components, an additive complex Gaussian white noise component  $\sigma \mathbf{n}_c$  models errors primarily due to thermal noises in the patient [13, 18, 19, 44]. The term  $\mathbf{n}_c$  denotes a standard (zero mean, unit variance) complex Gaussian vector of length  $N$ . The factor  $\sigma$  scales the noise resulting in a variance of  $\sigma^2$ . In general, the parameters of this model  $a$ ,  $b$ ,  $\vartheta$ , and  $\sigma$  are unknown and are different for each pixel time-series. This model was compared to actual fMRI time-series and these assumptions are in good agreement with actual data. Figure (2.2) shows the real part, imaginary part and phase of one time series from real fMRI



(a)



(b)

Figure 2.1. (a) Activation-baseline pattern for an fMRI experiment; (b) Modeling the activated pixel as a system.

experiment and it is illustrative of the constant phase idea in the model.

The modeling of fMRI signals is a very complicated process [24, 29, 38]. Our model is still rough and not complete. The noise *structure* in fMRI is very complicated. Our model does not try to capture more complicated disturbances present in fMRI data such as other nuisance components, for example, due to physiologic motions. For the sake of simplicity and demonstrating our method and ideas, the noise is assumed to be *white Gaussian*. The *whiteness* assumption does not change the problem essentially, since we can always use Cholesky factorization [56] to whiten the signal and leave the detection problem unchanged, although, of course, estimation of the covariance of the noise is a challenging problem in itself. Usually parametric tests, to which discussions in this chapter are confined, assume a *Gaussian* model for the underlying time series. Several researchers have challenged this assumption and have used nonparametric tests, including the Kolmogorov-Smirnov test, Kruskal-Wallis test, and Wilcoxon signed ranks test, for a summary, see [69].

I succeed in deriving the GLRT directly from this complex model in Equation (2.1):

$$L_0 = \frac{\mathbf{x}^H \mathbf{x} - N[\alpha_1^2 + \alpha_2^2]}{\mathbf{x}^H \mathbf{x} - \frac{N}{2}[\alpha_1^2 + \alpha_2^2 + \beta_1^2 + \beta_2^2 + \sqrt{(2\alpha_1\alpha_2 + 2\beta_1\beta_2)^2 + (\alpha_1^2 + \beta_1^2 - \alpha_2^2 - \beta_2^2)^2}]},$$

where  $^H$  denotes complex conjugate transpose and

$$\begin{aligned}\alpha_1 &= \frac{\mathbf{x}_R^T \mathbf{1}}{\sqrt{N}}, \\ \beta_1 &= \frac{\mathbf{x}_R^T \mathbf{r}}{\sqrt{N}}, \\ \alpha_2 &= \frac{\mathbf{x}_I^T \mathbf{1}}{\sqrt{N}}, \\ \beta_2 &= \frac{\mathbf{x}_I^T \mathbf{r}}{\sqrt{N}}.\end{aligned}$$

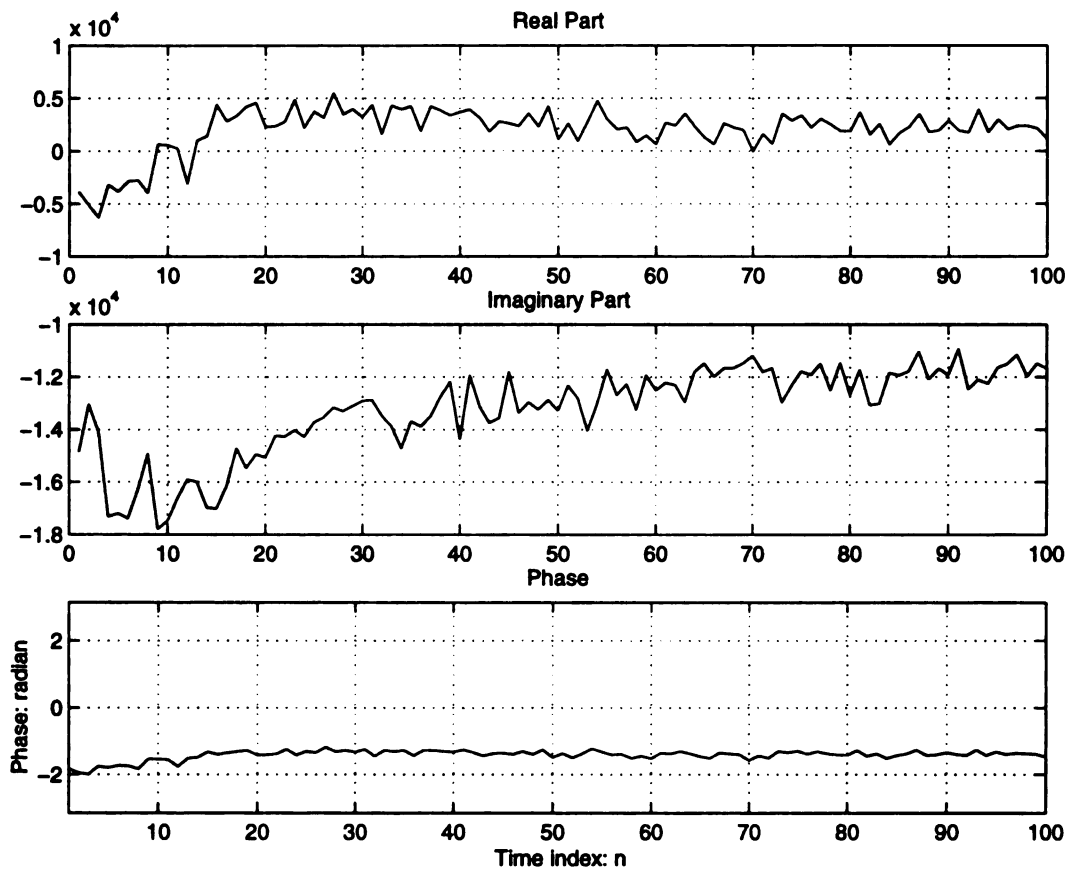


Figure 2.2. One time series from a real fMRI experiment. (a) Real part; (b) Imaginary part; (c) Phase.

This form, however, is not well suited for mathematical analysis. I thus turned to analysis in real domain. Because complex numbers can be interpreted as pairs of real numbers, I therefore re-express the complex model (2.1) as a  $2N \times 1$  dimensional real-valued model:

$$\mathbf{y} = \mathbf{S}\boldsymbol{\phi} + \mu\mathbf{H}\boldsymbol{\phi} + \sigma\mathbf{n}, \quad (2.2)$$

where  $\mu \equiv b/a$  and

$$\mathbf{y} = \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{r} & \mathbf{0} \\ \mathbf{0} & \mathbf{r} \end{bmatrix}, \quad \boldsymbol{\phi} = \begin{bmatrix} a \cos \vartheta \\ a \sin \vartheta \end{bmatrix}, \quad \mathbf{n} = \begin{bmatrix} \mathbf{n}_{cR} \\ \mathbf{n}_{cI} \end{bmatrix}.$$

The subscripts  $R$  and  $I$  denote real and imaginary parts, respectively. The phase-coupling in the complex model is manifest in this real model as a nonlinear coupling between parameter  $\mu$  and parameter  $\boldsymbol{\phi}$ . The reader is reminded here that  $\mu$  is the ratio of reference signal intensity to baseline signal intensity.

Note here that this nonlinear model stands in marked contrast to the classical linear regression model:

$$\mathbf{y} = \mathbf{S}\boldsymbol{\phi}_1 + \mu\mathbf{H}\boldsymbol{\phi}_2 + \sigma\mathbf{n}, \quad (2.3)$$

where  $\boldsymbol{\phi}_1$  and  $\boldsymbol{\phi}_2$  are independent. In the following it is shown that the CC test of Lai and Glover [36] can be derived from the linear model above. It is our contention that the nonlinear model is a more accurate representation of the physical fMRI problem, and, indeed, the new GLRT based on the nonlinear model outperforms the CC test.

## 2.3 Generalized Likelihood Ratio Tests

The likelihood ratio test (LRT) [33] is an optimal method for deciding which of two hypotheses (competing data models) best describes a set of observed data. The data model corresponding to each hypothesis is given by a probability density function

(pdf). Unfortunately, however, to implement the LRT, the pdf's under each hypothesis must be completely specified. The corresponding test is called *simple hypothesis test* problem. This is not the case in fMRI. In the fMRI case, we have two hypotheses:  $H_0$ , BOLD response absent ( $\mu = 0$ ), and  $H_1$ , BOLD response present ( $\mu \neq 0$ ). Under hypothesis  $H_0$ , the vector  $\phi$  and the noise power  $\sigma^2$  are unknown. Under hypothesis  $H_1$ ,  $\phi$ ,  $\sigma^2$  and  $\mu$  are unknown. Due to the unknowns, in fMRI we have what is called a *composite* hypothesis test.

There are two standard approaches to composite hypothesis testing. The *Bayesian* approach prescribes *a priori* pdf's for the unknown parameters themselves; and the likelihoods are integrated against these pdf's to eliminate the dependence of the LRT on the unknown parameters. Another approach, the *generalized likelihood ratio test* (GLRT) is often preferable to Bayesian approaches due to its ease of implementation and less restrictive assumptions. Specifically, the GLRT does not require the specifications of *a priori* probability distributions for the unknown parameters [33, 56]. For these reasons, this chapter focuses on the GLRT.

The idea of GLRT comes from the LRT used in simple hypothesis testing which means the pdf for each assumed hypothesis is completely known. Let  $p_{H_i}(\mathbf{x}; \Theta_i)$ ,  $i = 0, 1$ , denote the pdf's corresponding to the two hypotheses. Recall that  $\mathbf{x}$  denotes the data. The argument  $\Theta_i$  denotes *known* parameters that specify the precise form of the pdf. For example,  $\Theta_i$  may represent the mean vector and covariance matrix of a multivariate Gaussian density. The LRT decides  $H_1$  if

$$L(\mathbf{x}) = \frac{p_{H_1}(\mathbf{x}; \Theta_1)}{p_{H_0}(\mathbf{x}; \Theta_0)} > \eta,$$

where  $\eta$  is the threshold, which can be chosen to achieve a desired  $P_f$ . The likelihood ratio (LR)  $L(\mathbf{x})$ , a function of the data  $\mathbf{x}$ , is called the *test statistic*.

The GLRT is also based on the LR, but in the composite case the parameters are

unknown. The key idea in the GLRT is to replace the unknown parameters by their maximum likelihood estimates (MLE's). In general, the GLRT decides  $H_1$  if

$$L(\mathbf{x}) = \frac{p_{H_1}(\mathbf{x}; \hat{\Theta}_1)}{p_{H_0}(\mathbf{x}; \hat{\Theta}_0)} > \eta,$$

where  $\hat{\Theta}_1$  is the MLE of  $\Theta_1$  assuming  $H_1$  is true, and  $\hat{\Theta}_0$  is the MLE of  $\Theta_0$  assuming  $H_0$  is true. The MLE of a parameter is simply the value of the parameter that maximizes the corresponding pdf (*i.e.*, the value that makes the observed data most likely).

The GLRT has no optimality property, in general, but it asymptotically approaches the uniformly most powerful (UMP) test among invariant tests [7]. For more details on maximum likelihood estimation and the GLRT, see [33]. In the following sections we review the basic MC and CC tests and introduce the new nonlinear test for fMRI detection using the GLRT.

## 2.4 Method 1: Magnitude Correlation Detection

The magnitude of MRI data is known to be Rician distributed [41]. To see this, note that  $x_j$  in Equation (2.1), the  $j$ th observation in the time series can be written as

$$x_j = (a + br_j) \cos \theta + \sigma n_{Rj} + i[(a + br_j) \sin \theta + \sigma n_{Ij}].$$

The two terms being independent,  $z_j \equiv |x_j|$  is Rician distributed (see Definition (3) and Equation (5.1) in Appendix III). However, for large values of ratio  $a/\sigma$  (ratio of baseline component intensity to noise standard deviation) the Rician density can be well approximated as a Gaussian distribution because

$$\begin{aligned} z_j \equiv |x_j| &= \sqrt{[(a + br_j) \cos \theta + \sigma n_{Rj}]^2 + [(a + br_j) \sin \theta + \sigma n_{Ij}]^2} \\ &= \sqrt{(a + br_j)^2 + \sigma^2(n_{Rj}^2 + n_{Ij}^2) + 2(a + br_j)\sigma(n_{Rj} \cos \theta + n_{Ij} \sin \theta)} \end{aligned}$$

$$= (a + br_j) \sqrt{1 + \frac{2\sigma(n_{Rj} \cos \theta + n_{Ij} \sin \theta)}{a + br_j} + \frac{\sigma^2}{(a + br_j)^2} (n_{Rj}^2 + n_{Ij}^2)}.$$

Note that  $n_{Rj} \cos \theta + n_{Ij} \sin \theta$  is nothing but another Gaussian random variable. We denote it as  $n_j$ . Also note that  $n_j$  and  $n_k$  are independent for  $j \neq k$ .  $n_{Rj}^2 + n_{Ij}^2$  is a  $\chi_2^2$  random variable. Under the assumptions that  $a \gg \sigma$  and  $\mu = b/a$  is very small, the third term under the square root sign is much smaller than the second one and therefore can be neglected. In this case, application of the binomial expansion

$$\sqrt{1+x} \approx 1 + \frac{1}{2}x \quad |x| \ll 1,$$

to the above equation leads to

$$z_j \approx a + br_j + \sigma n_j.$$

Hence, the following Gaussian approximationa is commonly assumed for fMRI detection [3]:

$$\mathbf{z} \approx a\mathbf{1} + b\mathbf{r} + \sigma\mathbf{n}, \quad (2.4)$$

where here  $\mathbf{z} = |\mathbf{x}|$ ,  $\mathbf{n}$  is (real) Gaussian distributed, and with  $b = 0$  under  $H_0$  and  $b \neq 0$  under  $H_1$ . Hence, in this case  $\Theta_0 = [a \ \sigma^2]$  and  $\Theta_1 = [a \ b \ \sigma^2]$ . Bear in mind that this approximation does not accurately model the data when  $a/\sigma$  is small, as we shall see later in some examples.

The GLRT for the above detection problem results in the following test statistic [56, 57]:

$$t_1(\mathbf{z}) = (N-1)[L_1(\mathbf{z}) - 1] = (N-1) \frac{\|P_{\mathbf{r}} P_{\mathbf{1}}^\perp \mathbf{z}\|^2}{\|P_{\mathbf{r}}^\perp P_{\mathbf{1}}^\perp \mathbf{z}\|^2}, \quad (2.5)$$

where

$$L_1(\mathbf{z}) = \frac{\|P_1^\perp \mathbf{z}\|^2}{\|P_{1\mathbf{r}}^\perp \mathbf{z}\|^2}.$$

If  $t_1(\mathbf{z}) > \eta_1$ , then we decide  $H_1$ ; otherwise choose  $H_0$ . On the assumption that  $\mathbf{z}$  is truly Gaussian,  $t_1(\mathbf{z})$  is distributed as  $F_{1,(N-1)}(\text{SNR})$  [57], where  $\text{SNR} \equiv \mu^2 a^2 / \sigma^2$ . Refer to Appendix II in Chapter 5. The detector has CFAR property <sup>†</sup> Therefore, we can choose a threshold  $\eta_1$  to achieve a desired  $P_f$  irrespective of the signal to noise ratio, which is generally unknown *a priori*. This detector is called the magnitude correlation (MC) detector because the test statistic  $t_1(\mathbf{z})$  is proportional to the correlation between the magnitude data  $\mathbf{z}$  and the reference signal  $\mathbf{r}$ .

Unfortunately, the Gaussian approximation in Equation (2.4) is unreasonable when  $a/\sigma \leq 3$ . In fact, when  $a/\sigma$  is small, the distribution of test statistic  $t_1(\mathbf{z})$  is not known, nor whether the MC detector has CFAR property. So determination of a proper threshold to obtain a desired  $P_f$  is theoretically very difficult. How to solve this problem will be explained together with numerical results. Moreover, in this case, one expects the performance of the MC test to suffer. This is indeed the case as shown later by numerical results.

## 2.5 Method 2: Complex Correlation Detection

Recently, Lai and Glover proposed a *complex correlation* (CC) test based on the complex data, in order to take advantage of phase information in the data and improve the detectability of fMRI responses [36].

Here, the CC test statistic is shown to be also  $F$ -distributed and also has a CFAR

---

<sup>†</sup>Note that the MC test statistic has a central  $F_{1,N-1}$  distribution under  $H_0$  ( $\mu = 0$ ).

property. Recall the linear model

$$\mathbf{y} = S\phi_1 + \mu H\phi_2 + \sigma \mathbf{n}. \quad (2.6)$$

The unknown parameters in this case are  $\Theta_0 = [\phi_1^T \ \phi_2^T \ \sigma^2]$  and  $\Theta_1 = [\phi_1^T \ \phi_2^T \ \mu \ \sigma^2]$ . The GLRT based on this model in fact coincides with the CC test [36] and is given by

$$t_2(\mathbf{y}) = (N - 1)[L_2(\mathbf{y}) - 1] = (N - 1) \frac{\|P_H P_S^\perp \mathbf{y}\|^2}{\|P_{SH}^\perp P_S^\perp \mathbf{y}\|^2} = (N - 1) \frac{\|P_H P_S^\perp \mathbf{y}\|^2}{\|P_H^\perp P_S^\perp \mathbf{y}\|^2} \quad (2.7)$$

where

$$L_2(\mathbf{y}) = \frac{\|P_S^\perp \mathbf{y}\|^2}{\|\mathbf{y}\|^2 - \|P_S \mathbf{y}\|^2 - \|P_H \mathbf{y}\|^2} = \frac{\mathbf{y}^T P_S^\perp \mathbf{y}}{\mathbf{y}^T P_{SH}^\perp \mathbf{y}}. \quad (2.8)$$

If  $t_2(\mathbf{y}) > \eta_2$ , then we decide  $H_1$ ; otherwise, choose  $H_0$ . This test is called the complex correlation (CC) test because it is based on the correlation between reference signal and real and imaginary components of the complex data. The pdf of  $t_2(\mathbf{y})$  is non-central  $F_{2,2(N-1)}(\text{SNR})$ , where again  $\text{SNR} = \mu^2 a^2 / \sigma^2$ , and thus the detector has the CFAR property <sup>†</sup>

Despite this desirable property, the drawback to this test is that it is based on a less accurate model. The CC test focuses only on the response component of the data and ignores the baseline component. The DC component does not contain information relevant to the response itself, but it does contain important information about the phase. Although the phase is a *nuisance* parameter in the testing problem, more accurate knowledge of the phase can improve the detectability of the fMRI response. As noted previously, the phase-coupling between the nuisance and signal response components of the data dictates the nonlinear model in Equation (2.2). Therefore, I next derive a new GLRT based on this more accurate model.

---

<sup>†</sup>Note that the CC test statistic has a central  $F_{2,2(N-1)}$  distribution under  $H_0$  ( $\mu = 0$ ).

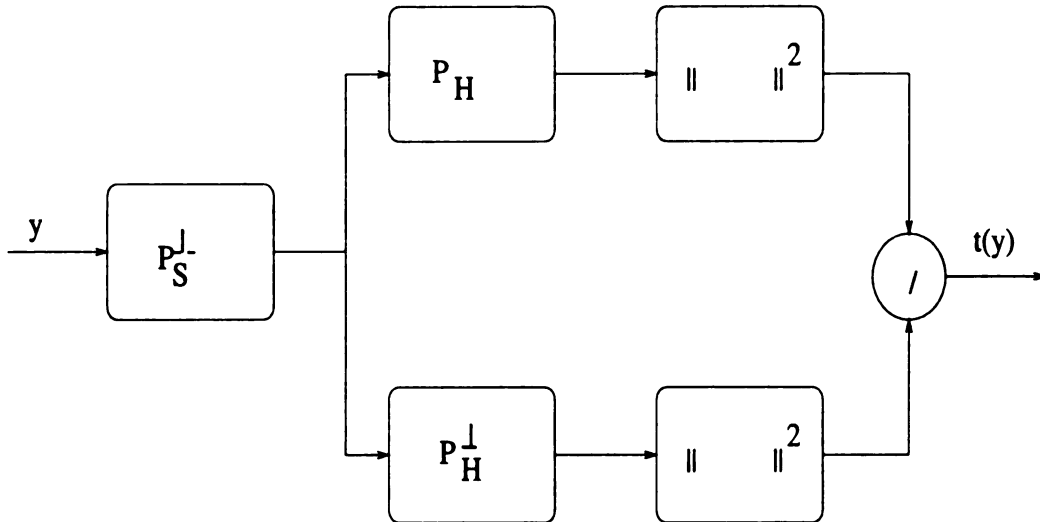


Figure 2.3. Structure of matched subspace detector.

Before going to next section, I'd like to summarize a little bit. The MC detector and CC detector actually have the same structure as shown in Figure (2.3) adapted from [57], which is known in the signal processing literature as a matched subspace detector [57]. First the data are projected onto a low-rank subspace by removing interference. The projector is also termed an *interference rejecting* or *null steering filter* [57]. Then the resulting data are further projected onto another low-rank subspace that is matched to the signal component, and energy is taken. This projector is usually called a *matched subspace filter*. Since the noise level is unknown, this energy is then compared with the energy in the component orthogonal to signal subspace. The ratio is computed and compared with a threshold for a decision [57].

## 2.6 Method 3: Nonlinear GLR Detection

In this section, a new nonlinear test based on the GLRT principle is found that incorporates the phase information contained in the baseline component [48]. The unknown parameters in model (2.2) are  $\Theta_0 = [\phi^T \ \sigma^2]$  and  $\Theta_1 = [\phi^T \ \mu \ \sigma^2]$ , under  $H_0$

and  $H_1$ , respectively. Recall that the phase coupling introduces a nonlinear coupling between the parameters  $\mu$  and  $\phi$  under  $H_1$ . This nonlinearity makes the MLE's very difficult to compute, but, remarkably, a closed-form solution for the GLRT statistic which is derived in the following subsection does exist:

$$t_3(\mathbf{y}) = [L_3(\mathbf{y}) - 1](N - 1), \quad (2.9)$$

where

$$L_3(\mathbf{y}) = \frac{2\|P_S^\perp \mathbf{y}\|^2}{\|P_H^\perp \mathbf{y}\|^2 + \|P_S^\perp \mathbf{y}\|^2 - \sqrt{\|P_H \mathbf{y}\|^4 + \|P_S \mathbf{y}\|^4 + 2\|P_H \mathbf{y}\|^2 \|P_S \mathbf{y}\|^2 \cos 2\varphi}} \quad (2.10)$$

with

$$\cos \varphi(\mathbf{y}) = \frac{(\theta_1, \theta_2)}{\|\theta_1\| \|\theta_2\|} = \frac{\mathbf{y}^T \frac{HS^T}{N} \mathbf{y}}{\|P_H \mathbf{y}\| \|P_S \mathbf{y}\|} \quad (2.11)$$

and  $\theta_1$  and  $\theta_2$  are two sufficient statistics

$$\theta_1(\mathbf{y}) = \frac{1}{N} H^T \mathbf{y}, \quad \text{and} \quad \theta_2(\mathbf{y}) = \frac{1}{N} S^T \mathbf{y}. \quad (2.12)$$

As usual, given a specified threshold  $\eta_3$ , we decide  $H_1$  if  $t_3(\mathbf{y}) > \eta_3$ , and  $H_0$  otherwise.

### 2.6.1 Derivation of the GLRT Test Statistic

In my derivations, I assume  $\mathbf{1}^T \mathbf{r} = 0$  and  $\mathbf{r}^T \mathbf{r} = N$  without loss of generality, so  $\mathbf{S}^T \mathbf{H} = \mathbf{H}^T \mathbf{S} = \mathbf{0}_{2 \times 2}$ ,  $\mathbf{S}^T \mathbf{S} = \mathbf{H}^T \mathbf{H} = N \mathbf{I}_{2 \times 2}$ . The second condition can always be satisfied since we can always normalize the reference signal without changing the problem essentially. The first condition is more difficult to meet. But the same ideas of decomposing one orthogonal projection operator into two oblique projection operators as in [4, 57] can be utilized to achieve the result, even under more compli-

cated nuisance and signal (reference) structures, *i.e.*, the nuisance component is not just a constant baseline  $\mathbf{1}$ , nor is the signal component just one reference signal  $\mathbf{r}$ . Possibly, both of them may be adaptively selected from real data to contain several components (see discussions in Chapter 4). However, the analysis of the test statistic's properties (such as invariance), and hence the determination of the threshold are much more complicated than presented in this dissertation. Even so, the properties of the corresponding linear detector may still be used as a guide.

Let  $\hat{\sigma}_0$  and  $\hat{\sigma}_1$  denote the MLEs of noise variance under hypotheses  $H_0$  and  $H_1$ , respectively. Recall that the GLRT is based on  $\frac{p_{H_1}(\mathbf{x}; \hat{\Theta}_1)}{p_{H_0}(\mathbf{x}; \hat{\Theta}_0)}$ . It is easy to show that this statistic reduces to

$$L_3(\mathbf{y}) = \min(\hat{\sigma}_0^2) / \min(\hat{\sigma}_1^2). \quad (2.13)$$

Recall that under Gaussian distribution the maximum likelihood estimate is the same as the least square estimate. Therefore, the calculation of  $\min \hat{\sigma}_0^2$  is straightforward:

$$\min \hat{\sigma}_0^2 = \| P_S^\perp \mathbf{y} \|^2 \quad (2.14)$$

However, determining  $\min(\hat{\sigma}_1^2)$  is much more difficult due to the nonlinear coupling between the two unknowns  $\mu$  and  $\phi$ . To circumvent this difficulty, I first decompose  $\mathbf{y}$  and  $\mathbf{y} - \mathbf{S}\phi - \mu\mathbf{H}\phi$  into three orthogonal components, *i.e.*,

$$\mathbf{y} = P_{SH}^\perp \mathbf{y} + P_S \mathbf{y} + P_H \mathbf{y},$$

and then

$$\begin{aligned} \hat{\sigma}_1^2 &= \| \mathbf{y} - \mathbf{S}\phi - \mu\mathbf{H}\phi \|^2 \\ &= \| P_{SH}^\perp \mathbf{y} + (P_S \mathbf{y} - \mathbf{S}\phi) + (P_H \mathbf{y} - \mu\mathbf{H}\phi) \|^2 \\ &= \| P_{SH}^\perp \mathbf{y} \|^2 + \| P_S \mathbf{y} - \mathbf{S}\phi \|^2 + \| P_H \mathbf{y} - \mu\mathbf{H}\phi \|^2 \end{aligned}$$

$$\begin{aligned}
&= \| P_{SH}^\perp \mathbf{y} \|^2 + \| \mathbf{S}\theta_2 - \mathbf{S}\phi \|^2 + \| \mathbf{H}\theta_1 - \mu\mathbf{H}\phi \|^2 \\
&= \| P_{SH}^\perp \mathbf{y} \|^2 + N[\theta_1^T \theta_1 + \theta_2^T \theta_2 + (1 + \mu^2)\phi^T \phi - 2(\theta_2 + \mu\theta_1)^T \phi]
\end{aligned}$$

where  $\theta_1$  and  $\theta_2$  are given by Equation (2.12).

Now  $\min \hat{\sigma}_1^2$  is equivalent to  $\min J = (1 + \mu^2)\phi^T \phi - 2(\theta_2 + \mu\theta_1)^T \phi$ . Setting partial derivatives of  $J$  with respect to  $\mu$  and  $\phi$  to zero results in:

$$\begin{cases} \frac{\partial J}{\partial \mu} = 2\mu\phi^T \phi - 2\theta_1^T \phi = 0, \\ \frac{\partial J}{\partial \phi} = 2(1 + \mu^2)\phi - 2(\theta_2 + \mu\theta_1) = 0. \end{cases}$$

We then get

$$\hat{\mu} = \frac{\theta_1^T \hat{\phi}}{\hat{\phi}^T \hat{\phi}}, \quad (2.15)$$

$$\hat{\phi} = \frac{\theta_2 + \hat{\mu}\theta_1}{1 + \hat{\mu}^2}. \quad (2.16)$$

So now

$$\min \hat{\sigma}_1^2 = \| P_{SH}^\perp \mathbf{y} \|^2 + N(\theta_1^T \theta_1 + \theta_2^T \theta_2) - N(1 + \hat{\mu}^2)\hat{\phi}^T \hat{\phi}.$$

Furthermore, note that

$$N(\theta_1^T \theta_1 + \theta_2^T \theta_2) = \| P_H \mathbf{y} \|^2 + \| P_S \mathbf{y} \|^2, \text{ and } \| P_{SH}^\perp \mathbf{y} \|^2 + \| P_S \mathbf{y} \|^2 + \| P_H \mathbf{y} \|^2 = \| \mathbf{y} \|^2,$$

which gives

$$\min \hat{\sigma}_1^2 = \| \mathbf{y} \|^2 - N(1 + \hat{\mu}^2)\hat{\phi}^T \hat{\phi} = \| \mathbf{y} \|^2 - \frac{N}{1 + \hat{\mu}^2}[\theta_2^T \theta_2 + 2\hat{\mu}\theta_1^T \theta_2 + \hat{\mu}^2 \theta_1^T \theta_1].$$

Eliminating  $\hat{\phi}$  from Equations (2.15) and (2.16) (or setting the derivative of the

above equation with respect to  $\hat{\mu}$  to zero) shows that  $\hat{\mu}$  must satisfy:

$$\frac{\theta_2^T \theta_2 + 2\hat{\mu}\theta_1^T \theta_2 + \hat{\mu}^2 \theta_1^T \theta_1}{1 + \hat{\mu}^2} = \frac{\theta_1^T \theta_2 + \hat{\mu}\theta_1^T \theta_1}{\hat{\mu}}. \quad (2.17)$$

Using this equation,  $\min \hat{\sigma}_1^2$  can be further simplified:

$$\min \hat{\sigma}_1^2 = \|\mathbf{y}\|^2 - N(\theta_1^T \theta_1 + \frac{\theta_1^T \theta_2}{\hat{\mu}}) \quad (2.18)$$

This equation is important in our derivation of an asymptotic expression for  $L_3(\mathbf{y})$  in the following subsection.

From Equation (2.17),  $\hat{\mu}$  satisfies quadratic equation  $\mu^2 + c\mu - 1 = 0$  with

$$c = \frac{\theta_2^T \theta_2 - \theta_1^T \theta_1}{\theta_1^T \theta_2}.$$

Since  $\mu_1 \mu_2 = -1$ , there are two solutions of opposite signs:

$$\hat{\mu} = -\frac{c}{2} \pm \sqrt{1 + \left(\frac{c}{2}\right)^2}.$$

However, from Equation (2.18), to make sure  $\hat{\sigma}_1^2$  is minimal,  $\hat{\mu}$  must have the same sign as  $\theta_1^T \theta_2$ , and so the unique solution for  $\hat{\mu}$  is:

$$\hat{\mu} = -\frac{c}{2} + \sqrt{1 + \left(\frac{c}{2}\right)^2}.$$

Substitution of  $\hat{\mu}$  into Equation (2.18) yields the right solution for  $\min \hat{\sigma}_1^2$ . Finally, from Equations (2.13), (2.14) and (2.18) I get the closed form expression for  $L_3(\mathbf{y})$  as given by Equation (2.10).

Instead of using  $L_3(\mathbf{y})$  directly, I use Equation (2.9) as the test statistic, which is suggested by the form of test statistic  $t_2(\mathbf{y})$  for the CC detector. The main reasons are

to facilitate the determination of the proper threshold and to get a good comparison between the three different detectors. It will become much clearer when we study the asymptotic property of  $t_3(\mathbf{y})$ .

Having established the form of the test statistic, the next question is naturally how to choose the threshold, which is a very difficult problem. In order to answer this question we need know the pdf of this statistic. Unfortunately, unlike the CC test, a closed form for the pdf of  $t_3(\mathbf{y})$  is not known to me at this stage. I can, however, show that the pdf of  $t_3(\mathbf{y})$  is a function of  $\mu$  and  $a^2/\sigma^2$  alone.

### 2.6.2 Invariance of GLRT Test Statistics

In this subsection, I employ theories on invariance [46, 56] to prove that the pdf of  $t_3(\mathbf{y})$  is a function of only  $\mu$  and  $a/\sigma$ . In order not to interrupt the continuity of description, the ideas and principles on invariant test are put in Appendix V of Chapter 5.

The difficult part in using invariant theory is to find an appropriate set of transformations which fully exploit the structure of the signal to be detected. Since our problem is nonlinear, finding this set of transformation is not an easy matter. Actually discovery of this set of transformations comes simultaneously with its proof. The following theorem may be regarded as an extension of the results in [56, 57], which only deal with the linear model of Equation (2.3).

**Theorem:** *The family of distributions of  $\mathbf{y}$  defined by*

$$\mathbf{y} = \mu \mathbf{H}\boldsymbol{\phi} + \mathbf{S}\boldsymbol{\phi} + \sigma \mathbf{n} \quad (2.19)$$

*where  $\mathbf{n}$  is Gaussian distributed as  $N(0, \mathbf{I})$ , is invariant to the group of transformations defined by:*

$$G = \{g(\mathbf{y}) : g(\mathbf{y}) = cQ_S Q_H \mathbf{y}\} \quad (2.20)$$

where the two orthogonal matrices

$$\begin{cases} Q_S &= U_S Q U_S^T + P_S^\perp &= \frac{S}{\sqrt{N}} Q \frac{S^T}{\sqrt{N}} + I - \frac{SS^T}{N}, \\ Q_H &= U_H Q U_H^T + P_H^\perp &= \frac{H}{\sqrt{N}} Q \frac{H^T}{\sqrt{N}} + I - \frac{HH^T}{N}, \end{cases}$$

$c$  is any constant and  $Q$  is any  $2 \times 2$  orthogonal matrix,  $U_S$  and  $U_H$  are defined in an obvious way.

Under the above transformation,  $g(\mathbf{y})$  is explicitly expressed as

$$g(\mathbf{y}) = \mu_1 \mathbf{H} \phi_1 + \mathbf{S} \phi_1 + \sigma_1 \mathbf{n} \quad (2.21)$$

with the induced transformation  $\tilde{G}$  given by:

$$\begin{cases} \mu_1 &= \mu \\ \phi_1 &= cQ\phi \\ \sigma_1 &= c\sigma. \end{cases}$$

Note that  $Q_S$  and  $Q_H$  are two orthogonal matrices (rotational matrices [56]). The geometrical meaning of this transformation is thus consecutive rotations of the original signal within the  $S$  plane (defined as the subspace spanned by the columns of  $S$ ) and then within the  $H$  plane followed by scaling that introduces unknown variance.

The following is presented to make the proof as applicable as possible to a general case.

**Proof:**

$$Q_H \mathbf{y} = Q_H S \phi + \mu Q_H H \phi + \sigma Q_H \mathbf{n} = S \phi + \mu H \phi' + \sigma Q_H \mathbf{n},$$

since

$$Q_H S = (U_H Q U_H^T + P_H^\perp) S = P_H^\perp S = S,$$

because  $H^T S = 0$ , i.e.,  $U_H^T S = \frac{H^T}{\sqrt{N}} S = 0$ .

Similarly,

$$Q_S H = H,$$

while

$$\begin{aligned} H\phi' &= Q_H H\phi = (U_H Q U_H^T + P_H^\perp) H\phi = U_H Q U_H^T H\phi \\ &= U_H (U_H^T U_H)^{-1} U_H^T U_H Q U_H^T H\phi \\ &= H (H^T H)^{-1} H^T Q_H H\phi \end{aligned} \tag{2.22}$$

due to the fact that  $U_H^T U_H = \mathbf{I}$ .

Therefore,

$$Q_S Q_H \mathbf{y} = Q_S (S\phi + \mu H\phi' + \sigma Q_H \mathbf{n}) = S\phi'' + \mu H\phi' + \sigma Q_S Q_H \mathbf{n}$$

where, similarly to Equation (2.22),

$$S\phi'' = Q_S S\phi = S(S^T S)^{-1} S^T Q_S S\phi. \tag{2.23}$$

It turns out that  $\phi'$  in Equation (2.22) and  $\phi''$  in Equation (2.23) are coincident, which is the fascinating and devious part for finding the above set of transformations:

$$\phi' = (H^T H)^{-1} H^T Q_H H\phi = \frac{1}{N} \frac{H^T H}{\sqrt{N}} Q \frac{H^T}{\sqrt{N}} H\phi = Q\phi,$$

due to the fact that  $\mathbf{H}^T \mathbf{H} = N \mathbf{I}_{2 \times 2}$ . Similarly,

$$\phi'' = (S^T S)^{-1} S^T Q_S S \phi = Q \phi = \phi'.$$

Regarding the transformed form of noise  $\mathbf{n}$ , since  $Q_S$  and  $Q_H$  are both orthogonal matrices,  $Q_S Q_H \mathbf{n}$  has exactly the same covariance matrix form as that for  $\mathbf{n}$ . So,  $g(\mathbf{y}) = c Q_S Q_H \mathbf{y}$  assumes the form as expressed in Equation (2.21). **QED**.

It follows that  $(\mu, \|\phi\|^2/\sigma^2)$  or more simply  $(\mu, a^2/\sigma^2)$  is a set of maximal invariant parameters under  $\bar{G}$  (see [7, 46, 56]) for details on invariance principles).

Furthermore, it is easy to verify that  $t_3(\mathbf{y})$  is invariant to the transformation group  $G$ . To see this, first of all, we observe that  $t_3(\mathbf{y})$  or  $L_3(\mathbf{y})$  is a function of only  $\|\mathbf{y}\|$ ,  $\|\theta_1(\mathbf{y})\|$ ,  $\|\theta_2(\mathbf{y})\|$  and  $\cos \varphi$  (equivalently  $\theta_1^T \theta_2$ ). We already know that  $Q_S$  and  $Q_H$  are both orthogonal matrices. Moreover, noting that

$$H^T Q_S = H^T, H^T Q_H = Q H^T, S^T Q_S = Q S^T, S^T Q_S Q_H = Q S^T Q_H = Q S^T,$$

we have invariance for the norms

$$\|\theta_1(\mathbf{y})\| = \|\theta_1(Q_S Q_H \mathbf{y})\|, \quad \|\theta_2(\mathbf{y})\| = \|\theta_2(Q_S Q_H \mathbf{y})\|, \quad \|\mathbf{y}\| = \|Q_S Q_H \mathbf{y}\|.$$

and

$$\theta_1^T(\mathbf{y}) \theta_2(\mathbf{y}) = \theta_1^T(Q_S Q_H \mathbf{y}) \theta_2(Q_S Q_H \mathbf{y}).$$

A constant  $c$  in both numerator and denominator of  $L_3(\mathbf{y})$  and  $\cos \varphi$  does not change the original quantities. Therefore,  $t_3(\mathbf{y})$  is invariant to the transformation group  $G$ .

Hence, from Theorem 3 in Appendix V of Chapter 5, the pdf of  $t_3(\mathbf{y})$  is a function of  $\mu$  and  $a/\sigma$  alone (instead of all four model parameters  $a, \sigma, \vartheta$ , and  $\mu$ ).

This is a desirable and very useful feature, since this result shows that the test is only a function of two key variables whereas in general a test could depend on all four of the unknown parameters  $(a, \mu, \vartheta, \sigma^2)$ . In particular, the GLRT is invariant to the unknown phase  $\vartheta$ . The invariance property saved me a lot of work when I performed numerical simulations in the next stage.

Unfortunately though, the dependence of the test on  $a^2/\sigma^2$  in addition to significant parameter  $\mu$  implies that, in general, this test does not have the CFAR property. Hence, selection of a threshold  $\eta_3$  to achieve a desired  $P_f$  is still very difficult.

### 2.6.3 Upperbound of GLRT Test Statistics

However, some interesting relationships exist between the nonlinear GLRT test and the CC test, which suggest some possibilities for threshold selection. Let us compare Equation (2.8) with Equation (2.10). Note that if  $\varphi = 0$  in Equation (2.10), then  $L_3$  and  $L_2$  coincide. It is precisely through the term  $\cos 2\varphi$  that the effect of phase coupling comes into play. Actually, from Equation (2.10), the upper bound of  $L_3(\mathbf{y})$  is easily seen to be coincident with  $L_2(\mathbf{y})$ .

Therefore, one method of threshold selection is to choose the threshold slightly smaller than that determined for the CC test statistics, which has the  $F_{2,2(N-1)}$  distribution. Furthermore, if the true parameter  $\mu$  under  $H_1$  is small, which is the case for most fMRI detection problems, then we show in the following subsection that  $L_3(\mathbf{y})$  asymptotically (as  $N$ , the length of the time-series, increases) has the same distribution as  $L_2(\mathbf{y})$ . This provides a more solid foundation for this threshold choice method. Under the guidance of this rough method, extensive numerical simulations lead to more accurate threshold selection method in Subsection (2.6.5).

### 2.6.4 Asymptotic of GLRT Test Statistics

In detection problem, our concern is low SNR case. In our situation, we assume that the true parameter  $\mu$  under  $H_1$  is small, *i.e.*,  $\mu \rightarrow 0$ . By the asymptotic property of maximum likelihood estimates, as  $N \rightarrow \infty$ ,  $\hat{\mu} \rightarrow \mu$ . In order to get a more accurate approximation of  $\hat{\mu}$ , we use Equation (2.16) combined with  $\hat{\mu} \rightarrow \mu \rightarrow 0$  (as  $N \rightarrow \infty$ ) and get  $\hat{\phi} \rightarrow \theta_2$ , so from Equation (2.15) (as  $N \rightarrow \infty$ ),

$$\hat{\mu} \approx \frac{\theta_1^T \theta_2}{\theta_2^T \theta_2}, \quad (2.24)$$

which is the maximum likelihood estimate for the corresponding linear model in Equation (2.3).

Substituting Equation (2.24) into Equation (2.18) we have

$$\min \hat{\sigma}_1^2 = \|\mathbf{y}\|^2 - N(\theta_1^T \theta_1 + \theta_2^T \theta_2) = \|\mathbf{y}\|^2 - \|P_S \mathbf{y}\|^2 - \|P_H \mathbf{y}\|^2. \quad (2.25)$$

Noting that

$$P_{SH}^\perp = P_S^\perp P_{SH}^\perp P_S^\perp, \quad (2.26)$$

$$P_S^\perp - P_{SH}^\perp = P_S^\perp P_H P_S^\perp = P_H = P_H P_S^\perp, \quad (2.27)$$

$$P_{SH}^\perp P_S^\perp = P_H^\perp P_S^\perp = P_{SH}^\perp = I - P_S - P_H. \quad (2.28)$$

we get, from Equations (2.13), (2.14), (2.25), as  $N \rightarrow \infty$ ,

$$L_3(\mathbf{y}) \approx L_2(\mathbf{y}).$$

This implies that  $t_3(\mathbf{y})$  asymptotically has the same distribution as  $t_2(\mathbf{y})$ , *i.e.*, non-central  $F_{2,2(N-1)}(\text{SNR})$ , and thus asymptotically has the CFAR property with small  $\mu$ .

### 2.6.5 Threshold Selection Based on Numerical Simulation

Another method is to try to determine the exact thresholds via Monte Carlo simulation. Some of the results of our simulations are given in the next subsection. Here, we summarize the conclusions. Extensive Monte Carlo simulation reveals that the GLRT test is also CFAR when  $a/\sigma \geq 1$ , which is the case for most, if not all, fMRI experiments. More importantly and more interestingly, *to achieve the desired  $P_f$ , the proper threshold of our GLRT detector is almost exactly one half that of the corresponding threshold required for an  $F_{1,(N-1)}$  distributed test statistic.* This is confirmed by extensive Monte Carlo simulation.

### 2.6.6 Distribution of GLRT Test Statistics

The observation in above subsection may lead to *Nan's Conjecture* formulated as follows.

In mathematical terms, we have

$$P_f = \int_{\eta}^{\infty} p_{t_3|H_0}(t_3) dt_3 \approx \int_{2\eta}^{\infty} f_0(t) dt, \quad (2.29)$$

where  $f_0$  denotes the density of an  $F_{1,(N-1)}$  distributed statistic. Differentiating the above equation with respect to  $\eta$  leads us to an exciting result that the density of the test statistic  $t_3$  under  $H_0$  is related to the  $F_{1,N-1}$  density by the approximation

$$p_{t_3|H_0}(t_3) \approx 2f_0(2t_3). \quad (2.30)$$

This implies that a very accurate threshold can be selected using standard  $F_{1,(N-1)}$  distribution tables [1].

Actually, on careful observation of the following Tables (2.1-2.3), we find that  $P_d$ 's

for GLRT and for MC detectors are the same, when  $a/\sigma$  is large, *i.e.*,

$$P_d = \int_{\eta}^{\infty} p_{t_3|H_1}(t_3) dt_3 \approx \int_{2\eta}^{\infty} f_1(t) dt, \quad (2.31)$$

where  $f_1$  denotes the density of an  $F_{1,(N-1)}$ (SNR) distributed statistic, since for large  $a/\sigma$ , the approximation for the MC model in Equation (2.4) is quite valid and hence the  $P_d$  for MC detector is given by the right integral of above equation. Differentiating the above equation with respect to  $\eta$  leads us to another result that the density of the test statistic  $t_3$  under  $H_1$  is related to the  $F_{1,N-1}$ (SNR) density by the approximation

$$p_{t_3|H_1}(t_3) \approx 2f_1(2t_3). \quad (2.32)$$

When  $\mu = 0$ ,  $f_1$  is coincident with  $f_0$ . Summarizing two cases, I suspect that

$$p_{t_3}(t_3) \approx 2f_1(2t_3), \quad (2.33)$$

for  $a/\sigma > 1$ . Although theoretical proof for this conjecture is difficult to achieve, it is still worth investigating. Because the pdf of this statistic does not depend on the angle  $\vartheta$ , the expression of  $\mathbf{y}$  simplifies greatly and may direct us to its final solution.

I caution here that numerical simulations reveal that the conclusions in Subsections (2.6.5) and (2.6.6) all break down for  $a/\sigma < 1$ .

## 2.7 Comparisons of the Three Detectors

In order to compare the performances of the three detectors, I run extensive Monte Carlo experiments. Because originally we only know that the performance of GLRT depends on two parameters  $a/\sigma$  and  $\mu \equiv b/a$ , it is necessary to study the performance for different values of  $a/\sigma$ . However, as mentioned above, for  $a/\sigma \geq 1$ , the Monte

|            | Threshold | 4.70 | 3.43 | 6.85 |
|------------|-----------|------|------|------|
| $a/\sigma$ | $\mu$     | CC   | GLRT | MC   |
| 1          | .3162     | .72  | .80  | .44  |
| 3.162      | .1        | .72  | .80  | .78  |
| 10         | .03162    | .72  | .80  | .80  |

Table 2.1.  $P_d$  with  $P_f = 0.01, N = 120$

|            | Threshold | 3.75 | 2.58 | 5.15 |
|------------|-----------|------|------|------|
| $a/\sigma$ | $\mu$     | CC   | GLRT | MC   |
| 1          | .3162     | .82  | .88  | .58  |
| 3.162      | .1        | .82  | .88  | .87  |
| 10         | .03162    | .82  | .88  | .88  |

Table 2.2.  $P_d$  with  $P_f = 0.025, N = 120$

Carlo analysis suggests that the GLRT is essentially CFAR.

In Tables (2.1-2.3), we compare the detection rates  $P_d$  of the three tests under three  $P_f$  specifications. The  $P_f$ 's are selected to be representative of those commonly used in fMRI. In these tables, the first row contains the thresholds corresponding to the preselected  $P_f$ . In order to see the functional dependence of  $P_d$  on  $\text{SNR} \equiv \mu^2 a^2 / \sigma^2$ , I deliberately select  $\mu$  so that SNR is the same for three different  $a/\sigma$  cases. In order to get as accurate results as possible, for each value of  $P_d$  (and  $P_f$ ),  $10^5$  simulations are run and the average is taken as true result.

The most difficult element of the Monte Carlo analysis, except in the CC test case, is the determination of proper thresholds to achieve a desired false-alarm rate with each detectors. The CC test is  $F_{2,2(N-1)}$  distributed under  $H_0$ , and therefore the proper threshold is very easily determined from standard tables [1].

Because the GLRT is not known to possess the CFAR property, the proper threshold will, in general, depend on  $a/\sigma$ . For a given value of  $a/\sigma$ , the threshold needed to achieve a desired false-alarm rate can be determined via Monte Carlo analysis and trial-and-error over a range of thresholds under the guidance of the rough method

|            | Threshold | 3.03 | 1.96 | 3.92 |
|------------|-----------|------|------|------|
| $a/\sigma$ | $\mu$     | CC   | GLRT | MC   |
| 1          | .3162     | .88  | .93  | .69  |
| 3.162      | .1        | .88  | .93  | .92  |
| 10         | .03162    | .88  | .93  | .93  |

Table 2.3.  $P_d$  with  $P_f = 0.05, N = 120$

described in Subsection (2.6.3). This is precisely how the thresholds were determined for the results given in Tables (2.1-2.3). Remarkably, however, the Monte Carlo analysis revealed that both the MC test and the GLRT are essentially CFAR so long as  $a/\sigma > 1$ , which is almost always true in fMRI. Moreover, the Monte Carlo analysis supports the use of some very simple rules for threshold selection.

First, in the case of the MC test, for very large values of  $a/\sigma$  the magnitude data are very well approximated as Gaussian. Therefore, in such situations, the MC test is (approximately)  $F_{1,(N-1)}$  distributed under  $H_0$  and the proper threshold can be determined again from standard tables [1]. Because the Monte Carlo simulations show that for  $a/\sigma \geq 1$  the MC test is essentially CFAR, the proper threshold may be determined from  $F_{1,(N-1)}$  distribution for all cases in which  $a/\sigma \geq 1$ . The derivation of approximation model (2.4) for MC detector also supports this, although not strictly.

Second, the similar performances of the GLRT and MC test for large  $a/\sigma$  suggest the possibility of a relationship between the GLRT statistic and the  $F_{1,(N-1)}$ . This intuition led to the discovery that the proper threshold for the nonlinear GLRT can be selected as one half the threshold required to achieve the desired  $P_f$  for a  $F_{1,(N-1)}$  distributed statistic.

The results in the three tables show clearly that our GLRT detector performs best for all three (low, medium, high)  $a/\sigma$  cases. The CC detector performs better than the MC detector in the low  $a/\sigma$  case. However, as the DC component becomes more and more dominant over the noise, the GLRT and MC test significantly outperform

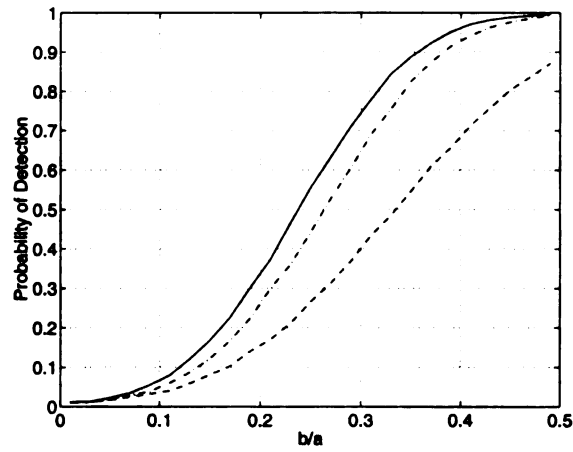
the CC test.

Finally, note that the detection rate of the CC test is constant for fixed  $\text{SNR} = \mu^2(a/\sigma)^2$ . This is expected because the CC test statistic is non-central  $F_{2,2(N-1)}(\text{SNR})$  under  $H_1$ . Remarkably, note that the dependence of detection rate of our GLRT detector also depends only on SNR. The same is not true of the MC test, whose performance drops severely as  $a/\sigma$  decreases.

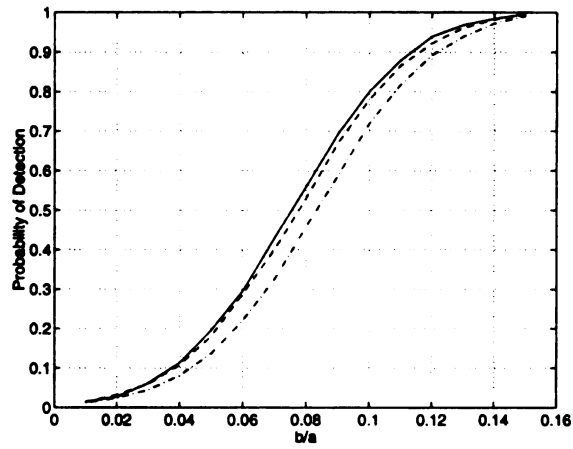
I also illustrate in Figure (2.4) the results using three performance curves ( $P_D$  versus response strength  $\mu = \frac{b}{a}$  curves) with  $N = 120, P_f = .01$ . Therefore, the thresholds are chosen as in Table (2.1). Solid (—) line for GLRT; dash-dot(-.) line for CC; dashed (— —) line for MC. Figure (2.4a) shows the case for  $a/\sigma = 1.0$ , which is low, so we expect MC detector to suffer. It is indeed the case: the GLRT curve is at the top, CC curve is in the middle and the bottom one is for MC. Figure (2.4b) shows the case for  $a/\sigma = 3.162$ , which is large. In this case, the top one is for GLRT, the middle one for MC, the bottom one for CC. It shows the MC detector begins to surpass the CC detector but is still inferior to the GLRT. Figure (2.4c) indicates the case for  $a/\sigma = 10$ , which is quite large, and so the MC and GLRT detectors have almost the same performance as shown in the figure: the GLRT and MC curves coalesce to one in the top while the CC detector remains at the bottom. All three curves clearly demonstrate that our GLRT detector is always the best.

## 2.8 A Simulated fMRI Study

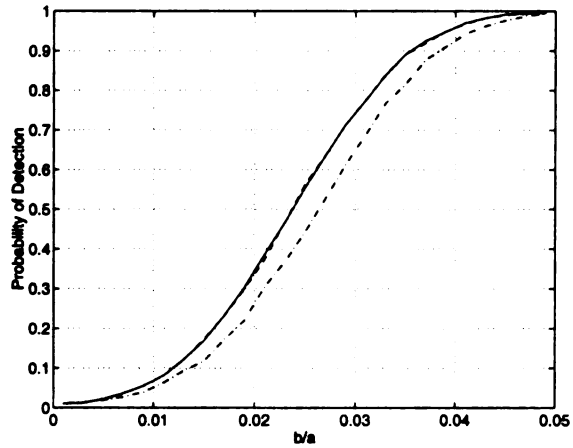
One fMRI experiment is simulated to illustrate pixel-wise detection efficiency by the above three detectors. The results are shown in Figure (2.5). Figure (2.5a) shows one slice image of the brain ( $64 \times 64$  pixels) with simulated activation region highlighted. A  $9 \times 9$  voxel region in the lower right corner of the brain (indicated in white) is selected to be active in this simulation. For the simulation, a time series with length



(a)



(b)



(c)

Figure 2.4. Three curves comparing the performance of three detectors. (a)  $a/\sigma = 1$ ; (b)  $a/\sigma = 3.162$ ; (c)  $a/\sigma = 10$ .

$N = 120$  is simulated for each voxel. The reference signal  $\mathbf{r}$  is a square wave with period 10. The fluctuation of reference signal about constant level is  $\pm 10\%$ , *i.e.*,  $\mu = 0.1$ . The noise variance in each time series is set so that  $a/\sigma = 3.162$ . For each time series, the phase is a constant. Spatially, the phase has a fluctuation (Gaussian noise with zero mean 0.1 variance) about a constant phase of  $\pi/3$ . The reader is referred to Figure (3.13a) in Chapter 3 for the correlation image for this simulated experiment.

The desired false-alarm rate in this example is chosen to be  $P_f = 0.01$ , and thus the three thresholds for CC, GLRT and MC detectors are 4.70, 3.43, and 6.85, respectively. The MC test, CC test, and GLRT test are compared in Figures (2.5b-d). The actual detection rates observed in this simulation, given in the caption of Figure (2.5), are in excellent agreement with the tabulated Monte Carlo results.

However, we note that in the activation maps shown in Figure (2.5), there is activation "detected" outside the brain. This is a serious problem which results from the fact that spatial information is completely ignored. This is the main defect of pixel-wise detection. Therefore next chapter is devoted to dealing with the problem of how to utilize spatial information to further enhance detection efficiency.

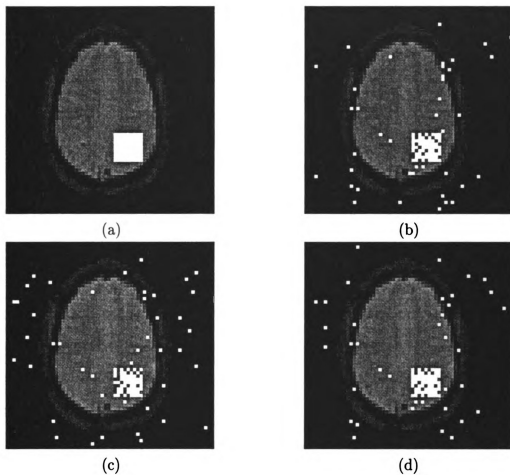


Figure 2.5. A simulated fMRI experiment illustrating pixel-wise detection by three detectors. (a) Brain image with simulated activation region highlighted; (b) MC test results:  $P_d = 0.77$ ; (c) CC test results:  $P_d = 0.70$ ; (d) GLRT results:  $P_d = 0.79$ .

# CHAPTER 3

## Multi-scale Detection for fMRI

### 3.1 Overview

The last chapter focused on pixel-wise detection for fMRI, which is most common in practice [3, 24, 25, 36, 55, 60, 64]. However, these techniques do not take advantage of mutual information among neighboring pixels. Ignoring such spatial information can reduce detection accuracy. Utilizing spatial information may enhance our detection accuracy. For example, in Figure (2.5) activity is “detected” in areas outside the brain — an erroneous decision that could be avoided by incorporating anatomical information in the decision rule. Furthermore, it may be quite possible that connected region of activation is larger than individual pixel dimensions. In other words, activated areas in reality tend to occur in clusters of neighboring pixels. Thus, limiting testing to individual pixels imposes artificial boundaries in the analysis process that may weaken the detection performance. On the other hand, if there is strong indication that a large group of pixels, which may be thought of as one large pixel at a very coarse (spatial) scale, is active, then the individual pixels inside this group may be more likely to be active themselves. Hence comes the idea of (spatial) scale and incorporating spatial correlation into the fMRI detection process.

### 3.1.1 Spatial Modeling and Outline for the Method

In light of the ideas above, the pixel-wise detection is oversimplistic. Therefore, it is necessary to develop detection methods that take advantage of spatial correlation. There are many approaches to attacking the problem, for example, cluster analysis [20, 26, 27] and independent component analysis (ICA) [43]. Detection methods using Bayesian strategies have also been recently proposed for fMRI [21, 35, 17]. Just as in the pixel-wise detection strategies, we need to model each time series; when we incorporate spatial correlation, we also need develop spatial models of the fMRI data. This is by no means easy. The recent works of [35, 17] use Markov random field (MRF) models to model the spatial relationships in fMRI data.

These methods all have their shortcomings. The clustering and independent component analysis techniques are somewhat *ad hoc* and do not enable explicit modeling assumptions about spatial correlation. The existing Bayesian methods mentioned above are all restricted to modeling only the finest scale (highest resolution). Such methods tend to be very computationally demanding, and are often difficult to analyze and interpret. Therefore, we will put forward a novel multi-scale modeling and detection framework that incorporates spatial correlation information and is much more amenable to analysis and optimization.

More specifically, this chapter will present *a two-step approach for fMRI detection*. First, a new multi-scale *image segmentation* algorithm is proposed to decompose the correlation image into several different regions, each of which is of homogeneous statistical behavior. Second, each homogeneous region will be classified independently as active or inactive using detection methods analogous to the *pixel-wise test* described in Chapter 2.

### 3.1.2 General Setting of Bayesian Image Segmentation

In a general setting, the fMRI activation mapping may be viewed as a particular image segmentation problem. We are given a random *continuous noisy image*  $Y$  which must be segmented into a *discrete image*  $X$  consisting of regions of distinct statistical behavior. For example, in fMRI, image  $Y$  may be composed of the correlation values (specifically, the correlation between amplitude time series and the reference signal) at each pixel location; the image  $X$  may be just a binary detection map, as in Chapter 2.

To simplify the presentation, I will modify the notation slightly from that used in previous chapters. From now on, I will adopt some notation from [10]. Symbols without subscript refer to the whole image field. Individual pixels in the image  $X$  are denoted by  $X_k$  where  $k$  is a point of a one-dimensional (1-D) or two-dimensional (2-D) lattice, depending on the context. The collection of lattice points at scale  $j$  is denoted as  $S^j$ . Random quantities are usually denoted by upper-case letters. For notational ease, however, lower-case letters may denote the stochastic quantities or corresponding deterministic realizations, which should be distinguishable also from contexts.

We assume that each observed pixel in image  $Y$  is dependent on a corresponding unobserved label in  $X$ . Each label specifies one of  $M$  possible states, each with its own statistical behavior. In our case,  $M = 2$  indicating “active” and “inactive”. However, in general, we may need to segment the correlation image into regions with homogeneous statistical behavior, so  $M \geq 2$ .

The dependence of observed pixels on their labels is specified through the conditional distribution of  $Y$  given  $X$ , *i.e.*,  $p_{y|x}(y|x)$ . The function  $p_{y|x}(y|x)$  is called the likelihood function. In fMRI, we are essentially interested in inverting this relationship; that is, we would like to determine  $p_{x|y}(x|y)$ , the probabilistic description of the

unknown activation map given the observed data. This calculation is facilitated by introducing *a priori* knowledge about the size and shapes of regions, modeled by a *prior* distribution  $p_x(x)$ .

By Bayes formula, we will estimate  $X$  given observed image  $Y = y$  as:

$$\begin{aligned}\hat{x} &= \arg \max_x p(X = x|Y = y) \\ &= \frac{p(X=x)p(Y=y|X=x)}{p(Y=y)} \\ &= \frac{p(X=x)p(Y=y|X=x)}{\sum_{\text{all } x} p(X)p(Y=y|X)} \\ &\propto p_x(x)p_{y|x}(y|x).\end{aligned}$$

where upper case letters denote random quantities and lower case letters denote the deterministic realizations.

This is the so-called *maximum a posteriori* (MAP) estimator. The general framework for Bayesian image segmentation problem is shown in Figure (3.1) (adapted from [10]).

Despite the apparent simplicity of this estimator, remember that  $X$  is 2-D image of integer values, making optimization prohibitively difficult, in general. Further, specification of the *a priori* distribution  $p_x(x)$  is not straightforward, either.

### 3.1.3 Multi-scale Image Segmentation Methods and Advantages

Traditionally, statistical image segmentation has been accomplished using MRF models. The global statistical models in the MRF theory lead to substantially better segmentation results than those of simpler, local methods [23, 50]. The theory of MRF models provides a powerful framework for studying nonlinear interactions among different features [50]. Under the MAP criterion, it leads to the minimization of a global

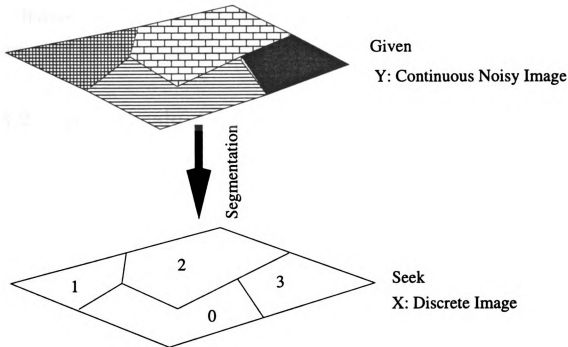


Figure 3.1. Illustration of image segmentation.

energy function which is very computationally expensive to carry out [28, 23, 45].

Recently, an alternative to the classical MRF approach to image segmentation was proposed in [9, 10]. This approach is based on modeling the discrete field  $X$  as a multi-scale Markov chain. In the sequel, it is called Algorithm I.

The multi-scale hidden Markov model (MHMM) proposed in [15] and its extension [49] may also be used to deal with the segmentation problem. Instead of the discrete field  $X$ , the states of the wavelet coefficients at different scales are modeled as a Markov chain. In the sequel, this approach is called Algorithm II.

According to the results and conclusions of previous work [10, 11, 40, 49], this kind of multi-scale modeling not only captures the key inter-scale *physical dependency* present in natural signals and images, it also leads to computationally efficient (usually scale-recursive) algorithms. A more precise explanation of *physical dependency* will be given later in the review of Algorithm I.

However, before proceeding to the detailed description of algorithms, I will first give a brief introduction to multi-scale analysis.

## 3.2 Multi-scale Analysis

In general, multi-scale analysis refers to the study of behavior of signals or images at various spatial and/or temporal resolutions [8, 32, 53, 54, 63, 65].

### 3.2.1 1-D Multi-resolution Analysis (MRA)

Let us begin by considering the multi-scale or *multi-resolution* analysis of a finite energy 1-D signal. There does exist 2-D MRA. However, it is not used in this dissertation and thus omitted. Finite energy signals are those signals belonging to the space of square-integrable functions on the real line,  $L^2(R)$ , see Appendix I in Chapter 5.

An MRA of  $L^2(R)$  is defined to be a sequence  $\{V^j | j \in Z\}$  of closed subspaces of  $L^2(R)$  satisfying the following properties [32, 53, 65]:

1)

$$V^j \subset V^{j+1}, \quad (3.1)$$

2)

$$f(x) \in V^j \iff f(2x) \in V^{j+1}, \quad (3.2)$$

3) for all  $f \in L^2(R)$ ,

$$\lim_{j \rightarrow -\infty} \|P_V^j f\| = 0, \quad \text{or } \lim_{j \rightarrow -\infty} V^j = \{0\} \quad (3.3)$$

$$\lim_{j \rightarrow \infty} \|P_V^j f - f\| = 0, \quad \text{or } \lim_{j \rightarrow \infty} V^j = L^2(R) \quad (3.4)$$

where  $P_V^j$  is the orthogonal projection on the space  $V^j$ , which can be considered as an approximation of  $f$  at the scale of  $2^{-j}$ . In the above definition,  $j$  indexes the scale or resolution of analysis — a smaller  $j$  corresponds to a lower resolution of analysis. A resolution of  $2^j$  corresponds to a scale of  $2^{-j}$ . The first limit implies that as resolution gets smaller and smaller ( $2^j \rightarrow 0$ ), the approximation is just 0, all information (particularly its details) about  $f$  is lost. The second limit implies that the approximation  $P_V^j f$  converges to the true signal  $f$  as resolution gets higher and higher ( $2^j \rightarrow \infty$ ).

4) There exists a function  $\phi(t) \in L^2(R)$  such that  $\{\phi(t - k) | k \in Z\}$  form an orthonormal basis for  $V^0$ .  $k$  indicates the time (or spatial) location of analysis.  $\phi(t)$  is called the *scaling function* of the associated MRA, which plays an essential role in MRA. Then it follows that the family  $\{\phi_k^j(t) = 2^{j/2} \phi(2^j t - k) | k \in Z\}$  is an orthonormal basis of  $V^j$  for all  $j \in Z$ . The factor  $2^{j/2}$  is introduced for normalization:  $\|\phi_k^j(t)\| = \|\phi(t)\| = 1$ .

The nesting of  $V^j \subset V^{j+1}$  implies that  $\phi \in V^0$  can be expressed as a linear combination of  $\{\sqrt{2}\phi_k^1 | k \in Z\}$ :

$$\phi(t) = \sum_k h_k \sqrt{2} \phi(2t - k), \quad (3.5)$$

where coefficients  $\{h_k = \langle \phi(t), \sqrt{2}\phi(2t - k) \rangle | k \in Z\}$  constitute *scaling filter*.

In addition to  $V^j$ ,  $W^j$  is defined as the orthogonal complement of  $V^j$  in  $V^{j+1}$ , i. e.,  $V^{j+1} = V^j \oplus W^j$  for any  $j \in Z$ .  $W^j$  represents the additional information that is necessary to pass from an approximation at resolution  $2^j$  to an approximation at the higher resolution  $2^{j+1}$ . A direct consequence is  $W^i \perp W^j, i \neq j$ . For example, since

---

For sake of generality, some literature first introduce a function  $b(t) \in V^0$  such that  $\{b(t - k), k \in Z\}$  is a *Riesz basis* for  $V^0$ . Then another function  $\phi(t) \in L^2(R)$  can be constructed from  $b(t)$  such that  $\{\phi(t - k) | k \in Z\}$  forms an orthonormal basis for  $V^0$ . The content of this dissertation is not on these issues of function construction, therefore I adopt a more direct and simpler definition.

$W^{i+1} \perp V^{i+1} = W^i \oplus V^i$ , we have  $W^{i+1} \perp W^i$ .

As in the case of  $V^0$ , there exists another function  $\psi \in W^0$ , which is called the *wavelet function* and can be constructed from  $\phi$ , such that the family  $\{\psi(x-k)|k \in Z\}$  forms an orthonormal basis for the space  $W^0$ . To understand how one can generate  $\psi(t)$  from  $\phi(t)$ , consider the following. Because  $W^j \subset V^{j+1}$ , wavelet function  $\psi \in W^0$  can also be expressed as a linear combination of  $\{\sqrt{2}\phi_k^1|k \in Z\}$ :

$$\psi(t) = \sum_k g_k \sqrt{2}\phi(2t - k), \quad (3.6)$$

where  $g_k, k \in Z$  is chosen to be  $^\dagger(-1)^k h_{1-k}$  and constitutes the *wavelet filter*. It follows that the family  $\{\psi_k^j(t) = 2^{j/2}\psi(2^j t - k)|k \in Z\}$  is an orthonormal basis of  $W^j$  for all  $j \in Z$ .

Iteratively applying the relationship  $V^{j+1} = V^j \oplus W^j$ , we have

$$V^J = V^{J-1} \oplus W^{J-1} = V^{J-2} \oplus W^{J-2} \oplus W^{J-1} = \dots = V^{J_0} \oplus W^{J_0} \dots \oplus W^{J-1}, \quad (3.7)$$

As  $J \rightarrow \infty$ , we get

$$L^2(R) = \oplus_{j \geq J_0} W^j \oplus V^{J_0}. \quad (3.8)$$

Further, let  $J_0 \rightarrow -\infty$ , then we get

$$L^2(R) = \oplus_{j \in Z} W^j. \quad (3.9)$$

Equation (3.8) means that there exist  $c_k^{J_0}$ , called *scaling coefficients*, and  $\theta_k^j$ , called *wavelet coefficients*, such that any one-dimensional signal  $f \in L^2(R)$  can be repre-

---

It turns out that there exist other recipes for  $\psi$ . For example,  $g_k = (-1)^k h_{1-k+2n}, n \in Z$  or  $g_k = (-1)^{k-1} h_{-k-1}$  also work.

sented as:

$$f(t) = \sum_k c_k^{J_0} \phi_k^{J_0}(t) + \sum_{j=J_0}^{\infty} \sum_k \theta_k^j \psi_k^j(t), \quad (3.10)$$

where the coefficients  $c_k^{J_0} = \langle f, \phi_k^{J_0} \rangle$  and  $\theta_k^j = \langle f, \psi_k^j(t) \rangle$  are called the *discrete wavelet transform* (DWT) of  $f$ . The reader is reminded that  $\theta$  in this chapter denotes wavelet coefficients. By invoking Equations (3.5) and (3.6), there exist recursive relations:

$$c_k^j = \langle f, \phi_k^j \rangle = \sum_n h_{n-2k} c_n^{j+1}, \quad (3.11)$$

$$\theta_k^j = \langle f, \psi_k^j \rangle = \sum_n g_{n-2k} c_n^{j+1}. \quad (3.12)$$

In practice, there is a fundamental limit on the meaningful resolution when we sample a continuous (infinitely high resolution) temporal signal or spatial image. Therefore, we usually start with a scale subspace  $V^J$ , with  $J$  chosen to be large enough to represent the finest details of interest in a signal, since  $P_V^J f \approx f$  for large  $J$  (see Equation (3.4)). So we replace the semi-infinite sum in Equation (3.10) with a sum over a finite number of scales  $J_0 \leq j \leq J$ ,  $\{J, J_0\} \subset \mathbb{Z}$ , where  $J_0$  and  $J$  indicate the coarsest scale (or lowest resolution) and finest scale (or highest resolution), respectively.

This point can be paraphrased from another perspective as well. Because at high resolutions, the scaling functions are similar to Dirac delta functions (assuming  $\phi$  is localized and well-behaved, *i.e.*  $\lim_{t \rightarrow \infty} \phi(t) = 0$ ), since the time scale is compressed while the magnitude scale is enlarged,

$$j \rightarrow \infty, \quad 2^{j/2} \phi(2^j t - k) \rightarrow 2^{-j/2} \delta(t - k 2^{-j}),$$

which can be verified by checking the integration results of both sides. Therefore, for

$j$  sufficiently large, for example,  $j = J$ , we have

$$\begin{aligned}
c_k^J &= (f, \phi_k^J) \\
&= \int_{-\infty}^{\infty} f(t) 2^{J/2} \phi(2^J t - k) dt \\
&= \int_{-\infty}^{\infty} f(t) 2^{-J/2} \delta(t - k 2^{-J}) dt \\
&= 2^{-J/2} f(k 2^{-J}) = 2^{-J/2} f(k T_s).
\end{aligned}$$

In other words, the scaling coefficients are approximately proportional to signal samples at a sampling rate of  $T_s = 2^{-J}$ . So in practical computation of DWT, we start with an initial set of scaling coefficients  $c_k^J$ , which are assumed to represent an approximation to signal  $f$  at a certain scale  $2^{-J}$  (corresponding to the sampling period  $T_s = 2^{-J}$ ).

The wavelet and scaling coefficients at coarser scales  $j < J$  can be computed recursively using the lowpass scaling filter  $\{h_{-n}\}$  and highpass wavelet filter  $\{g_{-n}\}$ , but only even-indexed samples at filter outputs are retained (*downsampling*) according to Equations (3.11) and (3.12). This is called the *pyramidal algorithm* [65], the realization structure for which is depicted in Figure (3.3).

Similarly, for signal synthesis (inverse DWT), we have

$$c_k^{j+1} = \sum_m c_m^j h_{k-2m} + \sum_m \theta_m^j g_{k-2m}. \quad (3.13)$$

The synthesis operations can be implemented using filter banks as well, involving interpolation (*upsampling*) and the two filters  $\{h_n\}$  and  $\{g_n\}$ , as shown in Figure (3.4).

Under some conditions [32, 65] a filter  $\{h_n | n \in \mathbb{Z}\}$  corresponds to a valid MRA satisfying the aforementioned conditions. Determination of the set of coefficients  $\{h_n | n \in \mathbb{Z}\}$  (or corresponding scaling function  $\phi(t)$ ) is beyond scope of this disserta-

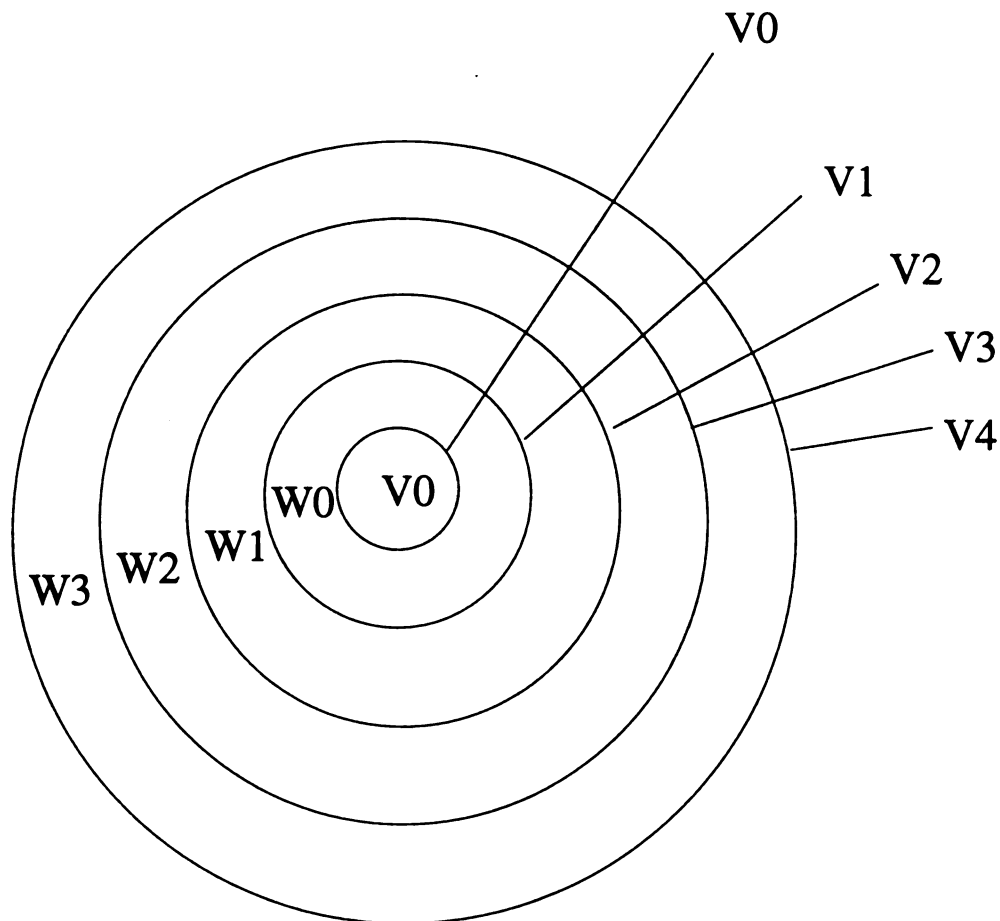


Figure 3.2. Nested scale spaces and wavelet spaces

tion. Here, we are only interested in the simplest DWT, *i.e.*, Haar wavelet transform, in which case,

$$\phi(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{otherwise,} \end{cases}, \quad \psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & \text{otherwise,} \end{cases}$$

and the analysis algorithm degenerates to much simpler form. See Section (3.4).

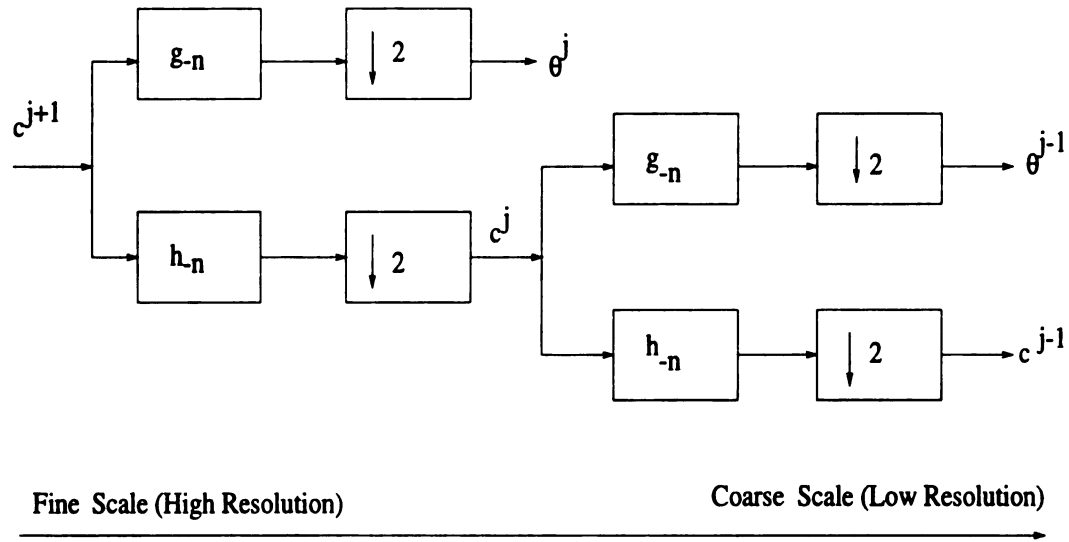


Figure 3.3. Computation of DWT by filter bank

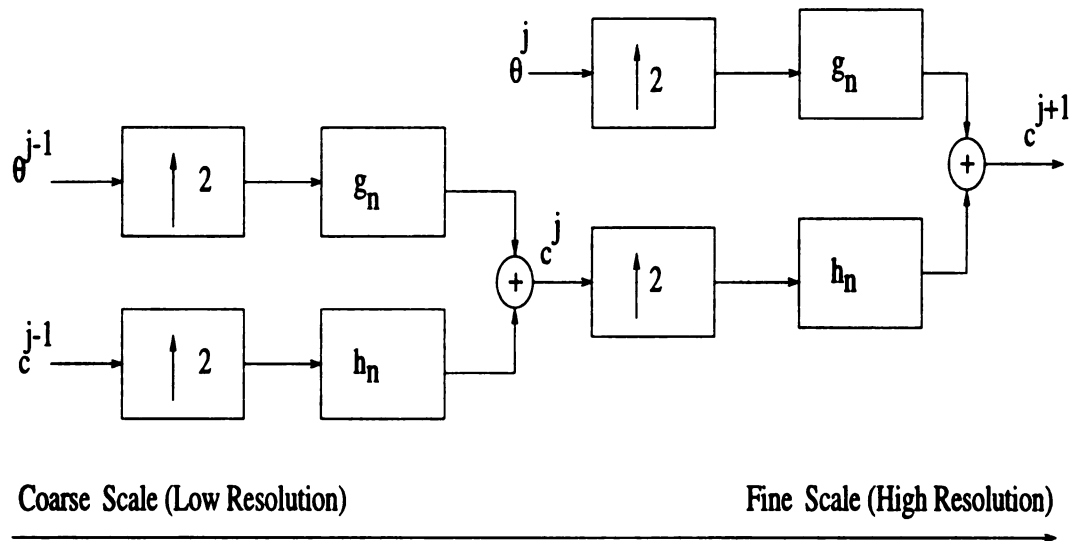


Figure 3.4. Synthesis by filter bank

### 3.2.2 Properties of the Discrete Wavelet Transform

Several attractive properties make the wavelet transform ideal for many applications in signal and image processing [15, 42]. Two key properties are *multi-resolution and locality*: each wavelet  $\psi_k^j(t)$  is only a dilated and translated version of the original mother wavelet  $\psi(t)$  and is localized simultaneously in time and frequency. A third property is the *compressive property*: the wavelet coefficients of real-world signals/images tend to be sparse. Two final properties are *clustering and persistence*: if one wavelet coefficient is large/small, then its adjacent coefficients within the same scale are very likely to also be large/small; and large/small values of wavelet coefficients tend to propagate across scales. The clustering property suggests that coefficients may have strong dependencies within scale, while persistence leads to dependencies across scale. The hidden Markov model used in Section (3.4) utilizes these properties.

It is very important to note that the above description is just *mathematical* multi-scale analysis of a signal/image. Actually, in many applications [10, 11, 40] with multi-scale analysis, it is the *physical nature* of the real signal (image) that is directly modeled via a multi-scale representation. In those cases, no wavelet transform is performed on the image at all. The motivation to do this is to utilize the nice multi-scale structure. The nicety is two-fold as summarized in Subsection (3.1.3) and is well embodied in the review of Algorithm I in Section (3.3) and Appendix V. This is a fundamental point to be kept in mind. Otherwise, when referred to multi-scale analysis, one may be misled to jump into the specific wavelet transform and gain nothing.

In other words, multi-scale analysis and wavelet analysis are not synonymous: wavelet analysis is just one form of multi-scale analysis. Of course, wavelet analysis has some benefits as well, one of which, as opposed to other multi-scale analysis

methods, is that it provides an orthogonal multi-scale decomposition that facilitates modeling and computation, as we shall see in Section (3.4).

### 3.3 Image Segmentation by Multi-scale Markov Modeling Discrete Field

In order to better understand the basic ideas of multi-scale modeling and to see clearly the distinction between Algorithm I and Algorithm II, I briefly outline the main points of Algorithm I [10] in this section. A more detailed mathematical derivation appears in Appendix VI of Chapter 5.

#### 3.3.1 Main Points of Algorithm I

First of all, let us refer to Figure (3.5) and make clear what the field  $X^j$  physically means. Each resolution is a level in a quad-tree in the 2-D case, so a lattice point at one resolution corresponds to four points at the next finer resolution. This group of four pixels in the continuous image  $Y$  is considered as a block and  $X^j$  denotes the field containing the labeling of each of the blocks at resolution  $j$ . We may assume that the finer segmentation  $X^{j+1}$  of  $X$  is an interpolated version of  $X^j$  [9]. Note in Figure (3.5),  $Y^J \equiv Y$  and  $X^J \equiv X$ .

The fundamental assumption in Algorithm I is that the sequence of  $X^j$  forms a first-order Markov chain [62], *i.e.*, the distribution of  $X^j$  given all coarser scale fields is only dependent on  $X^{j-1}$ :

$$P(x^j | x^l \quad l \leq j-1) = P_{x^j | x^{j-1}}(x^j | x^{j-1}). \quad (3.14)$$

This pyramid structure of the multi-scale random fields (MSRF) is depicted in Figure (3.5b).

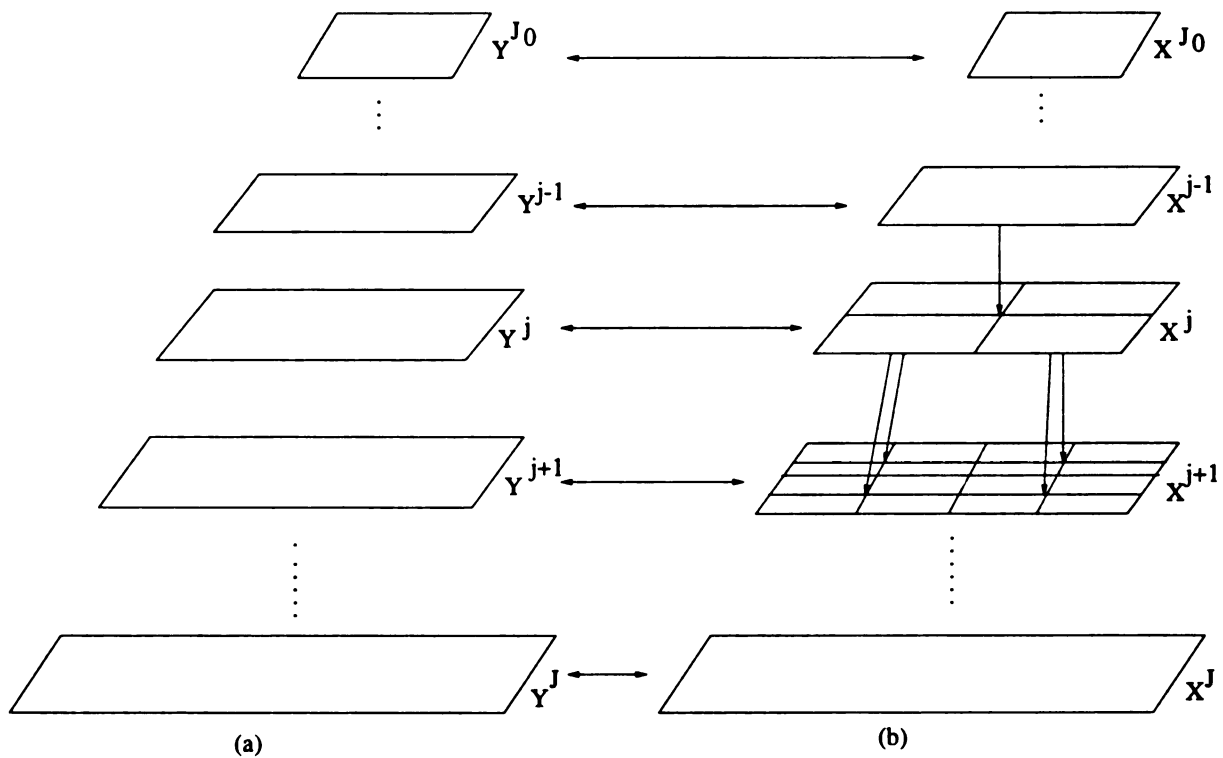


Figure 3.5. Pyramid structure of the MSRF. (a) Continuous image  $Y$  at different scales. (b) The random field  $X^j$  at each scale is causally dependent on the coarser scale field  $X^{j-1}$  above it.

The second important assumption is that each pixel  $X_k^j$  is only dependent on a local neighborhood of pixels at the next coarser scale.  $\partial k$  is used to denote the neighborhood of point  $k$ . The first choice of the neighborhood  $\partial k$  in [10] is a quad-tree structure in 2-D case or binary-tree structure in 1-D case, as depicted in Figure (3.6). Specifically, in the quad-tree structure each point is only dependent on a single point at the next coarser scale — its father  $d(k)$ . In words, if, by some means, we know that  $Y_k^j$  is active ( $X_k^j = 1$ ), then with a high probability, we may say that its four children are active as well. This probability is just the transitional probability density between individual pixels from a coarser scale to a finer scale. The transition probability that  $X_k^j$  has state  $m$  given that its father is in state  $m'$  is

$$p_{x_k^j | x_{\partial k}^{j-1}}(m|m') = q^j \delta_{m,m'} + \frac{1 - q^j}{M}. \quad (3.15)$$

This equation tells us two facts: 1) the probability that the labeling will remain the same from scale  $j - 1$  to  $j$  is  $q^j + \frac{1 - q^j}{M}$ , where  $q^j \in [0, 1]$ ; and 2) the probabilities that the child has any one of a number of different labels from its parent's are equally likely, *i.e.*,  $\frac{1 - q^j}{M}$ . The so-called “physical dependency” previously mentioned in Subsection (3.1.3) is embodied in the transition probability of the Markov chain.

This kind of choice of neighborhood structure leads to a simple and efficient algorithm — called sequential MAP (SMAP) algorithm — for the image/signal segmentation [10]. For more details, refer to Appendix VI of Chapter 5.

### 3.3.2 Simulation Result

Figure (3.7) shows the segmentation result for one 2-D image. In Figure (3.7) there are  $M = 2$  states. Each state corresponds to a Gaussian distribution, but has different mean and variance dictated by its label.

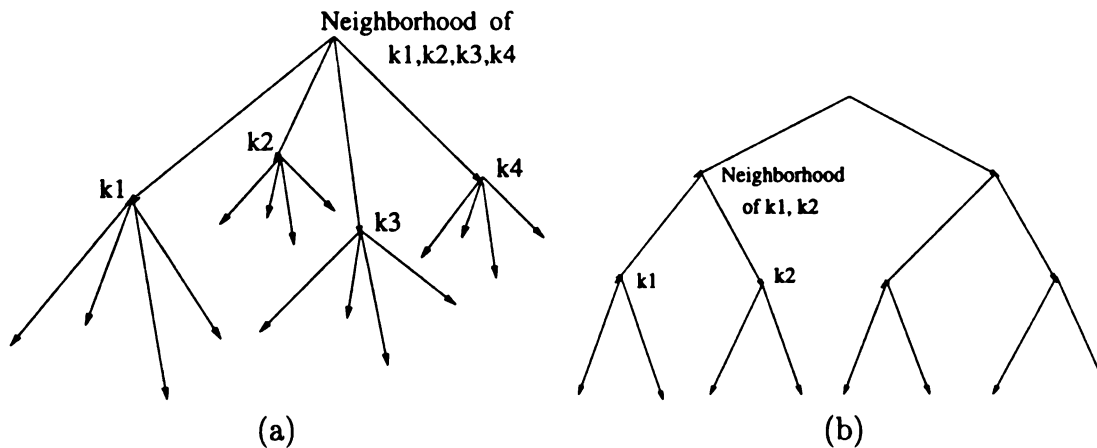
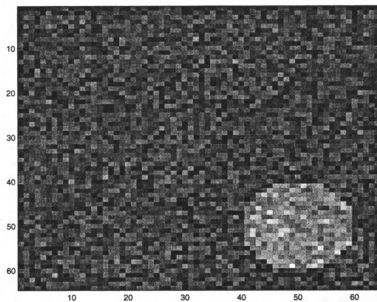


Figure 3.6. Neighborhood structure used in Algorithm I. (a) Quad-tree structure used for 2-D case; (b) Binary tree structure used for 1-D case.

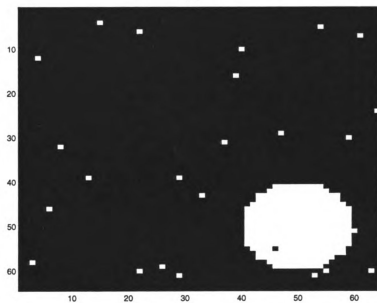
### 3.4 Image Segmentation by Multi-scale Hidden Markov Model (MHMM) of Wavelet Coefficients

An alternative to the multi-scale method of [10] based on wavelet analysis can also be used for image segmentation, which is a major contribution of this chapter. The main idea is to take advantage of the properties of DWT, as explained in Section (3.2.2). Its origination traces back to [15, 49]. The new method consists of two procedures: the first one is edge detection and the second is label estimation. In the following subsections, I will address the first step of edge detection in detail, since it is the core of our algorithm. I will then briefly explain the second step, which is much more straightforward, and provide some example applications.

The idea of applying wavelet analysis to edge detection is quite simple. Roughly speaking, wavelet coefficients represent the differences between signal/image approximations at different scales (or resolutions). Hence, they are actually a kind of differentiation, and intuitively, are well suited for edge detection.



(a)



(b)

Figure 3.7. One simulated image segmentation using Algorithm I. (a) Original noisy image; (b) Segmentation result.

### 3.4.1 Likelihood Function for fMRI Data

In this chapter, I deal exclusively with fMRI magnitude (rather than complex) time series which is approximately Gaussian distributed at reasonable SNR levels, as discussed in Chapter 2. Referring back to Equation (2.4), the correlation between the reference signal  $\mathbf{r}$  and the magnitude time series  $\mathbf{z}$ :

$$y = \mathbf{r}^T \mathbf{z} = \sum_{j=1}^N r_j z_j \quad (3.16)$$

is Gaussian distributed (provided the signal noise ratio is not very small). Once again, the reader is reminded that  $y$  in this chapter denotes correlation value.

Hence, the fMRI correlation image may be modeled as a 2-D Gaussian process. For simplicity, let us first consider the 1-D case. Extension to 2-D case is given in Subsection (3.4.7). Yielding to convention, we assume that the length of the correlation sequence is a power of 2. The observation model is:

$$y_k^J = \rho_k^J + w_k^J, \quad k = 0, \dots, 2^J - 1, \quad (3.17)$$

where  $\mathbf{y}^J \equiv \{y_k^J\}$  are the observations,  $\boldsymbol{\rho}^J \equiv \{\rho_k^J\}$  are “true” correlation values, and  $\{w_k^J\}$  are noise.

Now we are going to employ the Haar multi-scale analysis:

$$y_k^j = \frac{y_{2k}^{j+1} + y_{2k+1}^{j+1}}{\sqrt{2}}, \quad k = 0, \dots, 2^j - 1, \quad J_0 \leq j \leq J - 1.$$

The multi-scale analyses of  $\boldsymbol{\rho}$  and  $\mathbf{w}$  are defined in an analogous way. The binary tree structure of this multi-scale data analysis from scale  $j+1$  to scale  $j$  and then to scale  $j-1$  (fine-to-coarse) is shown in Figure (3.8) (adapted from [49]).

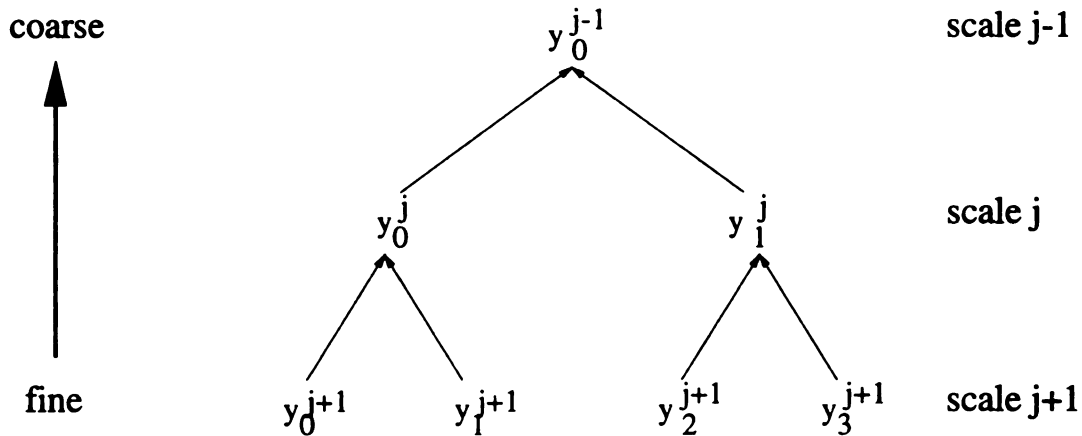


Figure 3.8. Binary data tree structure for Harr wavelet analysis.

It is straightforward to see that

$$y_k^j = \rho_k^j + w_k^j. \quad (3.18)$$

The noises  $\{w_k^j, k = 0, \dots, 2^j - 1\}$  are assumed to be independent, identically distributed (*i. i. d.*) as Gaussian random variables with zero-mean and variance  $\sigma$ . Because Haar wavelet transform is an orthonormal transform, it then follows that the preceding sentence is also true for any  $j$ , resulting in the likelihood function (refer to the above Equation (3.18)):

$$p(\mathbf{y}^j | \boldsymbol{\rho}^j) = \prod_{k=0}^{2^j-1} \mathcal{N}(y_k^j | \rho_k^j, \sigma^2), \quad J_0 \leq j \leq J \quad (3.19)$$

where  $\mathbf{y}^j \equiv \{y_k^j\}_{k=0}^{2^j-1}$  and similarly for  $\boldsymbol{\rho}^j$ ,  $\mathcal{N}(x | \rho, \sigma^2)$  denotes a Gaussian density with mean  $\rho$  and variance  $\sigma^2$  evaluated at the point  $x$ .

The relationship between a “parent” (*e.g.*,  $y_k^j$ ) and a “child” (*e.g.*,  $y_{2k}^{j+1}$ ) is very important in multi-scale data analysis. The parent-child conditional likelihood in our

case turns out to be:

$$p(y_{2k}^{j+1} | y_k^j, \boldsymbol{\rho}) = \mathcal{N} \left( y_{2k}^{j+1} | \frac{y_k^j}{\sqrt{2}} + \frac{\theta_k^j}{\sqrt{2}}, \frac{\sigma^2}{2} \right), \quad (3.20)$$

where

$$\theta_k^j = \frac{\rho_{2k}^{j+1} - \rho_{2k+1}^{j+1}}{\sqrt{2}} \quad (3.21)$$

is simply the Haar wavelet coefficient of true correlation  $\boldsymbol{\rho}$  at scale  $j$  and location  $k$ . This nice form of the likelihood suggests the use of a special kind of *a priori* model for the wavelet coefficients in Subsection (3.4.2), which complements the observation model and leads to closed-form expression for the *a posteriori* distribution of the states in Subsection (3.4.4).

The conditional density in Equation (3.20) is derived as follows.

$$\begin{aligned} y_k^j &= \frac{y_{2k}^{j+1} + y_{2k+1}^{j+1}}{\sqrt{2}} \\ \theta y_k^j &= \frac{y_{2k}^{j+1} - y_{2k+1}^{j+1}}{\sqrt{2}} \\ &= \frac{\rho_{2k}^{j+1} - \rho_{2k+1}^{j+1} + w_{2k}^{j+1} - w_{2k+1}^{j+1}}{\sqrt{2}} \\ &\equiv \theta_k^j + \theta w_k^j \end{aligned}$$

where  $\theta y_k^j$  and  $\theta w_k^j$  (read  $\theta y$  and  $\theta w$  as one symbol, not the multiplication of two symbols) are the Haar wavelet coefficients for observed correlation data  $\mathbf{y}$  and noise  $\mathbf{w}$ , respectively.

Summing up,

$$y_{2k}^{j+1} = \frac{y_k^j + \theta y_k^j}{\sqrt{2}} = \frac{y_k^j}{\sqrt{2}} + \frac{\theta_k^j}{\sqrt{2}} + \frac{\theta w_k^j}{\sqrt{2}}. \quad (3.22)$$

And since  $\boldsymbol{\rho}$  and  $\mathbf{w}$  are independent, so are  $\theta_k^j$  and  $\theta w_k^j$ . It can also be shown that  $y_k^j$  and  $\theta w_k^j$  are independent since:

$$\begin{aligned} E(y_k^j \theta w_k^j) &= \frac{1}{2} [E(w_{2k}^{j+1})^2 - E(w_{2k+1}^{j+1})^2] \\ &= \frac{1}{2} (\sigma^2 - \sigma^2) = 0. \end{aligned}$$

Since both  $y_k^j$  and  $\theta_k^j$  are independent of  $\theta w_k^j$ , and

$$\theta w_k^j \sim N(0, \sigma^2),$$

by the property of conditional likelihood [62, 61], Equation (3.20) ensues, which completes the derivation.

Further, the likelihood function in Equation (3.19) with  $j = J$  can be factorized as follows:

$$p(\mathbf{y} | \boldsymbol{\rho}) = p(\mathbf{y}^{J_0} | \boldsymbol{\rho}^{J_0}) \prod_{j=J_0}^{J-1} \prod_{k=0}^{2^j-1} p(y_{2k}^{j+1} | y_k^j, \theta_k^j), \quad (3.23)$$

where  $J_0$  is the coarsest scale for the analysis (usually we use  $J_0 = 0$ ),  $p(y_{2k}^{j+1} | y_k^j, \theta_k^j)$  is given by Equation (3.20) and  $p(\mathbf{y}^{J_0} | \boldsymbol{\rho}^{J_0})$  is given by Equation (3.19) with  $j = J_0$ . Note that  $\rho_0^0$  is the global average correlation data.

The factorization follows from the following lines of reasoning. Let us refer to Figure (3.9), which is an example of three scales. The key point is that the information contained in data at the finest scale  $\{y_0^2, y_1^2, y_2^2, y_3^2\}$  is completely the same as that in  $\{y_0^0, y_0^1, y_0^2, y_2^2\}$  (corresponding to the white dots in Figure (3.9)), therefore,

$$\begin{aligned} p(\mathbf{y}^2 | \boldsymbol{\rho}) &= p(y_0^0, y_0^1, y_0^2, y_2^2 | \boldsymbol{\rho}) \\ &= p(y_0^0 | \boldsymbol{\rho}) p(y_0^1, y_0^2, y_2^2 | y_0^0, \boldsymbol{\rho}) \end{aligned}$$

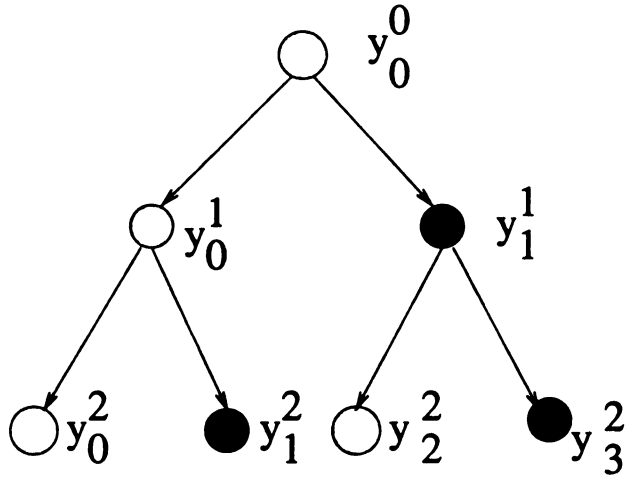


Figure 3.9. Propagation of wavelet coefficients (three scales)

$$\begin{aligned}
 &= p(y_0^0|\boldsymbol{\rho})p(y_0^1|y_0^0, \boldsymbol{\rho})p(y_0^2, y_1^2|y_0^0, y_0^1, \boldsymbol{\rho}) \\
 &= p(y_0^0|\boldsymbol{\rho})p(y_0^1|y_0^0, \boldsymbol{\rho})p(y_0^2|y_0^0, y_0^1, \boldsymbol{\rho})p(y_1^2|y_0^0, y_0^1, \boldsymbol{\rho}) \\
 &= p(y_0^0|\boldsymbol{\rho}_0^0)p(y_0^1|y_0^0, \boldsymbol{\rho})p(y_0^2|y_0^1, \boldsymbol{\rho})p(y_1^2|y_1^1, \boldsymbol{\rho})
 \end{aligned} \tag{3.24}$$

due to Equation (3.20), for example:

$$p(y_0^1|y_0^0, \boldsymbol{\rho}) = \mathcal{N}(y_0^1|y_0^0, \frac{\theta_k^j}{\sqrt{2}}).$$

Generalization of Equation (3.24) leads to Equation (3.23).

### 3.4.2 Key Point for Edge Detection and *a Prior* Distribution for the Wavelet Coefficients

Now let us consider (joint) *a prior* probability for the (unknown) wavelet coefficients  $\boldsymbol{\theta}$ . A simple approach is to model them as independent Gaussian mixture random variables. We move beyond this simple *a prior*, by specifying probabilistic dependencies between the *states* underlying the mixtures of parent and child wavelet coefficients.

To deduce discrete state estimation from continuous data, the *key point* for our algorithm is to associate the *continuous* wavelet coefficients with a 2-state (*discrete*) Markov chain as depicted in Figure (3.10), *i.e.*, each wavelet coefficient is described by a Gaussian Mixture Model (GMM). In Figure (3.10), each black node represents a continuous-valued wavelet coefficient  $\theta_k^j$ ; each white node represents the discrete hidden state variable  $s_k^j$  for the corresponding wavelet coefficient  $\theta_k^j$  (connected by a solid line to the state variable  $s_k^j$ ). To match the inter-scale coefficient dependencies, the hidden states are vertically linked across scale by dashed lines. Connections across scale capture the “parent-child” dependency inherent in the DWT of natural signals/images [49].

For our real problem of edge detection, the states of Markov chain are unknown (“hidden”) and represent the presence or absence of edges: state 0 indicates a homogeneous region; state 1 represents the existence of an edge. If we believe that the underlying signal is generally smooth with a few large edges, then the following modeling is intuitively reasonable. Specifically, consider two-state mixture model where state ‘0’ is a highly probable low-variance Gaussian density, indicative of a homogeneous region, while state ‘1’, corresponding to another less likely Gaussian density with a larger variance, indicates the presence of an edge (non-smooth area). Using this interpretation, we may *test for the presence of an edge simply by checking whether or not the following condition holds*:

$$p(s_k^j = 1|\mathbf{y}) > p(s_k^j = 0|\mathbf{y}), \quad (3.25)$$

If it holds, then we conclude there is an edge at scale  $j$  and location  $k$ .

Keeping these basic ideas in mind, let us now turn to a formal mathematical description. The MHMM is based on the modeling assumption that the value of each state  $s_k^j$  is *caused* by the value of its parent state  $s_{[k/2]}^{j-1}$ . This leads to the factorization

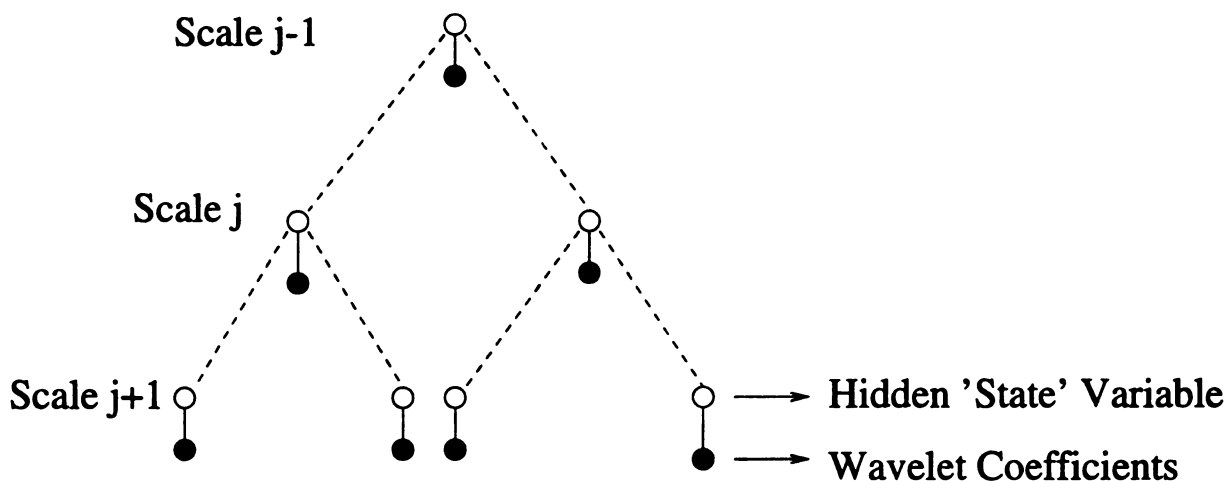


Figure 3.10. Wavelet-based HMM.

of the joint state probability function:

$$p(\mathbf{s}) = \prod_{j=J_0}^{J-1} \prod_{k=0}^{2^j-1} p(s_k^j | s_{\lfloor k/2 \rfloor}^{j-1}), \quad (3.26)$$

where  $\mathbf{s} \equiv \{s_k^j\}_{k=0,1,\dots,2^j-1}^{j=0,1,\dots,J-1}$  and  $p(s_0^0 | s_0^{-1}) \equiv p(s_0^0)$ . At the coarsest scale  $j = 0$ , there is no parent wavelet coefficient and so a *prior* is introduced for the state  $s_0^0$  of the wavelet coefficient  $\theta_0^0$ :  $\varrho_0^0(m) \equiv p(s_0^0 = m)$ .

Another property of HMMs [16, 52] in general is that, given their respective state values, all parameters  $\boldsymbol{\theta}$  are conditionally independent, which is also implied by the assumption of the model. That is,

$$p(\boldsymbol{\theta} | \mathbf{s}) = \prod_{j=J_0}^{J-1} \prod_{k=0}^{2^j-1} p(\theta_k^j | s_k^j), \quad (3.27)$$

where  $p(\theta_k^j | s_k^j)$  is assumed to be Gaussian as explained previously:

$$p(\theta_k^j | s_k^j = m) = \mathcal{N}(\theta_k^j | \mu_m^j, \tau_m^{j2}). \quad (3.28)$$

We regard the signal and its wavelet coefficients as realizations of a zero-mean random signal. Therefore, we assume  $\mu_m^j = 0$  for all  $m$  and  $j$ . In general, the variances  $\tau_m^{j2}$  are scale ( $j$ )-dependent, but in our experiment we set them the same at different scales. In this case, there are only two parameters  $\tau_0^2$  and  $\tau_1^2$ .

### 3.4.3 A Failure Modeling

An acute reader might raise the question: why do you abandon Algorithm I and proceed to think about Algorithm II? The first reason is that I need develop something new for my dissertation. The second reason is that actually we made a big mistake before proposing this two-step approach for image segmentation (edge detection followed by state estimation) in its present form.

At this point, I would like to further emphasize the meaning of  $X_k^j$  and its modeling in Equation (3.15) in Algorithm I. Algorithm I *places label  $X_k^j$  directly on some abstract of scaling coefficient  $Y_k^j$*  — refer back to Figure (3.5)). This is a crucial virtue and stands as a sharp contrast with our Algorithm II, which *places states on wavelet coefficients*. Since the states in Algorithm I reflects directly the image's classification labels, it is easy to physically understand the transition of states between upper and lower scales (see Equation (3.14) and Equation (3.15)), and thus solve *image segmentation* problem directly. Another feature of Algorithm I, as already stated in Subsection (3.2.2), is that it does not resort to any DWT at all, which is another contrast with our Algorithm II.

During the second stage of my research, at first we also hoped to segment the correlation image directly in one step to active and nonactive regions (instead of presently used edge detection and label estimation). The fundamental point was to deduce the labels of image pixels from the states of wavelet coefficients. The modeling was to use the state of wavelet coefficient at the next coarser scale to represent the labels of two image pixels at the next finer scale which correspond to that wavelet

coefficients; mathematically, to associate  $s_k^j$ , the state of wavelet coefficient  $\theta_k^j$  at scale  $j$ , with one of four values, depending on the two corresponding image pixels at the next finer scale  $j + 1$ :

$$s_k^j = \begin{cases} (0, 0) & \rho_{2k}^{j+1}, \rho_{2k+1}^{j+1} \text{ inactive (small)} \\ (0, 1) & \rho_{2k}^{j+1} \text{ inactive, } \rho_{2k+1}^{j+1} \text{ active} \\ (1, 0) & \rho_{2k}^{j+1} \text{ active, } \rho_{2k+1}^{j+1} \text{ inactive} \\ (1, 1) & \rho_{2k}^{j+1}, \rho_{2k+1}^{j+1} \text{ active (large)} \end{cases} \quad (3.29)$$

We also suspected that this approach is equivalent to Algorithm I. But after several months' trial, it turned out that our initial conjecture was wrong. The underlying reason is that *the states of wavelet coefficients DO NOT have the same physical interpretation as to represent the label of image pixels in Algorithm I*. Further, according to the compressive property of DWT, the states of wavelet coefficients should be just a few. However, think about the general situation of image segmentation in which there are  $M > 2$  possible labels for each image pixel. According to above modeling in Equation (3.29), there would be  $M^2$  labels for each wavelet coefficient, which is contradictory to the compressive property.

That conjecture cost me several months' time on analytical formulation and numerical simulations. However, our efforts ended up in vain. "Failure is the mother of success". So later I changed the meaning of the states of wavelet coefficients to that as described in the preceding Subsection (3.4.2), and accordingly changed Algorithm II to two-step approach: edge detection followed by label estimation. The successful results are shown in Subsection (3.4.8) and Section (3.5).

### 3.4.4 Solution for Joint *a Posteriori* State Probability

Let us continue our discussion with Algorithm II. Having set up the formulations for likelihood and *a prior*, we are now ready to determine the joint *a posteriori* density of the states  $\mathbf{s}$  given observations of correlation data  $\mathbf{y}$ . Note that:

$$\begin{aligned}
p(\mathbf{s} = \mathbf{m}|\mathbf{y}) &= \int p(\mathbf{s} = \mathbf{m}, \boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \\
&\propto \int p(\mathbf{y}|\mathbf{s} = \mathbf{m}, \boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{s} = \mathbf{m})p(\mathbf{s} = \mathbf{m})d\boldsymbol{\theta} \\
&= \prod_{j=J_0}^{J-1} \prod_{k=0}^{2^j-1} \int p(y_{2k}^{j+1}|y_k^j, \theta_k^j, s_k^j = m_k^j)p(\theta_k^j|s_k^j = m_k^j) \\
&\quad p(s_k^j = m_k^j|s_{\lfloor k/2 \rfloor}^{j-1} = m_{\lfloor k/2 \rfloor}^{j-1})d\theta_k^j \\
&= \prod_{j=J_0}^{J-1} \prod_{k=0}^{2^j-1} p(s_k^j = m_k^j|s_{\lfloor k/2 \rfloor}^{j-1} = m_{\lfloor k/2 \rfloor}^{j-1})L_k^j(s_k^j = m_k^j) \quad (3.30)
\end{aligned}$$

where  $m_k^j$  is one particular (deterministic) value assumed by random state variable  $s_k^j$  and  $L_k^j(s_k^j = m) \propto p(y_{2k}^{j+1}|y_k^j, s_k^j = m)$ , the essential ingredients for our estimation of the *a posteriori* states, are actually marginal likelihoods. From the likelihood function in Equation (3.20) and the *a prior* in Equation (3.28), we derive them to be:

$$\begin{aligned}
L_k^j(m) &= \int p(y_{2k}^{j+1}|y_k^j, \theta_k^j, s_k^j = m)p(\theta_k^j|s_k^j = m)d\theta_k^j \\
&\propto p(y_{2k}^{j+1}|y_k^j, s_k^j = m) \\
&= \mathcal{N}\left(y_{2k}^{j+1}|\frac{\mu_m^j}{\sqrt{2}} + \frac{y_k^j}{\sqrt{2}}, \frac{\tau_m^{j2} + \sigma^2}{2}\right), \\
&\quad J_0 \leq j \leq J-1, \quad k = 0, 1, \dots, 2^j-1, \quad m = 0, 1. \quad (3.31)
\end{aligned}$$

**Proof:** First recall Equation (3.22):

$$y_{2k}^{j+1} = \frac{y_k^j + \theta y_k^j}{\sqrt{2}} = \frac{y_k^j}{\sqrt{2}} + \frac{\theta y_k^j}{\sqrt{2}}.$$



where the first two terms are independent of the third one.

Further, recall Equation (3.28):

$$p(\theta_k^j | s_k^j = m) = \mathcal{N}(\theta_k^j | \mu_m^j, \tau_m^{j2}),$$

i. e.,

$$\begin{aligned} \theta w_k^j &\sim N(0, \sigma^2) \\ \theta_k^j | s_k^j = m &\sim N(\mu_m^j, \tau_m^{j2}) \end{aligned}$$

Therefore,

$$\begin{aligned} E(y_{2k}^{j+1} | y_k^j, s_k^j = m) &= \frac{\mu_m^j}{\sqrt{2}} + \frac{y_k^j}{\sqrt{2}} \\ Var(y_{2k}^{j+1} | y_k^j, s_k^j = m) &= \frac{\tau_m^{j2} + \sigma^2}{2} \end{aligned}$$

So,  $L_k^j(m) \propto p(y_{2k}^{j+1} | y_k^j, s_k^j = m)$  (where  $y_k^j$  is regarded as constant), one marginal density function, has the closed-form representation in Equation (3.30). **QED.**

### 3.4.5 Marginal *a Posteriori* State Probability Calculation

After determining the *joint a posteriori* state probability, we can use an upward-downward probability propagation algorithm [15] to determine the marginal *a posterior* probability of state  $s_k^j$  for the wavelet coefficients  $\theta_k^j$ , and then use Equation (3.25) to test the presence of an edge.

In the upward-downward algorithm, the **Up Step** marginalizes the *a posterior* state probability recursively from the finest scale  $j = J - 1$  to the coarsest scale  $j = 0$ . At the end the *a posterior* state probabilities  $\{p(s_0^0 = m | \mathbf{y})\}_{m=0}^{M-1}$  are provided and partial marginalizations are also stored for use in the Down Step. The **Down**

**Step** computes the marginal *a posterior* state probabilities for each  $s_k^j$  recursively. Specifically, the upward-downward algorithm goes as follows [15, 49]. In the following,  $\varrho_k^j(m|n) \equiv p(s_k^j = m | s_{\lfloor k/2 \rfloor}^{j-1} = n)$  and  $\varrho_0^0(m) \equiv p(s_0^0 = m)$ .

### Upward-Downward Propagation Algorithm

#### Up Step

Beginning at  $j = J - 1$ , compute

$$q_k^j(n) = \sum_{m=0}^{M-1} \varrho_k^j(m|n) L_k^j(m), \quad (3.32)$$

Then for  $j = J - 2, \dots, 1$

$$q_k^j(n) = \sum_{m=0}^{M-1} q_{2k}^{j+1}(m) q_{2k+1}^{j+1}(m) \varrho_k^j(m|n) L_k^j(m) \quad (3.33)$$

and for  $j = 0$

$$q_0^0(n) = q_0^1(n) q_1^1(n) \varrho_0^0(n) L_0^0(n). \quad (3.34)$$

The final quantities  $\{q_0^0(m)\}$  are the (unnormalized) posterior state probabilities  $\{p(s_0^0 = m | \mathbf{y})\}_{m=0}^{M-1}$ .

#### Down Step

Beginning the posterior states probabilities at scale 0, set  $p_0^0(m) = q_0^0(m)$ . Then for  $j = 1, \dots, J - 2$

$$p_k^j(m) = \sum_{n=0}^{M-1} \frac{p_{\lfloor k/2 \rfloor}^{j-1}(n) \varrho_k^j(m|n) q_{2k}^{j+1}(m) q_{2k+1}^{j+1}(m) L_k^j(m)}{q_k^j(n)} \quad (3.35)$$

and for  $j = J - 1$

$$p_k^j(m) = \sum_{n=0}^{M-1} \frac{p_{\lfloor k/2 \rfloor}^{j-1}(n) \varrho_k^j(m|n) L_k^j(m)}{q_k^j(n)}. \quad (3.36)$$

The final quantities  $\{p_k^j(m)\}_{m=0}^{M-1}$  are the desired marginal *a posteriori* state probabilities  $\{p(s_k^j = m|\mathbf{y})\}_{m=0}^{M-1}$ .

This determines all the marginal *a posteriori* state probability. Then according to the criterion in Equation (3.25), if the *a posteriori* probability that state  $s_k^j$  is 1 is greater than the *a posteriori* probability that state  $s_k^j$  is 0, then we decide that there is an edge between  $\rho_{2k}^{j+1}$  and  $\rho_{2k+1}^{j+1}$ . Note, in this formulation we actually adopt the *Maximum Marginal a Posteriori* criterion.

### 3.4.6 Image Label Estimation

After the edges are determined, it is straightforward to formulate likelihood ratio test to estimate the label of each homogeneous region.

Consider the following multi-hypothesis problem. The observation data  $\mathbf{y} = [y_1 \ y_2 \ \cdots \ y_n]^T$  within each homogeneous region is Gaussian random vector of dimension  $n$ . The  $M$  hypotheses are

$$H_i : \mathbf{y} \sim N(\mathbf{m}_i, \mathbf{C}_i), \quad i = 1, 2, \dots, M, \quad (3.37)$$

where  $\mathbf{m}_i$  and  $\mathbf{C}_i$ , which are assumed known, are the mean vector and covariance matrix of the observation under the  $i$ th hypothesis ( $i = 1, 2, \dots, M$ ). Suppose each hypothesis is equally likely and minimum error criterion is adopted [61], the decision rule then boils down to

$$\text{choosing } H_j \text{ where } j = \arg \min_i \|\mathbf{y} - \mathbf{m}_i\|^2 + \ln|\mathbf{C}_i|, \quad (3.38)$$

where  $\|\mathbf{y} - \mathbf{m}_i\|^2 \equiv (\mathbf{y} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{y} - \mathbf{m}_i)$  here and  $|\mathbf{C}_i|$  is the determinant of  $\mathbf{C}_i$ .

In the following numerical simulation, the observed data within each homogeneous region are assumed to be  $i$ .  $i$ .  $d$ ; that is, in each region it is assumed

that  $\mathbf{m}_i = m_i \mathbf{1}$  and  $\mathbf{C}_i = \sigma_i^2 \mathbf{I}$ . Some important practical problems are the assignment of the *a priori* probability  $\varrho_0^0(m)$  for  $s_0^0$  and state transition probabilities  $\varrho_k^j(m|n) = p(s_k^j = m | s_{\lfloor k/2 \rfloor}^{j-1} = n)$ , as well as the variances  $\tau_0^2$  and  $\tau_1^2$  characterizing Gaussian mixture density in Equation (3.28). Selection of the parameters for the Gaussian distribution characterizing each homogeneous region is also crucial. These parameters may be estimated by a complicated E-M algorithm [61]. For the initial investigation, I set them empirically (by observation). It turns out that our experimental results are insensitive to the *a priori* probability  $\varrho_0^0(m)$  and transition probability  $\varrho_k^j(m|n)$ . This apparent robustness is a nice feature.

### 3.4.7 Extension to 2 Dimensions

The preceding descriptions in this section are all confined to 1-D case. Direct application of the above procedures (edge detection followed by image segmentation) in 1-D to 2-D is not very easy. Recall the core idea of our Algorithm II is to use the states of wavelet coefficients as indicators of edges. This is not easily extendable to 2-D images, since we have three sets of wavelet coefficients at each scale for 2-D DWT, reflecting signal intensity changes in three orientations (horizontal, vertical and diagonal) [53, 54]. How to use the wavelet coefficients to represent edges of 2-D images in our general framework is not straightforward and needs future consideration.

We can, however, extend the multi-scale analysis and MHMMs from 1-D sequence to 2-D images by the following method. Instead of taking the usual 2-D wavelet transform of the original image, we use the following conversion method. First we convert the original 2-D image into 1-D sequence, and then apply previous 1-D wavelet analysis to the resulting sequence. The conversion details are: first split the image vertically into two halves, then horizontally splitting each half into two quarters, and reiterate until each one is a  $1 \times 1$  pixel. Refer to Figure (3.11) for details. The merit of this conversion is that it retains the original spatial configuration. And by

|    |    |    |    |
|----|----|----|----|
| 1  | 2  | 3  | 4  |
| 5  | 6  | 7  | 8  |
| 9  | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 |

|   |   |   |   |   |    |    |    |   |   |   |   |    |    |    |    |
|---|---|---|---|---|----|----|----|---|---|---|---|----|----|----|----|
| 1 | 5 | 2 | 6 | 9 | 13 | 10 | 14 | 3 | 7 | 4 | 8 | 11 | 15 | 12 | 16 |
|---|---|---|---|---|----|----|----|---|---|---|---|----|----|----|----|

Figure 3.11. Conversion of a 2-D image to 1-D sequence

this conversion method, the essential computations are performed with regard to 1-D sequence and thus quite affordable.

### 3.4.8 A Simulation of the New Segmentation Method

Figure (3.12a-c) shows a simulated noisy image, the detected edges and the gray level of the segmented image by Algorithm II, respectively. A two-state MHMM is specified for this problem with the following parameter settings:

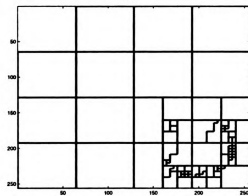
$$\begin{aligned}
\tau_0^2 &= 1, \\
\tau_1^2 &= 100, \\
\varrho_0^0(0) &= .9, \\
\varrho_k^j(0|0) &= .9, k = 0, \dots, 2^j - 1, j = 1, \dots, J - 1, \\
\varrho_k^j(0|1) &= .25, k = 0, \dots, 2^j - 1, j = 1, \dots, J - 1.
\end{aligned}$$

Figure (3.12c) demonstrates that the overall result is excellent.

In Figure (3.12b), there are some artificial edges (boundaries), because, for the sake of numerical stability, the whole image is divided into 16 subimages and our algorithm is actually applied to each sub-image. This parsing also brings another advantage: greatly reducing false edges in the final segmented image. For example, the region outside the brain usually has different statistical behavior as that inside



(a)



(b)



(c)

Figure 3.12. One simulated image segmentation by Algorithm II. a) Noisy image; b) Detected edges; c) Segmented image.

the brain. By this parsing, these two regions are almost treated separately, thereby greatly reducing boundaries between brain and air, which are most likely to be false edges (*i.e.*, edges beyond our interests of activeness). See the following fMRI data processing example in Figure (3.13).

### 3.5 Processing Results for fMRI Data by Two-Step Approach

As stated in the beginning of this chapter, the method for fMRI detection in this chapter involves a two-step procedure: multi-scale image segmentation will be first used to break the correlation image into different regions of homogeneous statistical behavior, each region will then be tested independently as active or inactive by single pixel detection method.

In order to see the potential of this method for fMRI detection, the following experiment is conducted to compare results from the combined effects of single pixel detection and image segmentation with results obtained in the last chapter based solely on pixel-wise detection.

Using the model introduced earlier in Equation (2.1), a simulated fMRI complex time series is generated at each pixel. In order to simulate the profile of the brain, the magnitudes of the baseline signal ( $a$ 's in Equation (2.1) in the complex time series roughly follow the magnitude data from a static brain image. Actually the original complex data used in this example are the same as those used in Figure (2.5). Next, the correlation value at each pixel is computed by correlating the magnitude time series with the reference — see Equation (3.16) — to produce Figure (3.13a).

Figure (3.13b) is the segmented result of the correlation image in Figure (3.13a) based on our Algorithm II in this chapter. There are  $M = 2$  labels: each pixel is

assigned to either 0 or 1 according to its label. The parameters in this example are set to be:

$$\begin{aligned}
\sigma^2 &= 1 \\
\tau_0^2 &= 1; \\
\tau_1^2 &= 100; \\
\varrho_0^0(0) &= .95; \\
\varrho_k^j(0|0) &= .95, k = 0, \dots, 2^j - 1, j = 1, \dots, J - 1; \\
\varrho_k^j(0|1) &= .05, k = 0, \dots, 2^j - 1, j = 1, \dots, J - 1; \\
m_0 &= 0; \\
m_1 &= 2.
\end{aligned}$$

In this example, I set  $\sigma_1$  and  $\sigma_0$  (variances for the Gaussian distributions characterizing two homogeneous regions) to be equal. The test criterion in Equation (3.38) reduces to simpler form in this case. The original simulated active region in Figure (2.5a) is a 9\*9 SQUARE (the coordinates are:  $y = 40, 41, \dots, 48; x = 34, 35, \dots, 42$ ); in Figure (3.13b) the white region is 10\*8 RECTANGLE (the coordinates are:  $y = 41, \dots, 48; x = 33, 34, \dots, 42$ ). They are in good agreement but not in perfect match. This is not surprising, since, in general, we cannot guarantee the segmentation step produces exactly the same GEOMETRY as the original simulated regions.

Next I consider applying single pixel detection technique in Chapter 2 to each of the above homogeneously (statistically) distributed region. The idea is to regard each homogeneous region as one large, *macro*-pixel: the average of all time series inside each *macro*-pixel is taken to be the new time series characterizing this *macro*-pixel; then apply single pixel detection (MC detection) method to the new time series individually to determine which of these macro-pixels is active and which one

is inactive. By this approach, the *micro-pixels* (original pixels in Figure (3.13a), in contrast to *macro-pixel*) corresponding to the black region in Figure (3.13b) are all inactive, which is expected since this region contains a large area outside the brain. The micro-pixels in Figure (3.13a) corresponding to the white region in Figure (3.13b) turn out to be all active. In other words, by this approach, only 9 pixels inside the square are missed while the 8 pixels outside the square are false-alarmed.

Now let us take a comparison between Figure (2.5) in Chapter 2 and Figure (3.13b) in this chapter. Recalling the results in last chapter, we see spurious activation regions outside the brain. However, the falsely alarmed regions disappear in Figure (3.13b) (except for 8 pixels outside the square) after combining image segmentation with single pixel detection. Pure single pixel detection methods failed to detect some active pixels inside the small square in Figure (2.5a). However, these regions (except for 9 pixels) are now correctly detected by combining image segmentation with single pixel detection.

The enhancement of detection efficiency is clearly visible and also easily understandable. Actually we are given spatial-temporal series. However, the pixel-wise detection method used in the last chapter only takes temporal information into account: spatial information is completely ignored. The image segmentation algorithm in this chapter exactly complements the pixel-wise detection and remedies its shortcoming: it utilizes the spatial correlation information inherent in the data. So it is no wonder that the detection performance improves after image segmentation.

To see the influence of different parameter setting on detection results, I produce another correlation image and corresponding detection map in Figure (3.13c) and Figure (3.13d). Figure (3.13c) is the correlation image produced completely by the same procedure but with  $a/\sigma = 10$  ( $a$  and  $\sigma$  here are two parameters in our model (2.1)). If we use the previous set of parameters, the detection result is the same as in Figure (3.13b). However, Figure (3.13d) is the corresponding detection map achieved

under a different set of parameters:  $\sigma^2 = 4, \tau_0^2 = 2, \tau_1^2 = 200$  (other parameters are the same as those used for (3.13b)). The detected region in this case is 8\*8 square (the coordinates are:  $y = 41, \dots, 48; x = 33, 34, \dots, 40$ ). The performance from this set of parameters is inferior to that from the previous set of parameters.

To my knowledge, the idea of multi-scale detection has not been applied to fMRI data processing yet, and therefore the method in this chapter is quite original and is promising to future real fMRI data processing.

One last point I'd like to make is that in Figures (3.13)b and (3.13)d the brain profiles are artificially overlapped, as are the cases in Figure (2.5).

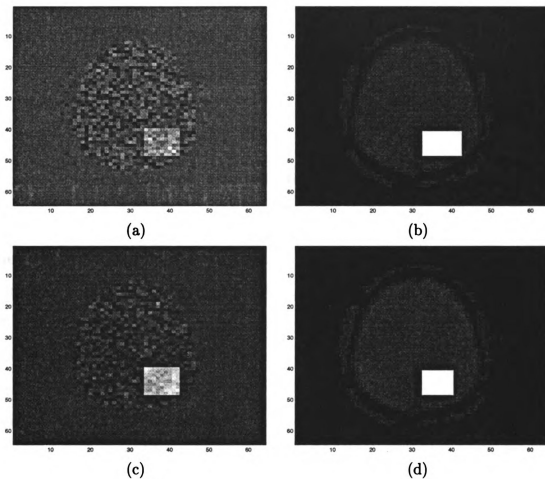


Figure 3.13. Processing results for fMRI data by two-step approach. a) One fMRI correlation image; b) Segmented image of (a), also final detection results by combinational use of image segmentation and single pixel detection; (c) Another correlation image; (d) Segmentation and detection result of (c).

# CHAPTER 4

## Discussions and Conclusions

### 4.1 For Pixelwise Detection

In Chapter 2, a novel nonlinear GLRT detector for fMRI using complex data is developed, and is compared to the commonly used MC test and the recently proposed CC test. The test statistic for the nonlinear GLRT detector has a closed-form expression. All three tests are roughly equal in terms of computational complexity. Theoretical analysis establishes an invariance property for the test statistic, and it is shown that the GLRT and the CC test are asymptotically equivalent (as the length fMRI time series increases). Monte Carlo analysis is used to demonstrate that the GLRT performs better than the MC test or CC test overall. Furthermore, the analysis reveals that a desired  $P_f$  can be achieved with the GLRT using thresholds selected from well-known distribution tables. The distributions of GLRT statistic at high baseline signal intensity under two hypotheses are approximated as well.

There are several avenues for future work within the GLRT framework. First, the noise structure in fMRI is very complicated. For simplicity and the purpose of demonstrating our method and ideas, we assume the noise is white and Gaussian. The whiteness assumption does not change the problem essentially, since given a *known* time-correlation structure we can always use the Choleksy factorization of the

noise covariance to whiten the data [56], producing a model with the same form as that used in Chapter 2. Hence, many of our conclusions are easily extended to more realistic noise models that incorporate random fluctuations due to the respiration and cardiac cycle and patient motions [5, 29, 38], provided that these components are known.

Second, more realistic (and necessarily more complicated) signal models can be used in the GLRT framework. For example, multi-parameter models of the reference signal  $\mathbf{r}$  could account for uncertainties in the BOLD response. Multi-parameter linear regression models of the response could be used within the GLRT framework to make the test more robust to such uncertainties.

Some difficulties that we face, however, are 1) the distribution of noise  $\mathbf{n}$  is usually unknown *a priori* — not as nice as Gaussian model which is adopted in this dissertation, and estimation of the noise covariance is a challenging issue even if it is Gaussian distributed; 2) the structures of nuisance components are quite spatially-varying and it is hard to distinguish the signal components from the nuisance components. How to determine adaptively the signal and nuisance components from real data is an important issue.

Actually, when I was studying on our nonlinear model, of course I pondered on the linear model in Equation (2.3). Specifically, the problem under my consideration was how to determine the best representations for signal subspace  $H$  and nuisance subspace  $S$  for (magnitude) time series directly from actual data. Ardekani *et al* went one step ahead of me. In [2], they partially solved the problem under my consideration (he also dealt with magnitude time series), that is, he devised one method for determining the best representation of nuisance component assuming the signal component has a known form. Completely solving the problem remains an important issue.

Finally, we close with a summary of our conclusions regarding complex domain

fMRI. First, at relatively high baseline signal intensity ( $a/\sigma > 3$ ), the simple MC test, which is very common in practice, performs quite well. Hence, in such regimes there is no compelling reason for testing based on the complex data. This is expected since the magnitude data is approximately Gaussian at high signal intensity, in which case the MC test is nearly optimal. In fact, in most typical fMRI experiments  $a/\sigma > 3$  and the MC test is adequate. However, at lower baseline signal intensity the performance of the MC test drops off dramatically, and in such situations complex data tests such as the new GLRT and CC test offer superior performance. Low signal intensity does occur as the spatial and/or temporal resolution of the fMRI study is increased. Most fMRI experiments work with limited resolution in order to avoid the low signal intensity problem. However, high resolution, low signal intensity fMRI may be useful in certain research or clinical paradigms, and in such cases we advocate the GLRT.

## 4.2 Consideration on Spatial Information

In the detection method of Chapter 3 I first segment image composed of correlation data which are assumed to be Gaussian. I then apply single pixel detection method to each homogeneous region. One disadvantage in real data processing is that we do not know *a priori*  $M$ , the definite number of homogeneous regions to which the correlation image is to be classified. We may turn the wheel around: first apply pixel-wise detection to get the values of test statistics at all pixels, and then apply image segmentation algorithm to decide active and non-active regions. One advantage in this direction is that we have a definite number of labels ( $M = 2$ ) when we perform image segmentation. However, in this case, the data (test statistics) are most likely to be  $F$  distributed and the nice factorization for likelihood function and parent-child transition in Chapter 3 breaks down. Hence applying ideas in our Algorithm II is much more difficult and entails further consideration in this case.

Considering utilizing spatial correlation information, this dissertation uses Bayesian image segmentation method. One practical difficulty in real data processing is that it is hard to determine and incorporate the *a priori* distribution. Other approaches involving spatial consideration can be used as well. For example, clustering analysis are gaining more recognition in this field [20, 26, 27]. Borrowing some ideas from array signal processing [47, 58, 70] may also be beneficial to fMRI detection. Formulation as decentralized detection problem is also a promising candidate [30, 51].

### 4.3 Epilogue

As stated in the very beginning of this dissertation, fMRI involves a lot of background in physiology, neurology. This dissertation deals with pure signal and image processing, because I lack *a priori* knowledge about the spatial varying nature of fMRI time series, the feature of physiological respiration, machine and/or head motion artifacts, etc. To fully validate and further refine the methodologies developed in this dissertation, comprehensive testing and evaluation with real fMRI data is necessary.

In view of the great complexity of fMRI data processing, insight and expertise of experts from other fields are very valuable — even indispensable — for successful, practical research on real data processing. However, one biggest disadvantage during my research is that I did not have enough communication and collaboration with specialists from other fields, let alone control over specific experiment. This is exactly what hindered me from real data processing in this dissertation. I hope these drawbacks can be remedied later in the future.

# CHAPTER 5

## Appendices

### 5.1 Appendix I: Notations and Some Results on Projection Matrix

Some mathematical conventions and notations used throughout this dissertation are established here.

All norms are the standard (Euclidean) 2-norm. Given two real one-dimensional (1-D) function  $f(t), t \in R$  and  $g(t), t \in R$ , their *inner product* is defined as

$$\langle f, g \rangle \equiv \int_R f(t)g(t)dt.$$

The norm of  $f$  is defined as

$$\| f \| = \sqrt{\langle f, f \rangle}.$$

Given a sequence  $h_n, -\infty < n < \infty$ , its norm is defined as

$$\| h \| = \sqrt{\sum_n h_n^2}.$$

$L^2(R)$  is defined to be the set of  $\{f \mid \|f\| < \infty\}$ .  $l^2(Z)$  is defined to be the set of  $\{h_n \mid \|h\| < \infty\}$ .

Given a real column vector  $\mathbf{x}$  of dimension  $N$ , its *norm* is defined as

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T \mathbf{x}} = \sqrt{\sum_{i=1}^N x_i^2}.$$

Let  $M$  denote an  $p \times q$  matrix. Let  $P_M$  denote the matrix that projects a vector onto the subspace spanned by the columns of  $M$ , i.e.,  $P_M = M(M^T M)^{-1} M^T$ , where the superscript  $T$  denotes matrix transposition. Let  $P_M^\perp$  denote the matrix projecting a vector onto the complementary subspace that is perpendicular to the subspace spanned by the columns of  $M$ , i.e.,  $P_M^\perp = I - P_M$ , where  $I$  denotes the  $p \times p$  identity matrix. There are several properties of a projection matrix [4]:

- 1) Idempotent:  $P^2 = P$ . The eigenvalues of a projection matrix are either 0 or 1.
- 2) Symmetry for orthogonal projection:  $P^T = P$ ;

## 5.2 Appendix II: Some Preliminary Results on $\chi^2$ and $F$ Distribution

**Def 1**[33]: Suppose  $\mathbf{x}$  is a  $N$  dimension column random vector  $\mathbf{x} \sim N(\mathbf{m}, \mathbf{I})$ , i.e., the  $x_i$ 's are independent and  $x_i \sim N(m_i, 1)$ . If  $x = \|\mathbf{x}\|^2 = \sum_{i=1}^N x_i^2$  then  $x \sim \chi_N^2(\lambda)$  where the *noncentrality parameter*  $\lambda = \|\mathbf{m}\|^2 = \sum_{i=1}^N m_i^2$ . When  $\lambda = 0$ , it is called *central  $\chi^2$  distribution*, otherwise *noncentral*.

**Lemma 1:** Suppose  $\mathbf{x}$  is a  $N$  dimension column random vector  $\mathbf{x} \sim N(\mathbf{m}, \sigma^2 \mathbf{I})$ , i.e., the  $x_i$ 's are independent and  $x_i \sim N(m_i, \sigma^2)$ , and  $P$  is a projection matrix of rank  $r$ , then

$$\frac{\|P\mathbf{x}\|^2}{\sigma^2} \sim \chi_r^2(\lambda).$$

where the noncentrality parameter  $\lambda = \frac{\|P\mathbf{m}\|^2}{\sigma^2}$ .

**Proof:** Since  $P$  is a projection matrix with rank  $r$ , it can be factorized as  $P = \mathbf{U}_{N \times r} \mathbf{U}_{N \times r}^T$  where  $\mathbf{U}$  is a full rank ( $= r$ ) orthogonal matrix, i.e.,  $\mathbf{U}^T \mathbf{U} = \mathbf{I}_{r \times r}$ . Then  $\|P\mathbf{x}\|^2 = \mathbf{x}^T P \mathbf{x} = \mathbf{x}^T \mathbf{U} \mathbf{U}^T \mathbf{x} = \|\mathbf{U}^T \mathbf{x}\|^2$ .  $\mathbf{U}^T \mathbf{x}$  is now a  $r \times 1$  Gaussian random vector due to the orthogonality of  $\mathbf{U}$ . So

$$\begin{aligned} E(\mathbf{U}^T \mathbf{x}) &= \mathbf{U}^T \mathbf{m} \\ \text{Var}(\mathbf{U}^T \mathbf{x}) &= E(\mathbf{U}^T \mathbf{x} \mathbf{x}^T \mathbf{U}) - E(\mathbf{U}^T \mathbf{x}) E(\mathbf{x}^T \mathbf{U}) \\ &= \mathbf{U}^T E(\mathbf{x} \mathbf{x}^T) \mathbf{U} - E(\mathbf{U}^T \mathbf{x}) E(\mathbf{x}^T \mathbf{U}) \\ &= \mathbf{U}^T (\mathbf{m} \mathbf{m}^T + \sigma^2 \mathbf{I}_{N \times N}) \mathbf{U} - \mathbf{U}^T \mathbf{m} \mathbf{m}^T \mathbf{U} = \sigma^2 \mathbf{I}_{r \times r}. \end{aligned}$$

In other words,  $\frac{\mathbf{U}^T \mathbf{x}}{\sigma} \sim N(\frac{\mathbf{U}^T \mathbf{m}}{\sigma}, \mathbf{I}_{r \times r})$ . Therefore the conclusion in the lemma follows from Def 1:  $\frac{\|P\mathbf{x}\|^2}{\sigma^2} = \|\frac{\mathbf{U}^T \mathbf{x}}{\sigma}\|^2$  with the noncentrality parameter  $\lambda = \frac{\|\mathbf{U}^T \mathbf{m}\|^2}{\sigma^2} = \frac{\|P\mathbf{m}\|^2}{\sigma^2}$ .

A direct consequence is that  $\lambda = 0$  when  $P\mathbf{m} = \mathbf{0}$ . In this case, since  $P\mathbf{m} = \mathbf{U}(\mathbf{U}^T \mathbf{m}) = \mathbf{0}$  and the  $r$  column vectors in  $\mathbf{U}_{N \times r}$  are independent, we must have  $\mathbf{U}^T \mathbf{m} = \mathbf{0}$ .

**Def 2 [33]:** If  $x = \frac{x_1/N_1}{x_2/N_2}$ , where  $x_1 \sim \chi_{N_1}^2(\lambda)$  and  $x_2 \sim \chi_{N_2}^2(0)$  and  $x_1, x_2$  are independent, then  $x$  has *F distribution*, denoted as  $x \sim F_{N_1, N_2}(\lambda)$ . When  $\lambda = 0$ , it is called *central F distribution*, otherwise *noncentral*.

**Lemma 2:** Suppose  $P$  and  $Q$  are two projection matrices with rank  $r_1$  and  $r_2$  respectively, and  $PQ = \mathbf{0}$ .  $\mathbf{x} \sim N(\mathbf{m}, \sigma^2 \mathbf{I})$ .  $Q\mathbf{m} = \mathbf{0}$ . Then

$$\frac{\|P\mathbf{x}\|^2}{\|Q\mathbf{x}\|^2} \sim F_{r_1, r_2}(\lambda),$$

where  $\lambda = \frac{\|P\mathbf{m}\|^2}{\sigma^2}$ .

**Proof:** Similar to the proof for Lemma 1, we can decompose  $P$  and  $Q$  as:  $P =$

$\mathbf{U}_1 \mathbf{U}_1^T$ ,  $\mathbf{Q} = \mathbf{U}_2 \mathbf{U}_2^T$ . And  $\mathbf{U}_1^T \mathbf{x}$  and  $\mathbf{U}_2^T \mathbf{x}$  are  $r_1 \times 1$  and  $r_2 \times 1$  Gaussian vectors and the latter vector has zero mean. In light of Lemma 1,  $\frac{\|\mathbf{P}\mathbf{x}\|^2}{\sigma^2} \sim \chi_r^2(\lambda)$ , and  $\frac{\|\mathbf{Q}\mathbf{x}\|^2}{\sigma^2} \sim \chi_r^2(0)$ . Further, their correlation is

$$\begin{aligned}
E[\mathbf{U}_1^T \mathbf{x} \mathbf{x}^T \mathbf{U}_2] &= \mathbf{U}_1^T E(\mathbf{x} \mathbf{x}^T) \mathbf{U}_2 \\
&= \mathbf{U}_1^T (\mathbf{m} \mathbf{m}^T + \sigma^2 \mathbf{I}) \mathbf{U}_2 \\
&= \sigma^2 \mathbf{U}_1^T \mathbf{U}_2 \\
&= \sigma^2 (\mathbf{U}_1^T \mathbf{U}_1) \mathbf{U}_1^T \mathbf{U}_2 (\mathbf{U}_2^T \mathbf{U}_2) \\
&= \sigma^2 \mathbf{U}_1^T \mathbf{P} \mathbf{Q} \mathbf{U}_2 \\
&= 0.
\end{aligned}$$

Therefore, the numerator and denominator are independent and the conclusion in Lemma ensues.

### 5.3 Appendix III: Definition of Rician, Rayleigh and $t$ Distribution

**Def 3** [33]: The pdf of  $X = \sqrt{X_1^2 + X_2^2}$ , where  $X_1 \sim N(\mu_1, \sigma^2)$  and  $X_2 \sim N(\mu_2, \sigma^2)$  are independent, is called *Rician* pdf. Its pdf is explicitly expressed as:

$$p_X(x) = \begin{cases} \frac{x}{\sigma^2} \exp[-\frac{1}{2\sigma^2}(x^2 + \alpha^2)] I_0(\frac{\alpha x}{\sigma^2}) & x > 0 \\ 0 & x < 0, \end{cases} \quad (5.1)$$

where  $\alpha^2 = \mu_1^2 + \mu_2^2$  and  $I_0(u)$  is the modified Bessel function of the first kind and order 0:

$$I_0(u) = \frac{1}{\pi} \int_0^\pi \exp(u \cos \theta) d\theta. \quad (5.2)$$

When  $\alpha^2 = 0$  it reduces to the *Rayleigh* PDF.

**Def 4** [59]: Suppose  $X \sim N(0, 1)$  and  $Y \sim \chi_n^2(0)$  are independent. Then the pdf of  $T \equiv \frac{X}{\sqrt{Y/n}}$  is called *t distribution*,  $T \sim t_n$ . Its pdf is explicitly expressed as

$$p_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad -\infty < t < \infty. \quad (5.3)$$

$\lim_{n \rightarrow \infty} p_T(t) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{t^2}{2})$ , i.e., as  $n$  is very large,  $t_n \rightarrow N(0, 1)$ .

## 5.4 Appendix IV: Analysis of *t* Test Used in fMRI Detection

Refer to Figure (5.1).  $x_i$ 's and  $y_i$ 's are all independent. The two hypotheses are:

$$H_0 : x_i \sim N(\mu_y, \sigma^2), \quad y_i \sim N(\mu_y, \sigma^2) \quad \text{versus} \quad H_1 : x_i \sim N(\mu_x, \sigma^2), \quad y_i \sim N(\mu_y, \sigma^2) \\ i = 1, 2, \dots, N.$$

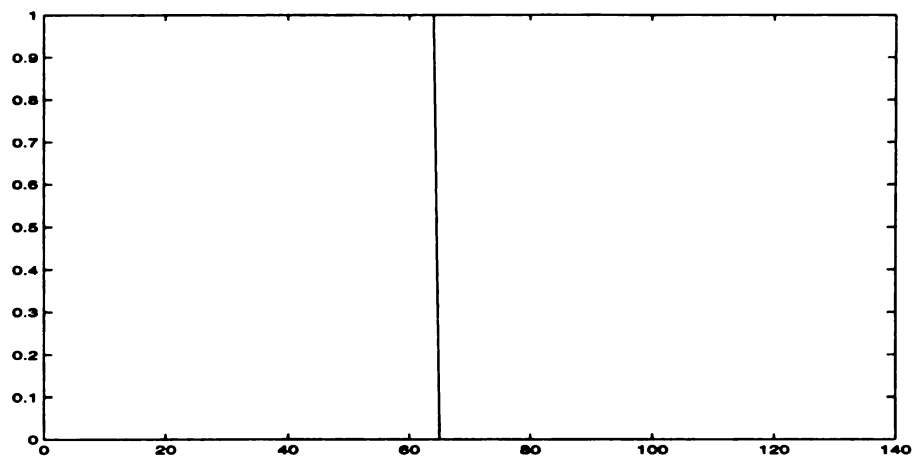
where the parameters  $\mu_x, \mu_y, \sigma^2$  are all unknown. Also by GLRT principle, the test statistics in this case turns out to be [14]:

$$t(\mathbf{x}, \mathbf{y}) = \frac{\frac{\bar{x} - \bar{y}}{\sqrt{2/N}}}{\sqrt{\frac{1}{2N-2} [\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{i=1}^N (y_i - \bar{y})^2]}} \quad (5.4)$$

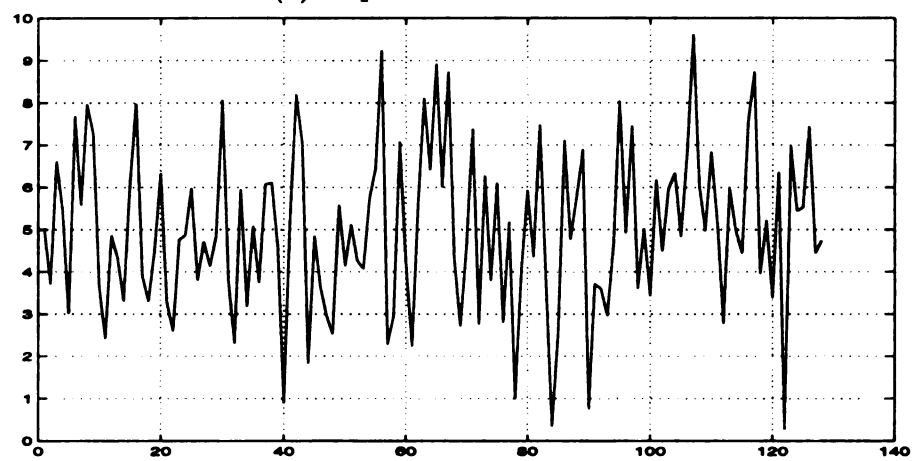
where  $\bar{x}$  and  $\bar{y}$  are the mean of  $x_i$ 's and  $y_i$ 's ( $i = 1, 2, \dots, N$ ).

Under  $H_0$ :  $\bar{x} \sim N(\mu_y, \frac{\sigma^2}{N})$  and  $\bar{y} \sim N(\mu_y, \frac{\sigma^2}{N})$ , so  $\bar{x} - \bar{y} \sim N(0, 2\frac{\sigma^2}{N})$ , i. e.,

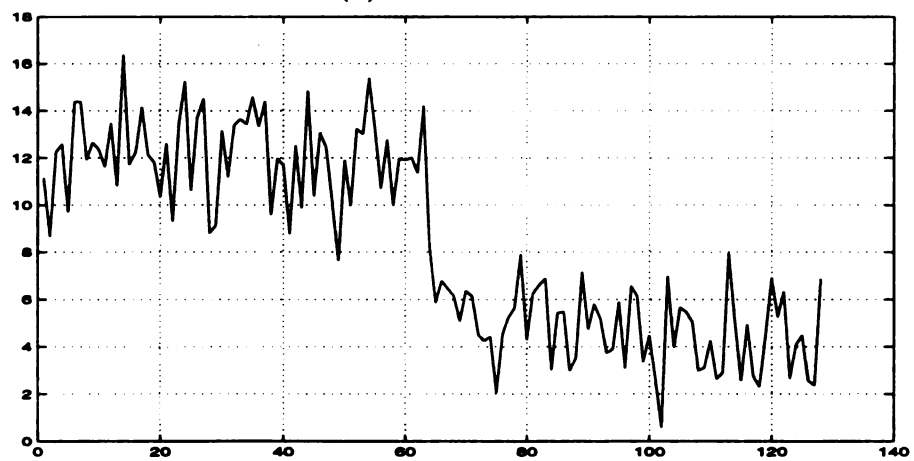
$$\frac{\bar{x} - \bar{y}}{\sqrt{\frac{2}{N}\sigma^2}} \sim N(0, 1). \quad (5.5)$$



(a) Experiment-Rest States



(b) No Activation



(c) Activation Present During Experiment Period

Figure 5.1. Experiment setup and two hypotheses for  $t$ -test used in fMRI detection.

On the other hand,

$$\sum_{i=1}^N \frac{(x_i - \bar{x})^2}{\sigma^2} = \frac{\|\mathbf{x} - \bar{x}\mathbf{1}\|^2}{\sigma^2} = \frac{\|\mathbf{x} - \frac{\mathbf{1}\mathbf{1}^T}{N}\mathbf{x}\|^2}{\sigma^2} = \frac{\|P_1^\perp \mathbf{x}\|^2}{\sigma^2} \sim \chi_{N-1}^2(0),$$

Similarly,

$$\sum_{i=1}^N \frac{(y_i - \bar{y})^2}{\sigma^2} = \frac{\|P_1^\perp \mathbf{y}\|^2}{\sigma^2} \sim \chi_{N-1}^2(0)$$

Hence,

$$\frac{\sum_{i=1}^N (x_i - \bar{x})^2 + \sum_{i=1}^N (y_i - \bar{y})^2}{\sigma^2} \sim \chi_{2N-2}^2(0). \quad (5.6)$$

Further, we observe that  $P_1^\perp \mathbf{x} \perp P_1 \mathbf{x} = \mathbf{1}\bar{x}$  and  $P_1^\perp \mathbf{y} \perp P_1 \mathbf{y} = \mathbf{1}\bar{y}$ . Therefore, the numerator and the denominator in Equation (5.4) are independent due to Gaussian assumption, and finally the  $t(\mathbf{x}, \mathbf{y})|H_0 \sim t_{2N-2}$  following Definition 4.

## 5.5 Appendix V: Principle of Invariant Test

In this Appendix I explain the idea behind invariant test [7, [56] and introduce some basic definitions and theorems [46].

In the hypothesis testing problems in Chapter 2, in addition to the *significant parameter*  $b$  or  $\mu$  on which we are testing, there are other unknown parameters called *nuisance parameters*, such as the amplitude of DC level  $a$  and the variance of noise  $\sigma^2$ . The nuisance parameters do not affect our decision, but their presence complicates the distribution of a given test statistic [7] and determination of appropriate threshold for the detector is entangled.

Therefore, a test is targeted which is unaffected by such nuisance parameters. We can do this by deliberately selecting a special class of transformations on the data so

that the distribution of the transformed data still belongs to the same family of distributions as the original, just with another set of parameters. After transformation, however, the significant parameters on which we are testing for different hypotheses still correspond to the same hypotheses as the original while the nuisance parameters are left free to change. Since the transformed data also support the original hypothesis with the significant parameters staying within the same region of the parameter space, it is a natural physical reasoning that if a test gives some decision on the original data, it has to output the same decision on the transformed data. This is what *invariant test* means. The basic idea is depicted in Figure (5.2) adapted from [7]. It turns out that these transformations possess a group structure [46].

This idea can be formulated more precisely in combination with our detection problem in Section (2.6). The observed data  $\mathbf{y}$  is regarded as a point in the *sample space*  $\Psi$  of random vectors with the same dimension as  $\mathbf{y}$ . It has the probability distribution  $P_{\Theta}$ ,  $\Theta \in \Omega$ , where  $\Theta$  is the parameter (vector-valued in our case) describing the distribution and lying in the *parameter space*  $\Omega$ . Under hypothesis  $H_1$ ,  $\Omega$  becomes  $\Omega_1$ ; under hypothesis  $H_0$ ,  $\Omega$  becomes  $\Omega_0$ . Thus if  $\Omega_0$  and  $\Omega_1$  form a partitioning of  $\Omega$ , the goal of the detection problem is then just to locate which partition  $\Theta$  lies in  $\Omega$ , *i.e.*, choose between the hypotheses as follows:

$$H_0 : \Theta \in \Omega_0 \quad \text{versus} \quad H_1 : \Theta \in \Omega_1.$$

From the above description, let  $g$  be a one to one transformation on the sample space  $\Psi$  such that  $g\mathbf{y}$  has the same distribution form as  $\mathbf{y}$  but is characterized by a different parameter  $\Theta'$ , that is, it is distributed as  $P_{\Theta'}$ ,  $\Theta' \in \Omega$ . This transformation thereby induces another transformation  $\bar{g}$  on the parameter space defined by  $\bar{g}\Theta = \Theta'$ . It is easy to see that this decision problem is invariant to the transformation  $g$  if the

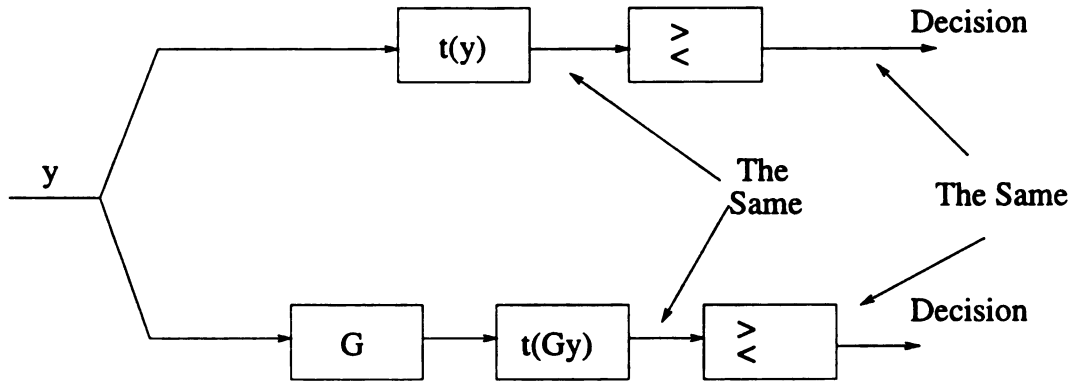


Figure 5.2. Invariant Test.

corresponding induced transformation  $\bar{g}$  maps each of the partitions of  $\Omega$  to itself:

$$\bar{g}\Omega_0 = \Omega_0 \qquad \bar{g}\Omega_1 = \Omega_1.$$

In fact, the *invariance of the test* may be achieved by requiring that the *test statistic be an invariant* to these transformation groups [7]; what this precisely means will be explained in the following definitions and is clarified by the proof of our theorem in Subsection (2.6.2).

We will know from the theory of invariance that all invariant tests can be characterized in this way, and it is possible to answer such questions as to whether they are CFAR, or if an optimum test exists among them. Further, restricting attention to such tests may bring us other advantages. For example, in many cases, the transformations  $g$  turn out to have such natural physical interpretations that, in practice, any test without the corresponding invariant property would not be acceptable, thereby substantially reducing the class of test statistics needed to be considered [57]. Also, as can be noted from our own problem in Section (2.6), there is a great reduction in the dimensionality of the parameters describing the performance of such tests.

Having understood the idea of invariant tests, we now give exact definitions for

relevant concepts and some theorems [46].

**Def 5:** A function  $t(\mathbf{y})$  is said to be invariant under a group of transformations  $G$  if  $t(g\mathbf{y}) = t(\mathbf{y})$  for all  $\mathbf{y} \in \Psi$  and all  $g \in G$ .

**Def 6:** The family of distributions  $\{P_{\Theta} : \theta \in \Omega\}$  is said to be invariant under  $G$  if every  $g \in G$ ,  $\Theta \in \Omega$  determine a unique element in  $\Omega$ , denoted by  $\bar{g}\Theta$ , such that when  $\mathbf{y}$  has distribution  $P_{\Theta}$ ,  $g\mathbf{y}$  has distribution  $P_{\bar{g}\Theta}$ . The ensemble of  $\bar{g}$  constitutes the induced group of transformations  $\bar{G}$ .

**Def 7:** A function  $\psi(\Theta)$  is said to be a *maximal invariant* under a group of transformations  $\bar{G}$  if it is invariant under  $\bar{G}$  and if  $\psi(\Theta_1) = \psi(\Theta_2)$  implies there exists some  $\bar{g} \in \bar{G}$  such that  $\Theta_2 = \bar{g}\Theta_1$ .

**Def 8:** Let the family of distributions  $\{P_{\Theta} : \Theta \in \Omega\}$  be invariant under  $G$ . The *problem of testing*  $H_0 : \Theta \in \Omega_0$  against  $H_1 : \Theta \in \Omega - \Omega_0$  is said to be invariant under  $G$  if  $\bar{g}\Omega_0 = \Omega_0$  for all  $\bar{g} \in \bar{G}$ .

**Thm 1:** A function is invariant under  $G$  if and only if it is a function of a maximal invariant under  $G$ .

**Thm 2:** If the family of distributions  $\{P_{\Theta} : \Theta \in \Omega\}$  is invariant under the group  $G$ , then  $\bar{G} = \{\bar{g} : g \in G\}$  is a group of transformations from  $\Omega$  to itself.

**Thm 3:** Suppose that the family of distributions  $\{P_{\Theta} : \Theta \in \Omega\}$  is invariant under the group  $G$ . If  $t(\mathbf{y})$  is invariant under  $G$  and  $\psi(\Theta)$  is a maximal invariant under the induced group  $\bar{G}$ , then the distribution of  $t(\mathbf{y})$  depends only on  $\psi(\Theta)$ .

## 5.6 Appendix VI: SMAP Algorithm for Multi-scale Image Segmentation

In Section (3.3), I only briefly outline the main points of Algorithm I [10]. More mathematical descriptions of this algorithm are given here. For a comprehensive

derivation, please refer to the original paper [10]. For the sake of completeness, some descriptions and equations in Section (3.3) are rewritten below.

### 5.6.1 *A Prior* Consideration

First of all, let us make clear what the field  $X^j$  physically means. Each resolution is a level in a quad-tree in the 2-D case, so a lattice point at one resolution corresponds to four points at the next finer resolution, as shown in Figure (3.5). This group of four pixels in the continuous image  $Y$  is considered as a block and  $X^j$  denotes the field containing the labeling of each of the blocks at resolution  $j$ . We may assume that the finer segmentation  $X^{j+1}$  of  $X$  is an interpolated version of  $X^j$  [9].

The fundamental assumption in Algorithm I is that the sequence of  $X^j$  forms a Markov chain [62], *i.e.*, the distribution of  $X^j$  given all coarser scale fields is only dependent on  $X^{j-1}$ :

$$P(x^j | x^l \quad l \leq j-1) = P_{x^j | x^{j-1}}(x^j | x^{j-1}). \quad (5.7)$$

This pyramid structure of the multi-scale random field (MSRF) is depicted in Figure (3.5).

Since  $Y$  (in Figure (3.5a),  $Y^J \equiv Y$ ) is exclusively dependent on  $X^J$ , where  $J$  is the finest scale index, it follows that the likelihood function is given by

$$P(y | x^j \quad n \leq J) = p(y | X^J) = p_{y|x^J}(y | x^J), \quad (5.8)$$

and then the joint distribution of  $X$  and  $Y$  may be expressed as product of recursive transition probabilities:

$$p(y, x) = p_{y|x^J}(y | x^J) \prod_{n=J}^{J_0+1} p_{x^j | x^{j-1}}(x^j | x^{j-1}) p_{x^{J_0}}(x^{J_0}), \quad (5.9)$$

where  $J_0$  is the coarsest scale of analysis.

## 5.6.2 Likelihood Function

Under the assumption that observed pixels are conditionally independent given their labels, the conditional density function for the image has the form

$$p_{y|x^J}(y|x^J) = \prod_{k \in S^J} p_{y_k|x_k^J}(y_k|x_k^J). \quad (5.10)$$

In [10], the authors confine themselves to models with two properties. First, the pixels in  $X^j$  are conditionally independent given the pixels in  $X^{j-1}$ . Second, each pixel  $X_k^j$  is only dependent on a local neighborhood of pixels at the next coarser scale. Use  $\partial k$  to denote the set of neighboring locations to  $k$ , then the transition distribution from coarse to fine scale assumes the form

$$p_{x^j|x^{j-1}}(x^j|x^{j-1}) = \prod_{k \in S^j} p_{x_k^j|x_{\partial k}^{j-1}}(x_k^j|x_{\partial k}^{j-1}), \quad (5.11)$$

where  $p_{x_k^j|x_{\partial k}^{j-1}}$  is the probability density for  $x_k^j$  given its neighbors at the coarser scale  $x_{\partial k}^{j-1}$ .

The second important assumption is that each pixel  $X_k^j$  is only dependent on a local neighborhood of pixels at the next coarser scale. The authors' first choice of the neighborhood  $\partial k$  is a quad-tree structure in 2-D case or binary-tree structure in 1-D case, as depicted in Figure (3.6). Specifically, in the quad-tree structure each point is only dependent on a single point at the coarser scale — its father  $d(k)$ . In words, if, by some means, we know that  $Y_k^j$  is active ( $X_k^j = 1$ ), then with a high probability, we may say that its four children are active as well. This probability is characterized by the transitional probability density between individual pixels (from a coarser scale to the next finer scale). The transition probability that  $X_k^j$  has state  $m$  given that its

father is in state  $m'$  is

$$p_{x_k^j|x_{\theta_k}^{j-1}}(m|m') = q^j \delta_{m,m'} + \frac{1 - q^j}{M}. \quad (5.12)$$

This equation tells us two facts: 1) the probability that the labeling will remain the same from scale  $j - 1$  to  $j$  is  $q^j + \frac{1-q^j}{M}$ , where  $q^j \in [0, 1]$ ; and 2) the probabilities that the child has any one of a number of different state from its parent's are equally likely, *i.e.*,  $\frac{1-q^j}{M}$ . The so-called “physical dependency” mentioned in Subsection (3.1.3) is embodied in the transition probability of the Markov chain.

An important property of the quad-tree structure is that the conditional distribution of  $Y$  given  $X^j$  has a product form that can be computed recursively:

$$p_{y|x^j}(y|x^j) = \prod_{k \in S^j} p_{y_k^j|x_k^j}(y_k^j|x_k^j), \quad (5.13)$$

where  $p_{y_k^j|x_k^j}$  is *defined* and computed recursively according to:

$$p_{y_k^{j-1}|x_k^{j-1}}(y_k^{j-1}|x_k^{j-1} = m) = \prod_{r \in d^{-1}(k)} \sum_{m'=1}^M p_{y_r^j|x_r^j}(y_r^j|m') p_{x_r^j|x_{\theta_r}^{j-1}}(m'|m), \quad k \in S^{j-1}, \quad (5.14)$$

where  $d^{-1}(k)$  denote the four (two) children of  $k$  in 2-D (1-D) case. Thus this kind of choice of neighborhood structure leads to simple and efficient algorithm for the image/signal segmentation [10].

### 5.6.3 Criterion and Solution

A Bayesian estimator minimizes the average cost of an erroneous segmentation:

$$\hat{x} = \arg \min_x E[C(X, x)|Y = y] \quad (5.15)$$

where  $C(X, x)$  is the cost of estimating the true segmentation,  $X$ , by the approximate segmentation  $x$ . Notice that  $X$  is random whereas  $x$  is deterministic. Expectation  $E$  is with respect to  $X$ .

The cost function used by the authors in [10] is deliberately selected to be

$$C(X, x) = 2^j,$$

where  $j$  is the unique scale such that  $X^j \neq x^j$ , but  $X^i = x^i$  for all  $i < j$ .

According to their cost function,  $\hat{x}$  turns out to be:

$$\begin{aligned}\hat{x} &= \arg \min_x \sum_{j=J_0}^J 2^j \{1 - P(X^i = x^i \mid i \leq j | Y = y)\} \\ &= \arg \max_x \sum_{j=J_0}^J 2^j P(X^i = x^i \mid i \leq j | Y = y).\end{aligned}$$

Since the random fields  $X^j$  form a Markov Chain, this estimate is computed recursively. Assuming that  $\hat{x}^i$  has been computed for  $i < j$ , and using this result to compute  $\hat{x}^j$ :

$$\hat{x}^{J_0} = \arg \max_{x^{J_0}} \log p_{x^{J_0}|y}(x^{J_0}|y), \quad (5.16)$$

$$\hat{x}^j = \arg \max_{x^j} \log p_{x^j|x^{j-1},y}(x^j|\hat{x}^{j-1}, y). \quad (5.17)$$

The recursion is started by determining the MAP estimate of the coarsest scale field given the observed data  $Y$ . The segmentation at each finer scale is computed as the MAP estimate of  $X^j$  given  $X^{j-1}$  and the image  $Y$ . It is therefore referred to as a sequential MAP (SMAP) estimator.

Assuming that  $X^{J_0}$  is uniformly distributed, they use Bayes rule and the Markov

properties of  $X$  to change the above form into another more easily computed form:

$$\hat{x}^{J_0} = \arg \max_{x^{J_0}} \log p_{y|x^{J_0}}(y|x^{J_0}), \quad (5.18)$$

$$\hat{x}^j = \arg \max_{x^j} \{\log p_{y|x^j}(y|x^j) + \log p_{x^j|x^{j-1}}(x^j|\hat{x}^{j-1})\}. \quad (5.19)$$

The first term in Equation (5.19) is the likelihood of the observed data  $y$  given the labeling at scale  $j$ . The second term carries the *a priori* information about the behavior of  $X$ .

In order to satisfy dynamic range requirement, a log likelihood function is defined at each point at each scale:

$$l_k^j(m) \equiv \log p_{y_k|x_k^j}(y_k^j|m). \quad (5.20)$$

A new recursion ensues on substituting the transition distribution of Equation (5.12) into Equation (5.14) and converting (5.14) to log likelihood functions:

$$l_k^J(m) = \log p_{y_k|x_k^J}(y_k^J|m), \quad (5.21)$$

$$l_k^{j-1}(m) = \sum_{r \in d^{-1}(k)} \log \{q^j \exp[l_r^j(m)] + \frac{1-q^j}{M} \sum_{m=1}^M \exp[l_r^j(m)]\}, \quad J_0 + 1 \leq j \leq J. \quad (5.22)$$

Finally, the SMAP segmentation may be efficiently computed by using the log likelihood functions:

$$\hat{x}^{J_0} = \arg \max_{x^{J_0}} \sum_{k \in S^{J_0}} l_k^{J_0}(x_k^{J_0}), \quad (5.23)$$

$$\hat{x}^j = \arg \max_{x^j} \sum_{k \in S^j} \{l_k^j(x_k^j) + \log p_{x_k^j|x_{\partial k}^{j-1}}(x_k^j|\hat{x}_{\partial k}^{j-1})\}, \quad J_0 + 1 \leq j \leq J. \quad (5.24)$$

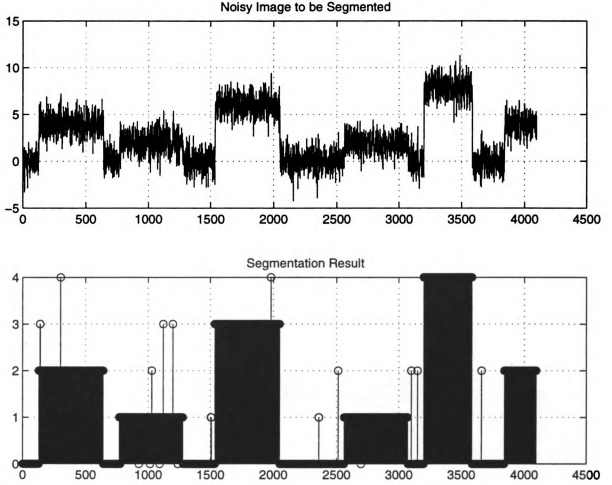


Figure 5.3. Segmentation Result Using SMAP Algorithm.

The estimate of individual pixel label is then easily evaluated by:

$$\hat{x}_k^{J_0} = \arg \max_{m \in [1, M]} l_k^{J_0}(m), \quad (5.25)$$

$$\hat{x}_k^j = \arg \max_{m \in [1, M]} \{l_k^j(m) + \log p_{x_k^j | x_{\partial k}^{j-1}}(m | \hat{x}_{\partial k}^{j-1})\}, \quad J_0 + 1 \leq j \leq J. \quad (5.26)$$

One simulation result for 2-D image is already given in Section (3.3). Here I give another example of segmenting 1-D sequence in Figure (5.3).

# **BIBLIOGRAPHY**

# BIBLIOGRAPHY

- [1] *Standard Mathematical Tables and Formulae*. 29th Edition, CRC Press, 1991.
- [2] Babak A. Ardekani, Jeff Kershaw, Kenichi Kashikura, and Iwao Kanno. Activation detection in functional MRI using subspace modeling and maximum likelihood estimation. *IEEE Trans. Medical Imaging*, 18(2):101–114, 2 1999.
- [3] P. A. Bandettini, A. Jesmanowicz, E. C. Wong, and J. S. Hyde. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.*, 30:161–173, 1993.
- [4] R.T. Behrens and L. L. Scharf. Signal rocessing applications of oblique projection operators. *IEEE Trans. Signal Processing*, 42(6):1413–1424, 1994.
- [5] J.W. Belliveau, D.N. Kennedy, B.R. Buchbinder R.C. McKinsty, R.M. Weiskoff, M. S. Cohen, J. M. Vevea, T.J. Brady, and B. R. Rosen. Functional mapping of the human visual cortex by magnetic resonance imaging. *Science*, 254(5032):716–719, 11 1991.
- [6] M. A. Bernstein, D. M. Thomasson, and W. H. Perman. Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. *Med. Phys.*, 16(5):813–817, 1989.
- [7] S. Bose and A. O. Steinhardt. A maximal invariant framework for adaptive detection with structured and unstructured covariance matrices. *IEEE Trans. Signal Processing*, 43(9):2164–2175, 1995.
- [8] A. Bruce, D. Donoho, and H-Y Gao. Wavelet analysis. *IEEE Spectrum*, (4):26–35, 10 1996.
- [9] C.Bouman and B.Liu. Multiple resolution segmentation of textured images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 13(2):99–113, 1991.
- [10] C.Bouman and M. Shapiro. A multiscale random field model for bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, 1994:

- [11] C.H.Forsgate, H. Krim, W.W. Irving, W.C.Karl, and A. S. Willsky. Multiscale segmentation and anomaly enhancement of sar imagery. *IEEE Trans. Image Processing*, 6(1):7–20, 1997.
- [12] Mark S. Cohen. *Rapid MRI and Functional Applications, In Brain Mapping: The Methods*, A. W. Toga and J. C. Mazziota (Editors). Academic Press, 1996.
- [13] R. T. Constable and R. M. Henkelman. Why MEM does not work in MR image reconstruction. *Magn. Reson. Med.*, 14:12–25, 1990.
- [14] R. Todd Constable, P. Skudlarski, and John C. Gore. An ROC approach for evaluating functional brain MR imaging and postprocessing protocols. *Magn. Reson. Med.*, 34:57–64, 1995.
- [15] M.S. Crouse, R.D.Nowak, and R.G.Baraniuk. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing*, 46(4):886–902, April 1998.
- [16] J. Deller, J. Proakis, and J. Hanson. *Discrete-Time Processing of Speech Signals*. Prentice Hall, Englewood Cliffs, NJ, 1993.
- [17] Xavier Descombes, Frithjof Kruggel, and D. Yves von Cramon. Spatio-temporal fmri analysis using markov random fields. *IEEE Trans. Medical Imaging*, 17(6):1028–1039, 1998.
- [18] W. A. Edelstein, P. A. Bottomley, and L. M. Pfeifer. A signal-to-noise calibration procedure for NMR imaging systems. *Med. Phys.*, 11:180–185, 1984.
- [19] W. A. Edelstein, G. Glover, C. Hardy, and R. Redington. The intrinsic signal-to-noise ratio in NMR imaging. *Magn. Reson. Med.*, 3:604–618, 1986.
- [20] S. Forman, J. C. Cohen, M. Fitzgerald, W. F. Eddy, M. A. Mintun, and D. C. Noll. Improved assessment of significant change in functional magnetic resonance fMRI: Use of cluster size threshold. *Magn. Reson. Med.*, 33:636–647, 1995.
- [21] L. R. Frank, R. B. Buxton, and E. C. Wong. Probabilistic analysis and functional magnetic resonance imaging data. *Magn. Reson. Med.*, 39:132–148, 1998.
- [22] K. J. Friston, P. Jezzard, and R. Turner. Analysis of functional mri time-series. *Human Brain Mapping*, 1:153–171, 1994.

- [23] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intell.*, 6(11):721–741, 1984.
- [24] C. R. Genovese. A time-course model for fMRI data. In *Proc. Intl. Symp. Magn. Reson. Med. Meeting*, page 1669, 1997.
- [25] C. R. Genovese, D. C. Noll, and W. F. Eddy. Estimating test-retest reliability of functional MR imaging i:statistical methodology. *Magn. Reson. Med.*, 38:497–507, 1997.
- [26] Xavier Golay, Spyros Kollias, Gautier Stoll, Dieter Meier, Anton Valavanis, and Peter Boesiger. A new correlation - based fuzzy logic clustering algorithm for fMRI. *Magn. Reson. Med.*, 40:249–260, 1998.
- [27] Cyril Goutte, Peter Toft, Egil Rostrup, Finn A. Nielsen, and Lars Kai Hansen. On clustering fMRI time series. *NeuroImage*, 9:298–310, 1999.
- [28] H.Derin and H.Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Patt. Anal. Mach. Intell.*, 9(1):39–55, 1987.
- [29] Xiaoping Hu, Tuong Huu Le, Todd Parrish, and Peter Erhard. Retrospective estimation and correlation of physiological fluctuation in functional MRI. *Magn. Reson. Med.*, 34:201–212, 1995.
- [30] W. W. Irving and J. N. Tsitsiklis. Some properties of optimal thresholds in decentralized detection. *IEEE Trans. Automatic Control*, 39(4):835–838, 1994.
- [31] Jian-Ming Jin. Electromagnetics in magnetic resonance imaging. *IEEE Antennas and Propagation Magazine*, 40(6):7–22, 12 1998.
- [32] Gerald Kaiser. *A Friendly Guide to Wavelets*. Birkhauser, Boston, Basel, Berlin, 1994.
- [33] S. M. Kay. *Fundamentals of Statistical Signal Processing. Detection Theory*. Prentice-Hall, New Jersey, 1998.
- [34] S. Kim, W. Richter, and K. Ugurbil. Limitations of temporal resolution in functional MRI. *Magn. Reson. Med.*, 37:631–636, 1997.
- [35] T. Kim, L. Al-Dayeh, P. Patel, and M. Singh. Bayesian processing for fMRI. In *Proc. Intl. Soc. Magn. Reson. Med.*, Sidney, Australia, 1998.

- [36] S. Lai and G. H. Glover. Detection of BOLD fMRI signals using complex data. In *Proc. Intl. Soc. Magn. Reson. Med. Meeting*, page 1671, 1997.
- [37] Nicholas Lange. Tutorial in biostatistics: Statistical approaches to human mapping by fMRI. *Statistics in Medicine*, 15:389–428, 1996.
- [38] Tuong Huu Le and Xiaoping Hu. Retrospective estimation and correlation of physiological artifacts in fMRI by direct extraction of physiological activity from MR data. *Magn. Reson. Med.*, 35:290–298, 1996.
- [39] R A Lerski. *Physical Principles and Clinical Applications of Nuclear Magnetic Resonance*. Paradigm Print, Gateshead, London, Great Britain, 1985.
- [40] M.R. Luettggen, W.C.Karl, A. S. Willsky, and R.R.Tenney. Multiscale representations of Markov random fields. *IEEE Trans. Signal Processing*, 41(12):3377–3396, 1993.
- [41] A. Macovski. Noise in MRI. *Magn. Reson. Med.*, 36(3):494–497, 1996.
- [42] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. Patt. Anal. Mach. Intell.*, 14(7):710–732, 1992.
- [43] M. J. McKeown, S. Makeig, G. G. Brown, T-P Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI data by blind separation into independent spatial components. *Human Brain Mapping*, 6:160–188, 1998.
- [44] E. R. McVeigh, R. M. Henkelman, and M. J. Bronskill. Noise and filtration in magnetic resonance imaging. *Med. Phys.*, 12:586–591, 1985.
- [45] M.Malfait and Dirk Roose. Wavelet-based image denoising using a Markov random field *a priori* model. *IEEE Trans. Image Processing*, 6(4):549–565, 1997.
- [46] J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.
- [47] Fangyuan Nan. SVD reconstruction algorithm and determination of signal number by frequency domain information. *Submitted to IEEE Trans. on Antenna and Propagation*, 9 1999.
- [48] Fangyuan Nan and Robert D. Nowak. Generalized likelihood ratio detection of functional MRI signal using complex data. *IEEE Trans. Medical Imaging*, 18(4):320–329, 4 1999.

- [49] R. Nowak. *Multiscale Hidden Markov Models for Bayesian Image Analysis, In Bayesian Inference in Wavelet Based Models, B. Vidakovic and P.Muller(Editors)*. Springer Verlag, New York, 1999.
- [50] Patrick Perez and Fabrice Heitz. Restriction of a markov random field on a graph and multiresolution statistical image modeling. *IEEE Trans. Inform. Theory*, 42(1):180–190, 1996.
- [51] G. Polychronopoulos and J. N. Tsitsiklis. Explicit solutions for some simple decentralized problems. *IEEE Trans. Aerospace and Electronic Systems*., 26(2):282–292, 1990.
- [52] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, 77:257–285, 1989.
- [53] Raghuveer M. Rao and Ajit S. Bopardikar. *Wavelet Transforms—Introduction to Theory and Applications*. Addison-Wesley Longman, Inc., Reading, Massachusetts, 1998.
- [54] O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Mag.*, pages 14–38, 10 1991.
- [55] U. E. Ruttimann, N. F. Ramsey, D. W. Hommer, P. Thevanaz, Ch. Lee, and M. Unser. Analysis of functional magnetic resonance images by wavelet decomposition. In *Proc. Intl. Conf. Image Proc.*, volume I, pages 633–636, 1995.
- [56] L. L. Scharf. *Statistical Signal Processing. Detection, Estimation, an Time Series Analysis*. Addison-Wesley, Reading, MA, 1991.
- [57] L. L. Scharf and B. Friedlander. Matched subspace detectors. *IEEE Trans. Signal Processing*, 42(8):2146–2157, 1994.
- [58] K. Sekihara and H. Koizumi. Detecting cortical activities from fMRI time-course data using MUSIC algorithm with forward and backward covariance averaging. *Magn. Reson. Med.*, 35:907–813, 1996.
- [59] K. S. Shanmugan and A. M. Breipohl. *Random Signals: Detection, Estimation and Data Analysis*. John Wiley and Sons, New York, 1988.
- [60] B. Siewert, B. M. Bly, G. Schlaug, D. G. Darby, V. Thangaraj, S. Warach, and R. Edelman. Comparison of the bold- and epistar-technique for functional brain imaging using signal detection theory. *Magn. Reson. Med.*, 36:249–255, 1996.

- [61] M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan. *Introduction to Statistical Signal Processing with Applications*. Prentice Hall, Englewood Cliffs, New Jersey, 1996.
- [62] H. Stark and J. W. Woods. *Probability, Random Processes and Estimation Theory for Engineers (2nd Edition)*. Prentice Hall, Englewood Cliffs, New Jersey, 1994.
- [63] Gilbert Strang. Wavelets. *American Scientist*, 82:250–255, May-June 1994.
- [64] S. C. Strother, I. Kanno, and D. A. Rottenberg. Principle component analysis, variance partitioning, and functional connectivity. *Journal of Cerebral Blood Flow and Metabolism*, 15:353–360, 1995.
- [65] M. Vetterli and Jelena Kovacevic. *Wavelets and Subband Coding*. Prentice Hall PTR, Englewood Cliffs, New Jersey, 07632, 1995.
- [66] A. Villringer and U. Dirnagl. Coupling of brain activity and cerebral blood flow: Basis of functional neuroimaging. *Cerebrovascular and Brain Metabolism Reviews*, 7:240–276, 1995.
- [67] Marinus T. Vlaardingerbroek and Jacques A. Den Boer. *Magnetic Resonance Imaging*. Springer, Verlag Berlin Heidelberg, 1996.
- [68] G. A. Wright. Magnetic resonance imaging. *IEEE Signal Processing Mag.*, 40(1):56–66, 1 1997.
- [69] Jinhua Xiong, Jia-Hong Gao, Jack L. Lancaster, and Peter T. Fox. Assessment and optimization of functional MRI analyses. *Human Brain Mapping*, 4:153–167, 1996.
- [70] Ilan Ziskind and Mati Wax. Maximum likelihood localization of multiple sources by alternating projection. *IEEE Trans. Acoust., Speech, Signal Processing*, 36(10):1553–1560, 10 1988.



MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02092 7160