

# fMRI Detection with Spatial Regularization

by

Wanmei Ou

Submitted to the Department of Electrical Engineering and Computer  
Science

in partial fulfillment of the requirements for the degree of

Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2005

© Massachusetts Institute of Technology 2005. All rights reserved.

Author .....  
Department of Electrical Engineering and Computer Science  
May 6, 2005

Certified by.....  
Polina Golland  
Assistant Professor  
Thesis Supervisor

Accepted by.....  
Arthur C. Smith  
Chairman, Department Committee on Graduate Students



# fMRI Detection with Spatial Regularization

by

Wanmei Ou

Submitted to the Department of Electrical Engineering and Computer Science  
on May 6, 2005, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Electrical Engineering and Computer Science

## Abstract

Functional Magnetic Resonant Imaging (fMRI) is a non-invasive imaging technique used to study the brain. Neuroscientists have developed various algorithms to determine which voxels of the images are active. Most of these algorithms, operating on the signal of each voxel separately, are referred to as the voxel-by-voxel detectors. Among those voxel-by-voxel detectors, paired T-test and General Linear Model (GLM) are the most popular. The Mutual Information (MI) based detector has recently been introduced. It is interesting to compare these three detectors' modelling assumptions, as well as their performance, in order to understand their advantages and shortcomings.

Due to the low signal-to-noise ratio (SNR), the voxel-by-voxel detectors usually result in fragmented activation pattern, which is not in agreement with our understanding of brain activation. The biological models of brain activation suggest that adjacent locations of the brain tend to be in the same activation state. We take advantage of these models and apply a Markov Random Field (MRF) spatial prior to the statistics provided by the voxel-by-voxel algorithms. MRF has been shown to be able to overcome the effect of over-smoothing, which is the major drawback of the conventional spatial regularization models such as the Gaussian smoothing model. We adopt Mean Field, a variational algorithm, to estimate the MRF solution. We show that Mean Field can provide reasonable approximation compared with the exact solver in the case of binary MRFs, while reducing the computations by one to two orders of magnitude in our simulated and real data sets. In addition, unlike the exact solver, it can handle multiple-state MRFs.

Inspired by atlas-based segmentation, we further refine the spatial regularization process by incorporating anatomical information, such as segmentation results from Magnetic Resonance Imaging (MRI), into the MRF prior. The extended MRF model encodes both tissue type and activation state. To our knowledge, our approach is the first spatial smoothing method that utilizes anatomical information without cortical surface extraction. To evaluate the smoothing models, we performed ROC and confusion matrix analysis on synthetic data. We also evaluate them by studying their ability to recover activation from significantly shorter time course using real data. In-

cluding anatomical information improves detection accuracy for both the Gaussian-smoothing-based detector and the MRF-based detector. The Gaussian-smoothing model provides poor results if we are interested in both positive and negative activation regions in the brain. Furthermore, the anatomically guided MRF-based detector improves the detection quality compared with the anatomically guided Gaussian-smoothing-based detector for standard fMRI in standard SNR quality.

Thesis Supervisor: Polina Golland

Title: Assistant Professor

## Acknowledgments

Nothing is created in a vacuum, and this thesis could not have been completed without the input and support of many people. First and foremost I would like to thank my research advisor Professor Polina Golland. She showed me different ways to approach a research problem as well as the need to be persistent to accomplish my goal. She is also the most responsible for helping me complete the writing of this thesis and the challenging research that lies behind it. Without her encouragement and constant guidance, I could not have finished this thesis. I thank her for being flexible, so attuned to her student, and so respectfully open to differences.

I would like to thank Dr. Lawrence Panych for providing fMRI data, Professor Alan Oppenheim for making my transition to graduate school smoother, and Dr. William Wells and Dr. Bruce Fischl for helping me during the course of my thesis. Eric, Marshall, and Thomas have been always available for intelligent and insightful discussion. Without Kilian's help of accurate segmentations of MR images, I would not have proved the idea in my thesis. Furthermore, I would like to thank the MIT Writing center for making sure my thesis is clear for general audiences.

This thesis was partially supported by the NIH National Center for Biomedical Computing Program, National Alliance for Medical Imaging Computing (NAMIC), Fund No. 1U54 EB005149, the NSF IIS 9610249 grant, the NIH R01 NS37922 grant, and the Shillman fellowship.

I would also like to thank my boyfriend, Jeffrey Hughes, for being a 24-hour oncall technician for fixing my computer and LaTeX problems. Last, but not least, I thank my parents, Qinghua Ou and Gen Y. Won, for giving me life in the first place, and for unconditional support and encouragement to pursue my interests, even when the interests went beyond boundaries of language, field, and geography.



# Contents

<b>1</b>	<b>Introduction</b>	<b>15</b>
<b>2</b>	<b>Background on fMRI</b>	<b>19</b>
2.1	Basics of MRI . . . . .	19
2.2	Basics of fMRI . . . . .	22
2.3	fMRI Applications . . . . .	24
2.3.1	Applications in Studies of Neuroscience . . . . .	24
2.3.2	Other fMRI Applications . . . . .	26
2.4	Future Development . . . . .	28
2.5	Summary . . . . .	29
<b>3</b>	<b>Voxel-by-voxel Detectors</b>	<b>31</b>
3.1	Background . . . . .	32
3.1.1	Paired T-test (T-test) . . . . .	32
3.1.2	Mutual Information (MI) . . . . .	36
3.1.3	General Linear Model (GLM) . . . . .	39
3.2	Experiments and Results . . . . .	43
3.2.1	T-test <i>vs.</i> MI . . . . .	43
3.2.2	T-Test <i>vs.</i> GLM . . . . .	45
3.3	Summary . . . . .	47
<b>4</b>	<b>Spatial Regularization for fMRI Detection</b>	<b>49</b>
4.1	Background Statistics . . . . .	51

4.1.1	Sufficient Statistic for the General Linear Model . . . . .	51
4.1.2	Sufficient Statistic for the Mutual Information Detector . . . . .	52
4.2	Markov Priors . . . . .	52
4.3	A Variational Algorithm: Mean Field . . . . .	54
4.4	Estimating Model Parameters . . . . .	58
4.5	Empirical Evaluation . . . . .	59
4.5.1	Synthetic Data Sets with Binary Activation States . . . . .	59
4.5.2	Synthetic Data Sets with Trinary Activation States . . . . .	62
4.6	Summary . . . . .	65
<b>5</b>	<b>Spatial Regularization with Anatomical Structure</b>	<b>67</b>
5.1	Anatomical Priors for Spatial Regularization . . . . .	68
5.2	Empirical Evaluation . . . . .	70
5.2.1	Synthetic Data Sets . . . . .	71
5.2.2	Real fMRI Experiments . . . . .	77
5.3	Summary . . . . .	79
<b>6</b>	<b>Discussion and Conclusions</b>	<b>85</b>
<b>A</b>	<b>Detailed Derivations</b>	<b>89</b>
A.1	The Independence of the Sample Mean and Sample Variance . . . . .	89
A.2	T-Distribution . . . . .	91
A.3	Confidence Intervals . . . . .	93
A.4	Paired T-test is a special case of GLM . . . . .	93



# List of Figures

1-1	<i>ROC curves for different smoothing techniques augmented with the anatomical information. False positive rate is shown on the log scale.</i>	17
3-1	<i>Graphical explanation of converting T-statistic to P-value . . . . .</i>	36
3-2	<i>Left: maximum intensity projections (axial, coronal and sagittal) of the activation maps detected by T-test (top) and MI (bottom). The colormap is used to show the negative logarithm of P-values. Right: joint scatter of P-values (log-scale) produced by T-test and MI. . . . .</i>	44
3-3	<i>Left: intensity histograms during task (<math>p_{Y U=1}(y)</math>, blue), during rest (<math>p_{Y U=0}(y)</math>, red), and over the entire run (<math>p_Y(y)</math>, black) for two voxels from the right plot in Fig. 3-2. Voxel A: <math>P_{T-test} = 10^{-15}</math>, <math>P_{MI} = 10^{-7}</math>; Voxel B: <math>P_{T-test} = 10^{-2}</math>, <math>P_{MI} = 10^{-5}</math>. Right: time course (black) and scaled protocol timeline (red) for the same voxels. . . . .</i>	45
3-4	<i>Left: maximum intensity projection (axial, coronal and sagittal) of the activation maps detected by T-test (top) and GLM (bottom). The colormap is used to show the negative logarithm of P-values. Right: joint scatter of P-values (log-scale) produced by T-test and GLM. . . . .</i>	46
3-5	<i>Time course (black), T-test protocol (red), and the GLM regressor (blue) for four voxels from Fig. 3-4. Voxel A<sub>1</sub>: <math>P_{T-test} = 10^{-2}</math>, <math>P_{GLM} = 10^{-13}</math>; Voxel A<sub>2</sub>: <math>P_{T-test} = 10^{-14}</math>, <math>P_{GLM} = 10^{-13}</math>; Voxel B<sub>1</sub>: <math>P_{T-test} = 10^{-13}</math>, <math>P_{GLM} = 10^{-16}</math>; Voxel B<sub>2</sub>: <math>P_{T-test} = 10^{-13}</math>, <math>P_{GLM} = 10^{-6}</math>. . . . .</i>	47
4-1	<i>Graphical model for <math>P_{\bar{X}, \bar{Z}}</math>. . . . .</i>	54

4-2	<i>ROC curves for different smoothing techniques, at two noise levels. False positive rate is shown on the log scale. . . . .</i>	60
4-3	<i>ROC curves for different smoothing techniques, at two noise levels. False positive rate is shown on the log scale. It is impossible to achieve 100% true detection rate, since false detections between positively and negatively active states are ignored. . . . .</i>	64
5-1	<i>Graphical model for <math>P_{\vec{U}, \vec{Z}, \vec{W}}</math>. . . . .</i>	68
5-2	<i>ROC curves for different smoothing techniques augmented with the anatomical information, at two noise levels. False positive rate is shown on the log scale. . . . .</i>	72
5-3	<i>ROC curves for different smoothing techniques augmented with the anatomical information, at two noise levels. False positive rate is shown on the log scale. It is impossible to achieve 100% true detection rate, since false detections between positively and negatively active states are ignored. . . . .</i>	73
5-4	<i>One slice from estimated activation maps for the same ground truth at 0.05% false positive rate. True detections are shown in yellow; false positive voxels are shown in red. . . . .</i>	80
5-5	<i>One slice from estimated activation maps for the same ground truth at 0.05% false positive rate. True positive and negative detections are shown in yellow and blue; false detections are shown in red. . . . .</i>	81
5-6	<i>Real fMRI study. One slice in the estimated activation map. (a) No spatial smoothing, using the entire time course. (b)-(f) Estimation based on the first five epochs of the time course using different spatial smoothing methods. . . . .</i>	82

5-7 *Real fMRI study. One slice in the estimated activation map. (a) No spatial smoothing, using the entire time course. (b)-(f) Estimation based on the first five epochs of the time course using different spatial smoothing methods. Positive and negative detections are shown in yellow and blue, respectively. . . . . 83*



# List of Tables

2.1	<i>Compare T1 and T2 Weighted Image . . . . .</i>	22
3.1	<i>Modelling assumptions made by different methods. . . . .</i>	32
4.1	<i>Detection performance of three selected detectors at two SNR levels. .</i>	64
5.1	<i>Detection performance of three selected detectors and their correspond- ing anatomically guided extensions at <math>-9</math> dB SNR. . . . .</i>	74
5.2	<i>Detection performance of three selected detectors and their correspond- ing anatomically guided extensions at <math>-6</math> dB SNR. . . . .</i>	75



# Chapter 1

## Introduction

Modelling brain activation based on fMRI has been an active research area since the genesis of its development. fMRI signal, obtained based on the different paramagnetic properties of oxygenated and deoxygenated blood, reflects the underlying information processing function mediated by neuronal activity. According to a pre-designed experimental protocol, a subject undergoes fMRI scans while alternating between performing tasks and resting. With fast encoding techniques, images are generated as fast as once per second.

Neuroscientists have developed various detectors to determine which voxels of the image are active. Most of these detectors, operating on the signal of each voxel separately, measure the difference in the signal obtained between the task and rest periods or the correlation between the signal and the experimental protocol. This kind of detector is usually referred to as a voxel-by-voxel detector.

Two most popular voxel-by-voxel detectors for activation detection in fMRI images are the paired T-test and the General Linear Model (GLM). They are parametric approaches with simple implementations (reliable software is widely available). Recently, several alternatives, such as Mutual Information (MI), have been proposed to relax some of the assumptions on the temporal structure of the signal and the statistics of noise. Combining the statistic of each voxel, we obtain a statistical parametric map (SPM). Researchers usually convert the statistic into a P-value if the distribution of the statistic is known; otherwise, it is converted into a Z-score. A voxel with

a P-value or Z-score beyond a certain threshold is considered an active voxel. As a result, an activation map is created for mapping brain function onto brain anatomy.

Due to a low signal-to-noise ratio (SNR) in fMRI signal, thresholded SPMs produced by various voxel-by-voxel detectors usually have a highly fragmented activation pattern, which is not in agreement with our understanding of brain activation. The biological models of brain activation suggest that adjacent locations of the brain tend to be in the same activation state. Many researchers take advantage of the general understanding of brain activation and smooth the signal spatially using a Gaussian filter prior to detection, but the resulting SPMs are usually overly-smoothed. An alternative spatial smoothing prior is a Markov Random Field (MRF), which has been shown to be able to overcome the effect of over-smoothing in other applications. We adopt Mean Field, a variational algorithms, to estimate the MRF solution. We show that Mean Field can provide reasonable approximation compared with the exact solver in the case of binary MRFs, while reducing the computations by one to two orders of magnitude in our simulated and real data sets. Compared with the Gaussian-smoothing-based detector, the MRF-based detector achieves a higher detection rate in our experiment using synthetic and real data sets and in binary and trinary activation configurations.

Atlas-based segmentation inspired our idea of further refinement of the spatial regularization process by incorporating anatomical information, such as segmentation results from Magnetic Resonance Imaging (MRI), into the MRF prior. Anatomical information biases the detections towards a correct location. The extended MRF model encodes both tissue type and activation state. Our approach is one of the first few spatial smoothing methods that utilize anatomical information in fMRI detection. On the other hand, it is straightforward to inject anatomical structure into the GLM detector and the Gaussian-smoothing-based detector. We also show that all three anatomically guided detectors can achieve high detection accuracy in experiments using synthetic (Fig. 1-1) or real fMRI data compared to their corresponding detectors without anatomical information. Furthermore, Fig. 1-1 also illustrates that the anatomically guided MRF-based detector improves the detection quality compared



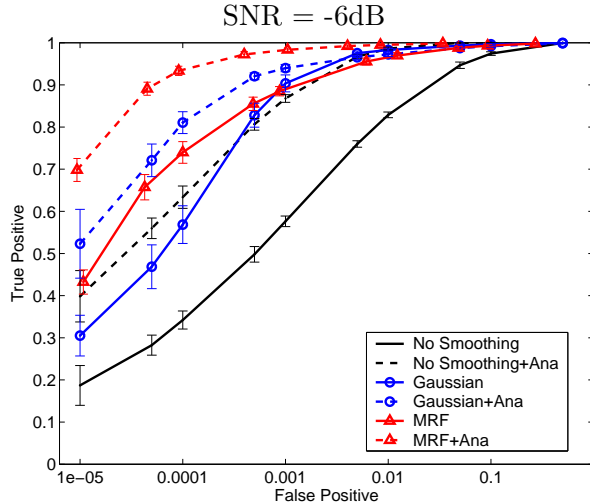


Figure 1-1: *ROC curves for different smoothing techniques augmented with the anatomical information. False positive rate is shown on the log scale.*

with the anatomically guided Gaussian-smoothing-based detector and the anatomically guided GLM detector for fMRI in standard SNR quality.

Contributions of this thesis include:

- In the context of a block-design fMRI study, we empirically compare selected voxel-by-voxel detectors and discuss their theoretical similarities and differences.
- We compare detection accuracy of an MRF-based detector with a Gaussian-smoothing-based detector in both simulated and real fMRI data sets.
- A fast variational approximation algorithm, Mean Field, is adopted for solving MRFs, and we compare its result with the exact solution in binary MRFs.
- We refine selected detectors by injecting anatomical information and compare their performance by performing ROC analysis and by presenting confusion matrices. The resulting activation maps generated by these detectors are also presented in order to give readers insight into the advantages of incorporating anatomical information into the detectors.

This thesis is organized as follows. Chapter 2 presents basic physical mechanisms and biological assumptions for generating fMRI images. Chapter 3 begins with the

modelling assumption and a mathematical derivation of various voxel-by-voxel algorithms for fMRI detection, including paired T-test, GLM, and MI. Then, it presents the empirical comparisons of selected pairs of algorithms using real fMRI data. Chapter 4 begins with the background of MRF and the derivation of the Mean Field algorithm in variational approximation point of view. Then, we present the results of performing spatial smoothing using MRF priors. Chapter 5 introduces the novel spatial smoothing method that incorporates anatomical structure. We discuss possible ways to incorporate anatomical structure into different conventional detectors and compare anatomically guided MRF-based detector's experimental results with other anatomically guided detectors. Discussion and conclusions follow this section.

# Chapter 2

## Background on fMRI

Functional Magnetic Resonance Imaging (fMRI) is a relatively new non-invasive technique used to study the brain. fMRI signals reflect differences in the paramagnetic properties of oxygenated and deoxygenated blood as a consequence of activation [20]. The development of fMRI is based on Magnetic Resonance Imaging (MRI), invented by Felix Bloch and Edward Purcell in the 1940s. MRI utilizes combinations of magnetic fields and radio waves to create three-dimensional internal images of soft tissue, including the brain, the spinal cord, and the muscle. The development of Echo Planar Imaging (EPI) enables fast MRI scans and has facilitated fMRI's development. This chapter begins with the background of MRI and fMRI. It then explains how psychological research and clinical practise benefit from fMRI, followed by a brief discussion of fMRI's future development.

### 2.1 Basics of MRI

MRI is used to generate static images of soft tissue, such as the brain, for anatomy identification. It is based on the principles of nuclear magnetic resonance: the vibration of molecules in a strong magnetic field. A nucleus of an atom has spin states of different energy levels which can be excited to make a transition from a low energy state to a higher state if the atom absorbs a photon whose energy matches the difference between the states. The nucleus can also make a transition from a high

energy state to a lower one by emitting a photon whose energy matches the difference between the states. MRI only focuses on the hydrogen atoms, among all different types of atoms in the body, because of their high concentration and large magnetic moment.

When placed in a magnetic field of strength  $B$ , a nucleus with a net spin can absorb a photon of frequency  $\nu$ .  $\nu$  depends on the gyromagnetic ratio,  $\gamma$ , of the particle. Eq. (2.1) illustrates this relationship.

$$\nu = \gamma B \tag{2.1}$$

Hydrogen,  $^1\text{H}$ , has a single unpaired proton and  $1/2$  net spin. The proton spins and creates magnetic field. Spin is used to characterize this fundamental property of a charged particle, such as the proton in this case. We can consider the spin of this proton as a magnetic moment vector, causing the proton to behave like a tiny magnet with a north and a south pole. The gyromagnetic ratio for hydrogen is  $\gamma = 42.58$  MHz/T. When the energy of the photon,  $E = h\gamma B$ , matches the difference between two spin states, an absorption occurs.

MRI is a non-invasive imaging technique. A subject lies inside an MRI scanner, with his/her head surrounded by a strong magnet and a radio wave transducer. The induced magnetic field causes the spin of a small amount of hydrogen protons (around  $10^{-4}\%$ ) within the subject's head to align with the magnetic field. In fact, each of these proton spins is slightly tilted away from the external magnetic field. The protons precess, or wobble, about the axis of the external field. The frequency of the precession is directly proportional to the strength of the magnetic field, and it is defined by the Larmor Equation (Eq. (2.1)), which is the underlining principle of magnetic resonance imaging. The rest of the proton spines are at random orientations, whose induced magnetic fields cancel out with each other. Therefore, the average spin of all protons is in the same direction as the external field.

A radio frequency pulse is then directed at the subject's head, exciting some of the hydrogen protons, which causes their spins' mis-alignment with the induced

magnetic field. If the radio frequency pulse is at the resonance frequency,  $\nu$ , then the protons can absorb the energy and jump to a higher energy state. Additionally, the protons start spinning in-phase. Once the pulse is removed, the mis-aligned protons emit radio waves as they realign themselves to the external magnetic field, and this process is referred to as the longitudinal relaxation. Additionally, the spinning starts to de-phase, and this process is referred to as the transverse relaxation. Various tissues have different relaxation time constants in the two relaxation processes. The following is the description of the time constants of the longitudinal and transverse relaxation.

T1, also called spin lattice<sup>1</sup> relaxation time, is the time constant that describes the return to equilibrium of the longitudinal (Z component) magnetization. The longitudinal component is in the same direction as the induced magnetic field. T1 relaxation is the process in which the energy absorbed by the excited protons is released back into the surrounding lattice, reestablishing thermal equilibrium.

T2, also called spin-spin relaxation time, is the time constant which describes the return to equilibrium of the transverse (X and Y components) magnetization. The transverse component is in the plane perpendicular to the longitudinal component. We can also consider T2 relaxation as the random interaction among spins that causes a cumulative loss in phase, resulting in an overall loss of signal.

T1 is always longer than T2. In general, T2 values are not related to the field strength, but T1 values are. T1 and T2 are unique for different tissue types, so it is sufficient to use one of them to represent the signal. The emitted radio waves can be captured by a receiver and used to construct an image of the soft tissue. TE is the time period between the initial perturbing radio frequency pulse and the center of the acquisition interval. TR stands for repetition time, or the elapsed time between successive RF excitation pulses. These facts permit the introduction of contrast in the images by appropriate manipulation of the TR and TE. Two types of contrasts are possible based on weighting with T1 and T2, as described below.

T1-weighted imaging emphasizes the differences on longitudinal relaxation rates by

---

<sup>1</sup>In this context, lattice means the surrounding tissue of a proton.

Table 2.1: *Compare T1 and T2 Weighted Image*

Image Type	TR	TE
T1 weighted	short (450-850 ms)	short (10-30 ms)
T2 weighted	long (2000+ ms)	long (60+ ms)

utilizing a short TR and short TE. The short TR allows a full recovery of tissues with a short T1 (i.e., fat) while allowing only partial recovery of tissues with a long T1 (i.e., CSF). The short TE used to obtain a T1-weighted image will allow minimal loss of transverse magnetization due to T2 relaxation. Basically, short TR increases T1 effects with the short TE minimizes T2 effects.

T2-weighted imaging emphasizes the differences on transverse relaxation rates by utilizing a long TR and long TE. The long TR allows tissues to reach complete longitudinal magnetization which will reduce T1 effects. The long TE will allow for the loss of transverse signal enhancing T2 effects. Fluid has a very long T2 and is frequently associated with pathology, so it is important to take advantage of the imaging parameters to accentuate signal differences.

When a certain type of tissue experiences a uniform magnetic field, energy emitted from different parts of the tissue are of identical frequency. Utilizing this fact, if a unique magnetic field is applied to each of the spin regions, an image of their positions can be constructed. Superimposing a spatially variant magnetic field, referred to as the gradient magnetic field, allows us to accomplish this.

## 2.2 Basics of fMRI

fMRI, a variation of MRI, is used to study dynamical changes in the brain. Besides the physical principles of MRI, fMRI takes advantage of two additional phenomena. First, it is known that neural activation causes an increase in blood flow. The blood is to provide extra oxygen and glucose for the active brain cells<sup>2</sup>. Second, iron

---

<sup>2</sup>However, for the first three to six seconds after activation occurs, the activated areas experience a relative decrease in oxygenated blood as oxygen is extracted by the active regional neurons after activation occurs, according to a recent neural study.

atoms by themselves cause distortions in the magnetic field, but iron atoms that are bound to oxygen do not. Blood contains iron, which is the oxygen-carrying part of hemoglobin. Therefore, de-oxygenated hemoglobin (“deoxyhemoglobin”) has different magnetic properties from oxygenated hemoglobin (“oxyhemoglobin”), where iron atoms are bound to oxygen. As large amounts of freshly oxygenated blood pour into any activated brain region, they cause reduction of the deoxyhemoglobin to oxyhemoglobin ratio. As a consequence, there is a small change in the magnetic field in the active region, and thus a small variation in the MRI signal. This type of image is called Blood-Oxygen-Level-Dependent (BOLD) fMRI. It is important to remember that fMRI is a measure of blood flow and not a direct measure of neural activity. BOLD-fMRI can provide some differentiation between types of tissue; however, the image resolution is limited.

fMRI is generated by a conventional MRI machine, but with T2\*-weighted imaging. T2\* decay is distinct from T2 decay in that T2 decay is the result of random fluctuations in the Larmor frequency at the molecular level, whereas T2\* decay results from larger scale variations in the applied static magnetic field. In practise the magnetic field is not spatially uniform. Spins on one side precess at a different frequency to those on the other side. Since the signal that is detected is the sum of all spins in the sample, the greater the variation in field that exists across the sample, the more rapidly the transverse magnetization will dephase. T2\* is shorter than T1 and ranges from a few milliseconds to tens of milliseconds. Strictly speaking, T2\* processes include the effect of T2 processes.  $\frac{1}{T2^*} = \frac{1}{T2} + \gamma\pi\Delta B_0$  demonstrates the relationship between T2 and T2\*.  $\Delta B_0$  is the field variation across the sample, and  $\gamma$  is the gyromagnetic ratio. The local T2\* is determined by the balance of deoxygenated to oxygenated haemoglobin. T2\* signal is retained longer in a region when it has more oxygenated blood. Thus, an area with more oxygenated blood will show up with higher intensity on T2\*-weighted images.

fMRI offers a better tradeoff between spatial resolution and sensitivity to brain activation. Compared with electrophysiological recording techniques such as electroencephalography (EEG) or event-related potential (ERP) measurements, which

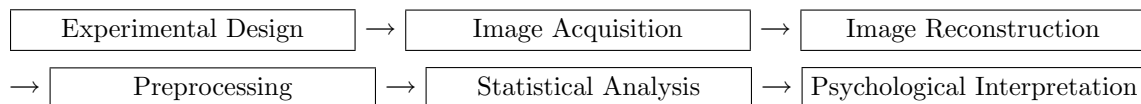
have a better temporal resolution, fMRI provides a better spatial localization and broader “windows” for electrical activity observation. Regional cerebral blood flow (RCBF) based brain mapping techniques such as single-photon emission computed tomography (SPECT) and positron emission tomography (PET) provide better localization, but are less sensitive to brain activation than fMRI. Also, fMRI is less physically demanding on subjects than SPECT or PET, which require injection of a small amount of radioactive drug.

## 2.3 fMRI Applications

fMRI has opened an era of functional studies of the brain. Neuroscientists take advantage of this imaging technique to map brain function into brain anatomy. Scientists are also starting to adopt fMRI in clinical applications, such as neurosurgical planning.

### 2.3.1 Applications in Studies of Neuroscience

With carefully designed psychological experiments, we can understand the neural basis for our behavior, from simple to cognitively complex functions. Simple tasks include stimuli based on presentation of light, colors, tones, chords, music, syllables, movement of fingers, and basic sensory perceptions. More complicated tasks include memory, attention, inhibition, face recognition, communications, etc. The goal is to functionally associate one or more brain regions with the tasks a subject performs. The following shows the general procedure of an fMRI study.



In today’s neural study, block design (box-car) and event-related design are two of the most common experimental techniques.



## **Block Design**

In a block design experiment, stimuli are usually presented at a constant pace over a fixed period of time, alternating with a fixed-length rest period in which no stimuli are presented. Blocks are also referred to as epochs. The experimental protocol is fixed regardless of the subject's response. Because a block is treated as a single unit for analysis, samples within a block should all belong to a single condition. It is important to note that a single condition may include more than one cognitive task. For example, a very popular experiment used in the working memory studies is the two-back word experiment, in which a subject taps his or her fingers together if the number presented is identical to the one that is two back. In this experiment, the subject has multiple cognitive tasks such as listening, memorizing the last two numbers, comparing the current number with the one that is two back, tapping fingers, and so on. Block design considers all of them as a single task condition.

Block design experiment is designed to locate regions with notable difference between these two conditions. Parts of such experiments are contrasted serially to determine the independent effects of the factors of primary interest. The simplest method to do this is the Paired T-test (T-test), which measures the statistical significance of the difference in mean image intensities obtained during the task condition and the rest condition. This assumes that the effects of activation of individual cognitive processing elements are linear additivity in a complex cognitive task. The validity of interpretation of the results therefore depends on the accuracy of this model. Mutual information (MI) based detector has a similar underlying principle as the T-test, but it does not require the linearly addition assumption. Considering each sample as a random variable, MI-based detector compares the different distributions of the task signal and the rest signal. Chapter 3 describes these two methods in detail.

## **Event-Related Design**

While block design has strong detection power, event-related design has the ability to estimate the shape of the hemodynamic response. Built on rapid imaging,

event-related design experiments maximize the opportunity to analyze the data by randomizing and mixing trials or by using differences in speed or accuracy of the subject's response. Event-related design does not generalize signal obtained under all physiological tasks as a single task condition. It assumes that the HRFs corresponding to various physiological tasks are different, signal is analyzed by task category. Event-related design trades improved specificity for reduced sensitivity.

Similarly to the block-design experiments, subjects perform task according to an experimental protocol in event-related experiments. Nevertheless, an event-related design experimental protocol does not necessary have a fixed inter-stimulus length. Instead, stimuli are presented according to a pre-designed varying-speed protocol, to a subject's response time to the previous stimulus, or to the accuracy of the subject's response. The event-related design is also called a single trial design because each trial is statistically independent of the other trials. To be statistically independent, different trial types must be intermixed so that it is impossible to predict the next trial type from the previous one.

### **2.3.2 Other fMRI Applications**

The field of fMRI is rapidly finding clinical application and is exciting scientists from a wide range of disciplines, including neurology, pathology, psychiatry, and psychology. fMRI holds considerable promise for unravelling the neuro-circuitry and metabolic pathways of psychiatric disorders in the immediate future and in helping in psychiatric diagnosis and treatment planning in the long term. Neurosurgical planning, understanding neurological disorders, and drug studies are some of the practices in which fMRI is heavily involved.

Being able to focus on the areas of the brain responsible for tremors or a seizure, fMRI helps make decisions regarding the advisability of surgery and the optimal surgical approach. fMRI can offer a three-dimensional surgical road map that is useful for planning the surgical procedure. During a surgical procedure, fMRI helps orient the neurosurgeon to both the structural and functional features of the anatomy within the limited view of the craniotomy. When intraoperative mapping by direct

cortical stimulation is still deemed necessary, fMRI directs the surgeon to the regions of interest, and shortens the length of the procedure. However, fMRI has potential limitations in the study of brain tumors. The presence of edema, tumor mass, or radiation-induced tissue damage affects the microvasculature surrounding the lesion. One must be careful in interpreting the data. The current most common applications of presurgical fMRI are sensorimotor and language mapping.

fMRI shows valuable significance in early diagnosis for neurological disorders such as schizophrenia, depression, and epilepsy. Applying identical task(s) to the patient group and the control group during fMRI scans, we can investigate which regions of the brain have malfunctions in the patient group. Those regions are critical for neurological disorder diagnostics. Current study shows that the differences are widely distributed in the brain, but a few consistent findings have emerged that hold promise for the development of clinically useful fMRI in psychiatry. For example, the anterior cingulate has been shown to be less active in schizophrenic patients [25], and some evidence shows that depression is associated with hyperactivity in the anterior cingulate cortex and that effective pharmacological treatment is associated with a normalization of activity [38]. In both of these examples, the difference between patients and controls was statistical. Using fMRI for individual diagnoses of mental disorders is not yet a reliable diagnostic test.

fMRI is improving our understanding of a variety of brain pathologies. For example, drug abuse has been shown to be without structural brain changes. Bloom's study [36] reports that nicotine produces a dose-related increase in fMRI signal in specific regions of the brain such as the nucleus accumbens, amygdala, cingulate, and frontal lobes. On the other hand, marijuana suppresses activation in the cerebellum, anterior cingulate, insula, and superior temporal gyrus, while enhancing activation in the globus pallidus, right middle frontal gyrus and areas of parietal cortex.

To conclude, fMRI allows us to see how a brain changes each second. It provides a tool for mapping brain function in neuroscience research. It also has great potential to help improve clinical practise.

## 2.4 Future Development

While researchers and engineers continue to validate current fMRI applications and to improve its reliability, it is expected that there will be a significant improvement in fMRI quality with the rapid developments in the following areas.

MR signal intensity and BOLD contrast are proportional to the external magnetic field strength. Scanners with a higher field strength can achieve sub-millimeter resolution in anatomical images. It can also increase the SNR level in T2\* weighting fMRI which leads to a better temporal resolution and improved sensitivity. One must be careful with the physiological impact from a strong magnetic field.

Another development is to incorporate dynamic perfusion measurements into fMRI imaging technique to quantify the BOLD signal, which is particularly important in fMRI diagnosis, prognosis, and treatment. There are two promising methods: arterial spin labelling (ASL) technique and dynamic perfusion imaging. ASL technique is a type of perfusion MRI that can be used for noninvasive measurement of the regional cerebral blood flow change associated with brain activation. In dynamic perfusion imaging, subjects have injections of exogenous contrast agents such as gadolinium chelate or super-paramagnetic iron oxide nano-particles. These agents can greatly expand the capability of fMRI and provide an accurate measurement of task-induced regional cerebral blood flow change. Either of these methods allows the calculation of relative cerebral blood volume maps which are directly related to the hemodynamic response.

An additional direction for fMRI future development is to integrate it with other functional imaging techniques such as EEG, Magnetic Resonance Spectroscopy (MRS), and Diffusion Tensor imaging (DTI). Combined fMRI and EEG recordings, which take advantage of the greater spatial resolution of fMRI and the superior temporal discrimination of EEG data acquisitions, enable extended forward modelling in fMRI time series. Multi-modality imaging not only enables validation of any one of scanning techniques, but also provides a multi-functional view of the brain.

## 2.5 Summary

This chapter briefly introduced the fMRI imaging technique, listed its wide-range applications, and described its future development. With fMRI, it is possible to link information about brain function directly to information about brain structure in order to understand the physical basis of the abstract mind. In the next chapter, we will discuss current popular detectors for analyzing fMRI signals.



# Chapter 3

## Voxel-by-voxel Detectors

As mentioned in the previous chapter, fMRI signals reflect differences in the paramagnetic properties of oxygenated and deoxygenated blood as a consequence of neural activity [20]. Several detectors, analyzing signal of each voxel separately, have been proposed for fMRI analysis over the years. We refer these detectors as the voxel-by-voxel detectors. Among them, the most basic detector, Paired T-test (T-test), measures the statistical significance of the mean voxel intensities obtained during task and rest conditions, treating each voxel separately. The General Linear Model (GLM) framework augments this approach with a hemodynamic response function, which is used to fit the recorded signal, leading to a more accurate modelling of the response (e.g., providing a mechanism for modelling delays in the signal relative to the onset of the task) [13, 20].

These two detectors have been widely used in fMRI analysis. Recent extensions and alternatives include relaxing the constant signal and the Gaussian noise assumptions of T-test by employing Mutual Information (MI) as a statistic [39], augmenting GLM with a model of temporal dependency in the noise [5], and incorporating spatial continuity constraints on the estimated signal to combat spurious false detection [7, 39].

This chapter compares the features and performance of several current fMRI detectors. The empirical comparison was performed on images from an fMRI study that focuses on an auditory memory task using a block-design protocol, with alternating

Detectors	Temporal Signal Model	Spatial Signal Model	Temporal Noise Model
T-test	constant within condition	–	i.i.d. Gaussian
MI	correlated with protocol	–	i.i.d.
GLM	HRF	–	i.i.d. Gaussian
GLM + Temporal	HRF	–	Gaussian
GLM + Spatial	HRF	MRF or Gaussian model	i.i.d. Gaussian
MI + Spatial	correlated with protocol	MRF or Gaussian model	i.i.d.

Table 3.1: *Modelling assumptions made by different methods.*

rest and task periods [40]. We present the activation maps and examine individual time courses when they offer insight into the differences in behavior of the detectors. In the next section, we provide a brief description of the detectors. Section 3.2 compares selected pairs of detectors with the goal of highlighting the fundamental differences between them.

## 3.1 Background

Many detectors for fMRI analysis in block-design studies essentially measure the difference between voxel response signals, indicated by image intensity, acquired during rest and task periods. Based on the statistical model specific to the method, the estimated differences are translated into significance indicators, such as P-values. Table 3.1 displays the differences in modelling assumptions made by the methods. This section contains a detailed description of the detectors compared in this work, including T-test, GLM, and the MI detector. Since there is no strong evidence of temporal correlation noise in our fMRI data, we did not perform a comparison for detectors that model temporal noise dependency. Details on the detectors that model temporal noise dependency can be found in [5, 11]. On the other hand, we performed extensive empirical evaluations of incorporation of spatial regularization for fMRI detection, and the results are reported in Chapter 4 and Chapter 5.

### 3.1.1 Paired T-test (T-test)

T-test measures the difference between the mean of the signals observed during task periods and those observed during rest periods.  $\vec{y}_{i_{task}}$  and  $\vec{y}_{i_{rest}}$  represent the samples



of voxel  $i$ 's temporal responses under task and rest conditions. Their corresponding sample means are  $\bar{y}_{i_{task}}$  and  $\bar{y}_{i_{rest}}$ , respectively. The detector treats the observed intensities as independent samples from two classes: task and rest. Assuming  $\vec{y}_{i_{task}}$  and  $\vec{y}_{i_{rest}}$  are independent Gaussian variables, we can use the T-distribution, also called the Student distribution [37, 32], to convert the resulting statistic into the corresponding P-value. Under certain assumptions, we can also derive the T-test from the GLM framework (see Appendix A.4).

The T-test's modelling assumption is that the samples drawn from the task and the rest conditions are independently Gaussian distributed.

$$\vec{y}_{i_{task}} \sim \mathcal{N}(\mu_{i_{task}}, \sigma_{i_{task}}^2) \text{ and } \vec{y}_{i_{rest}} \sim \mathcal{N}(\mu_{i_{rest}}, \sigma_{i_{rest}}^2)$$

where,  $\mu_{i_{task}}$  and  $\mu_{i_{rest}}$  are the population means of task and rest conditions;  $\sigma_{i_{task}}^2$  and  $\sigma_{i_{rest}}^2$  are the corresponding population variances.

According to probability theory, the sample means of these two conditions,  $\bar{y}_{i_{task}}$  and  $\bar{y}_{i_{rest}}$ , are also independently Gaussian distributed:

$$\bar{y}_{i_{task}} \sim \mathcal{N}(\mu_{i_{task}}, \frac{\sigma_{i_{task}}^2}{N_{task}}) \text{ and } \bar{y}_{i_{rest}} \sim \mathcal{N}(\mu_{i_{rest}}, \frac{\sigma_{i_{rest}}^2}{N_{rest}})$$

where,  $N_{task}$  and  $N_{rest}$  are the numbers of samples taken during task and rest conditions, respectively. The number of samples taken in an entire trial is  $N_T$ ,  $N_T = N_{task} + N_{rest}$ . It is important to note that according to the Law of Large Number, with sufficient samples, T-test is a valid method even though the samples do not have a Gaussian distribution.

If we let  $M_i = \bar{y}_{i_{task}} - \bar{y}_{i_{rest}}$ , then  $M_i \sim \mathcal{N}(\mu_{i_M}, \sigma_{i_M}^2)$ . Through a simple derivation, we can express  $\mu_{i_M}$  and  $\sigma_{i_M}^2$  in terms of means and variances in the task and rest conditions:  $\mu_{i_M} = \mu_{i_{task}} - \mu_{i_{rest}}$ , and  $\sigma_{i_M}^2 = \frac{\sigma_{i_{task}}^2}{N_{task}} + \frac{\sigma_{i_{rest}}^2}{N_{rest}}$ .

We are interested in whether these two groups of data,  $\vec{y}_{i_{task}}$  and  $\vec{y}_{i_{rest}}$ , have different population means. Hence, we set the hypothesis testing as the following:

$$\begin{aligned} H_0: \mu_{i_{task}} &= \mu_{i_{rest}} \\ H_1: \mu_{i_{task}} &\neq \mu_{i_{rest}}. \end{aligned}$$

The population means are unknown, and we replace them by sample means,  $\bar{y}_{i_{task}}$  and  $\bar{y}_{i_{rest}}$ . Under  $H_0$ , there is no difference in the sample means taken in the two conditions, which implies  $M_i|H_0 \sim \mathcal{N}(0, \sigma_{i_M}^2)$ .

We standardize  $M_i$  by letting

$$T_i = \frac{M_i - \mu_{i_M}}{\sigma_{i_M}}. \quad (3.1)$$

Therefore,  $T_i|H_0 \sim \mathcal{N}(0, 1)$ .

Since population variances,  $\sigma_{i_{task}}^2$  and  $\sigma_{i_{rest}}^2$ , are unknown, we replace them by sample variances, computed according to  $\hat{\sigma}^2 = \frac{\sum_i (x_i - \bar{x})^2}{|X| - 1}$ .  $\hat{\sigma}_{i_{task}}^2$  and  $\hat{\sigma}_{i_{rest}}^2$  are the sample variances for the task and the rest condition, respectively. Depending on whether we assume the two sets have identical population variances, there are two derivations for the T-test.

- Under the assumption that  $\sigma_{i_{task}}^2 = \sigma_{i_{rest}}^2 = \sigma_i^2$ , we can rewrite  $T_i$ 's definition (Eq. (3.1)) as

$$T_i = \frac{\bar{y}_{i_{task}} - \bar{y}_{i_{rest}}}{\sqrt{S_i^2 \left( \frac{1}{N_{task}} + \frac{1}{N_{rest}} \right)}} \quad (3.2)$$

where the pooled variance estimate  $S_i^2$  is defined as the following:

$$S_i^2 = \frac{1}{N_{task} + N_{rest} - 2} \left( \sum_{k \in task} (y_{ik} - \bar{y}_{i_{task}})^2 + \sum_{k \in rest} (y_{ik} - \bar{y}_{i_{rest}})^2 \right). \quad (3.3)$$

Under  $H_0$ ,  $T_i$  is T-distributed with  $N_{task} + N_{rest} - 2$  degrees of freedom. Appendix A provides the proof. Section 3.1.3 shows how to derive T-test formula using a GLM approach with particular assumptions. In other words, T-test is a special case of GLM.

- Alternatively, we proceed without the assumption that  $\sigma_{i_{task}}^2 = \sigma_{i_{rest}}^2$ . This becomes the Behrens-Fisher problem: finding a function of complete sufficient statistics. Mathematicians have developed different methods to solve this problem, such as the iterative likelihood-ratio test. Among these methods, Welch's

approach is the most widely used [6].

$$T_i = \frac{\bar{y}_{i_{task}} - \bar{y}_{i_{rest}}}{\sqrt{\frac{\hat{\sigma}_{i_{task}}^2}{N_{task}} + \frac{\hat{\sigma}_{i_{rest}}^2}{N_{rest}}}} \quad (3.4)$$

where,

$$E \left[ \frac{\hat{\sigma}_{i_{task}}^2}{N_{task}} + \frac{\hat{\sigma}_{i_{rest}}^2}{N_{rest}} \right] = \frac{\sigma_{i_{task}}^2}{N_{task}} + \frac{\sigma_{i_{rest}}^2}{N_{rest}} \quad (3.5)$$

and,

$$Var \left( \frac{\hat{\sigma}_{i_{task}}^2}{N_{task}} + \frac{\hat{\sigma}_{i_{rest}}^2}{N_{rest}} \right) = \frac{2\sigma_{i_{task}}^4}{N_{task}^2(N_{task} - 1)} + \frac{2\sigma_{i_{rest}}^4}{N_{rest}^2(N_{rest} - 1)}. \quad (3.6)$$

Here,  $\frac{\hat{\sigma}_{i_{task}}^2}{N_{task}} + \frac{\hat{\sigma}_{i_{rest}}^2}{N_{rest}}$  is approximately  $\chi^2$  distributed. By matching the first and the second moments,  $T_i|H_0$  is approximately T-distributed using Satterthwaites's approximation [6].

The second derivation of the T-test is more popular in fMRI analysis because there is no indication that the population variances of the samples taken in the task and rest conditions should be equal. In either case, we use the T-distribution to convert  $T_i$  into its corresponding P-value,  $p_i$ , which is used to set the threshold for graphical display of the activation maps:

$$p_i = 2P(T_i > |t_i| | H_0)$$

where,  $t_i$  is one observation. For example,  $t_i$  is a particular T-test result of voxel  $i$  according to Eq. (3.2) or Eq. (3.4). P-value indicates the probability of getting a T-statistic greater than the magnitude of the current observation  $t_i$  under the null hypothesis. We can obtain  $p_i$  from the T-distribution table.

In practice, a confidence level,  $\alpha$ , is selected as a threshold.  $2P(T_i > |t_i| | H_0) < \alpha$  indicates  $H_0$  is unlikely to explain this observation, and  $H_0$  is rejected. We can obtain  $t = t^*$  for a certain  $\alpha$  using the T-distribution table. Therefore, if the T-statistics of voxel  $i$ ,  $T_i$ , is greater than  $t^*$ , we label voxel  $i$  as an active voxel. We can also

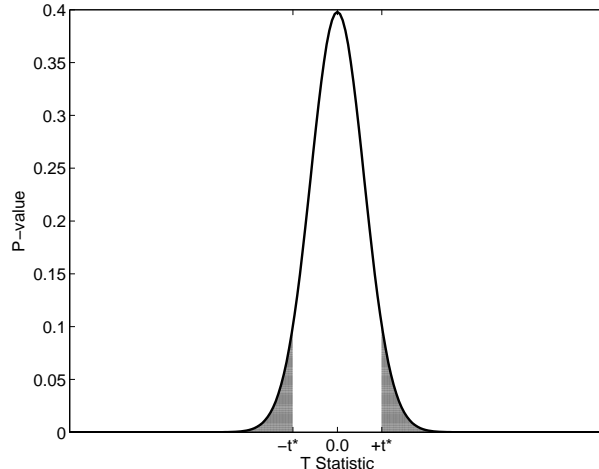


Figure 3-1: *Graphical explanation of converting T-statistic to P-value*

claim that we reject  $H_0$  with  $100(1 - \alpha)\%$  confidence. In fMRI detection,  $\alpha$  is chosen between  $10^{-7}$  and  $10^{-4}$ . Such a low threshold is chosen in order to suppress most false detections out side of the regions of interest (ROIs). In other areas,  $\alpha$  is usually set as 0.05. Either of these choices for  $\alpha$  is entirely arbitrary. Fig. 3-1 shows a graphical explanation of P-value.  $p_i$  corresponds to sum of the area of the shaded regions.

We can consider T-test as a significance test. It measures how likely, in terms of probability,  $\bar{y}_{i_{task}} - \bar{y}_{i_{rest}}$  deviates from zero under the probability density function of  $H_0$ . The false positive rate is  $\alpha$ . It is important to note that T-test does not provide any evidence indicating the alternative hypothesis is true.

### 3.1.2 Mutual Information (MI)

Mutual Information (MI) measures statistical dependence by capturing the differences between the probability distributions of the signal obtained during the task and the rest periods. MI-based activation detection [39] takes a non-parametric approach to model the dependencies between the experimental protocol of the study and the observed fMRI signal, and can therefore detect activations whose dependency on the protocol does not necessarily fit the model of T-test (i.e., non-linearities, partial activations, etc.).

Due to the trade-off between the statistical power and the generalization ability

of a model, MI-based activation maps typically exhibit lower significance values than parametric approaches such as T-test. While this method still assumes temporal independence in the noise, it removes the Gaussian assumption in modelling the noise statistics. MI treats the signal of voxel  $i$  as a stochastic entity,  $Y_i$ , and learns its underlying distribution from the observed data  $y_{i,t}$ ,  $t = [1, \dots, N_T]$ . A voxel is declared active if its intensity and the protocol are statistically dependent, leading to a hypothesis test:

$$H_0(\text{voxel } i \text{ is not active}) : Y_i, \Omega \sim p_{Y_i, \Omega}(y_i, \omega) = p_{Y_i}(y_i) \cdot p_{\Omega}(\omega)$$

$$H_1(\text{voxel } i \text{ is active}) : Y_i, \Omega \sim p_{Y_i, \Omega}(y_i, \omega)$$

where  $\Omega$  is the value of the protocol (0 - rest, 1 - task), and  $\omega_t$  is the protocol at time  $t$ . Since the signal samples of each voxel correspond to the same experimental protocol, there is no subscript for  $\Omega$  and  $\omega$  to indicate different voxels. This leads to a sufficient statistic that is proportional to MI:

$$\log \frac{p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega} | H_1)}{p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega} | H_0)} = \log \frac{\prod_{t=1}^{N_T} p_{Y_i, \Omega}(y_{it}, \omega_t)}{\prod_{t=1}^{N_T} p_{Y_i}(y_{it}) \cdot p_{\Omega}(\omega_t)} \quad (3.7)$$

$$= \sum_{t=1}^{N_T} \log \frac{p_{Y_i, \Omega}(y_{it}, \omega_t)}{p_{Y_i}(y_{it}) \cdot p_{\Omega}(\omega_t)} \quad (3.8)$$

$$= \sum_{t=1}^{N_T} \log p_{Y_i, \Omega}(y_{it}, \omega_t) - \sum_{t=1}^{N_T} \log p_{Y_i}(y_{it}) - \sum_{t=1}^{N_T} \log p_{\Omega}(\omega_t) \quad (3.9)$$

$$\approx N_T E[\log p_{Y_i, \Omega}(y_i, \omega)] - N_T E[\log p_{Y_i}(y_i)] - N_T E[\log p_{\Omega}(\omega)] \quad (3.10)$$

$$= N_T (-h(Y_i, \Omega) + h(Y_i) + H(\Omega)) \quad (3.11)$$

$$= N_T (H(\Omega) - H(\Omega | Y_i)) \quad (3.12)$$

$$= N_T (h(Y_i) - h(Y_i | \Omega)) \quad (3.13)$$

$$= N_T I(Y_i; \Omega). \quad (3.14)$$

Eq. (3.7) is the log likelihood ratio of the two hypotheses.  $I(\cdot; \cdot)$  is the mutual infor-

mation between two random variables.  $H(\Omega)$  is the entropy of the discrete random variable  $\Omega$ , and  $h(Y)$  is the differential entropy of the continuous random variable  $Y$ . We obtain Eq. (3.10) from the Weak Law of Large Numbers. Eq. (3.11) is based on the definition of entropy. We can rewrite Eq. (3.13) as the following:

$$\log \frac{p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega} | H_1)}{p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega} | H_0)} = N_T (h(Y_i) - P(\Omega = 0)h(Y_i | \Omega = 0) - P(\Omega = 1)h(Y_i | \Omega = 1)) \quad (3.15)$$

where,  $P(\Omega = 1)$  and  $P(\Omega = 0)$  indicate the how often task and rest conditions appear during the experiment.

From the basic information theory, we get  $0 \leq H(\Omega | Y_i) \leq H(\Omega) \leq 1$ . Hence,  $0 \leq I(Y_i; \Omega) \leq 1$ . We can consider that MI is a normalized measure of dependency between  $Y_i$  and  $\Omega$  with a high mutual information near 1 bit, indicating that the voxel is activated.

We estimate the unknown probability density function of  $Y_i$  using the Parzen density estimator.

$$h(Y_i) = - \int p_{Y_i}(y_i) \log p_{Y_i}(y_i) dy_i \quad (3.16)$$

$$= -E[\log p_{Y_i}(y_i)] \quad (3.17)$$

$$\approx -\frac{1}{N_T} \sum_{t=1}^{N_T} \log p_{Y_i}(y_{it}) \quad (3.18)$$

$$\approx -\frac{1}{N_T} \sum_{t=1}^{N_T} \log \hat{p}_{Y_i}(y_{it}) \quad (3.19)$$

$$= -\frac{1}{N_T} \sum_{t=1}^{N_T} \log \frac{1}{N_T} \sum_{s=1}^{N_T} K(y_{it} - y_{is}, \kappa). \quad (3.20)$$

The estimated entropy is affected by the choice of  $\kappa$ , which is commonly referred to as the kernel size or smoothing parameter. We can obtain  $\kappa$  by the leave-one-out maximum likelihood estimator [22].

In order to quantitatively compare MI with the parametric detectors, we used per-

mutations to convert the MI statistic into its corresponding P-value at each voxel [16]. In each iteration of the procedure, we permute the protocol values and re-estimate the MI of this new association between the protocol values and the activation intensity. After a large number of iterations, the histogram of MI values can be used to estimate the probability of obtaining the observed MI value (or higher) by chance, using one of such randomly generated associations. We use a fast procedure based on spacings estimates of entropy [26] for evaluating MI in this procedure.

### 3.1.3 General Linear Model (GLM)

GLM models the brain as a Linear Time Invariant (LTI) system with an impulse response function that reflects the hemodynamic properties of the brain tissue. A design matrix  $B = [B_1 \ B_2]$  is used for linear regression. GLM assumes the signal is a linear combination of a protocol-dependent component,  $B_1$ , and a protocol-independent component,  $B_2$ , such as physiological noise and drifting.

$B_1$  is constructed based on the experimental protocol and the assumed Hemodynamic Response Function (HRF). Two-gamma function [20]<sup>1</sup>, sinusoid functions, and Finite Impulse Response (FIR) functions are popular regressor models for HRF. Including assumptions of a specific form of HRF (e.g., two-gamma function) can increase the sensitivity of the detector, though it is important to note that the detector could not explain all the signals associated with the stimulus if the assumed shape was wrong. Low order polynomials and low frequency Fourier basis are often employed to model the protocol-independent component  $B_2$ .

Voxels are declared to be active if the significant power of the observed response is in the subspace of the active response. We will later show that the T- and the F-statistics are commonly used to evaluate the significance. GLM is usually formulated as

$$\vec{y}_i = B_1\vec{\beta}_{1i} + B_2\vec{\beta}_{2i} + \vec{\varepsilon}_i = B\vec{\beta}_i + \vec{\varepsilon}_i. \quad (3.21)$$

---

<sup>1</sup>A two-gamma function captures the fact that there is a small dip after the HRF has returned to zero:  $h(t) = (t/d_1)^{a_1} \exp(-\frac{t-d_1}{b_1}) - c(t/d_2)^{a_2} \exp(-\frac{t-d_2}{b_2})$  where,  $d_j = a_j b_j$  is the time to the peak, and  $a_1 = 6$ ,  $a_2 = 12$ ,  $b_1 = b_2 = 0.9s$ , and  $c = 0.35$

$\vec{y}_i$  is the signal at voxel  $i$ , noise  $\vec{\varepsilon}_i \sim \mathcal{N}(\vec{0}, \Sigma_i)$ , but  $\Sigma_i$  is unknown.  $\vec{\beta}_i = [\vec{\beta}_{1_i}, \vec{\beta}_{2_i}]$  is a vector of estimated amplitudes of the hemodynamic responses and the protocol-independent signals. Depending on the regression setting, either  $\vec{\beta}_{1_i}$ ,  $\vec{\beta}_{2_i}$ , or both may be a scalar. To get a generalized derivation, we assume  $\vec{\beta}_{1_i}$  and  $\vec{\beta}_{2_i}$  are vectors. Assuming that noise is white,  $\Sigma_i = \sigma_i^2 I$ , we can estimate  $\beta$  using the ordinary least square estimate<sup>2</sup>:

$$\hat{\vec{\beta}}_i = (B^T B)^{-1} B^T \vec{y}_i. \quad (3.22)$$

The residual error is defined as,

$$\vec{r}_i = \vec{y}_i - \hat{\vec{y}}_i \quad (3.23)$$

$$= \vec{y}_i - B \hat{\vec{\beta}}_i \quad (3.24)$$

$$= (I - B(B^T B)^{-1} B^T) \vec{y}_i = R \vec{y}_i \quad (3.25)$$

where,  $R$  is an idempotent matrix,  $RR = R$ . We can calculate the residual error variance as

$$\sigma_{r_i}^2 = \frac{\vec{r}_i^T \vec{r}_i}{N_T - N_\beta}. \quad (3.26)$$

where the normalization factor,  $N_T - N_\beta$ , is required because there are  $N_T - N_\beta$  degrees of freedom left after estimating  $\beta$ .  $N_T$  is the number of samples in a time course, and  $N_\beta$  is the number of regressors in  $\vec{\beta}_i$ . We can further show that  $R$  is not a full rank matrix. To do so, we first perform the singular value decomposition on  $B$ ,  $B = USV^T$ . Since  $B$  is an  $N_T \times N_\beta$  matrix,  $U$  is  $N_T \times N_T$ ,  $S$  is  $N_T \times N_\beta$ , and  $V$  is  $N_\beta \times N_\beta$ .  $U$  and  $V$  store orthogonal basis in each column entry for an  $N_T \times N_T$ -dimensional space and an  $N_\beta \times N_\beta$ -dimensional space, respectively.  $S$  stores  $N_\beta$  singular values along the diagonal of the first  $N_\beta$  rows.  $R$  can be decomposed as the following:

---

<sup>2</sup>Without the white noise assumption, Eq. (3.22) still provides an unbiased estimate of  $\vec{\beta}$ .



$$\begin{aligned}
R &= I - USV^T(VS^T U^T USV^T)^{-1}VS^T U^T \\
&= I - USV^T(VS^T SV^T)^{-1}VS^T U^T \\
&= I - USV^T V^{-T} \Lambda^{-1} V^{-1} VS^T U^T \\
&= I - US \Lambda^{-1} S^T U^T \\
&= I - U' U'^T \\
&= U'' U''^T.
\end{aligned} \tag{3.27}$$

We let the orthogonal matrix  $U = [\vec{u}_1, \vec{u}_2, \dots, \vec{u}_{N_\beta}, \dots, \vec{u}_{N_T}]$ .  $U'$  and  $U''$  are also  $N_T \times N_T$  matrices.  $U' = [\vec{u}_1, \vec{u}_2, \dots, \vec{u}_{N_\beta}, 0, \dots, 0]$  and  $U'' = U - U'$ .  $U'$  has rank  $N_\beta$ , and  $U''$  has rank  $N_T - N_\beta$ . Therefore,  $R$  is also a rank  $N_T - N_\beta$  matrix. This non-full rank matrix prevents inversion, so  $\Sigma_i$  cannot be derived from  $Cov(\vec{r}_i)$ , requiring alternative ways to model the noise covariance matrix, such as auto-regressive noise models and deterministic component identification. Detailed discussion on the noise models can be found in [5, 11].

After estimation, we perform hypothesis testing to detect any voxel whose signal corresponds to the protocol. Vector  $\hat{\vec{\beta}}_{1_i}$  or linear combination of the elements of this vector, depending on the test objective, is usually employed as the indicator for this evidence. A contrast matrix,  $C$ , is introduced according the test objective. For example, if we define activation as any element in  $\hat{\vec{\beta}}_{1_i}$  being significantly different from zeros, that means  $H_0 : \hat{\vec{\beta}}_{1_i} = \vec{0}$ . If  $\hat{\vec{\beta}}_{1_i}$  is a vector, we need to make a simultaneous test for each element. It can be proved that the confidence interval testing is F-distributed (Eq. (3.28)).

$$F_i = \frac{(C\hat{\vec{\beta}}_{1_i})^T (Cov(C\hat{\vec{\beta}}_{1_i}))^{-1} (C\hat{\vec{\beta}}_{1_i})}{N_{\beta_1}} \tag{3.28}$$

Under  $H_0$ ,  $F_i$  is F-distributed,  $F_i|H_0 \sim F_{N_{\beta_1}, N_T - N_\beta}$ . In this case,  $C$  is an  $N_{\beta_1} \times N_{\beta_1}$  identity matrix, and  $N_{\beta_1}$  is the number of regressors in  $\hat{\vec{\beta}}_{1_i}$ . On the other hand, if  $\hat{\vec{\beta}}_{1_i}$  is a scalar or the test objective is a linear combination of the elements in vector  $\hat{\vec{\beta}}_{1_i}$ ,

the confidence interval test follows T-distribution (Eq. (3.29)).

$$T_i = \frac{(C\hat{\vec{\beta}}_{1_i})}{\text{var}(C\hat{\vec{\beta}}_{1_i})} \quad (3.29)$$

Under  $H_0$ ,  $T_i$  is T-distributed,  $T_i|H_0 \sim T_{N_T - N_\beta}$ . The corresponding contrast matrix,  $C$ , is a row vector.

In addition, we can re-derive GLM using the maximum likelihood approach. The presence of the protocol-dependent signal indicates that the corresponding voxel is active due to the stimulus. This definition is equivalent to any element in  $\hat{\vec{\beta}}_1$  being significantly different from zeros.

$$\begin{aligned} H_0: \vec{y}_i &= B_2\vec{\beta}_{2_i} + \vec{\epsilon}_i \text{ (or } \vec{\beta}_{1_i} = \vec{0}) \\ H_1: \vec{y}_i &= B_1\vec{\beta}_{1_i} + B_2\vec{\beta}_{2_i} + \vec{\epsilon}_i \text{ (or } \vec{\beta}_{1_i} \neq \vec{0}) \end{aligned}$$

where,  $\vec{\epsilon}_i \sim \mathcal{N}(\vec{0}, \sigma_i^2 I)$ . We can express the signal by the sufficient statistic of the two hypotheses, the maximum log likelihood ratio  $z_i$ ,

$$z_i = \log \frac{\max_{\vec{\beta}_{1_i}, \vec{\beta}_{2_i}, \sigma_i^2} p(\vec{y}_i | H_1)}{\max_{\vec{\beta}_{2_i}, \sigma_i^2} p(\vec{y}_i | H_0)} \quad (3.30)$$

$$= \log \frac{\max_{\vec{\beta}_{1_i}, \vec{\beta}_{2_i}, \sigma_i^2} \mathcal{N}(\vec{y}_i; B_1\vec{\beta}_{1_i} + B_2\vec{\beta}_{2_i}, \sigma_i^2 I)}{\max_{\vec{\beta}_{2_i}, \sigma_i^2} \mathcal{N}(\vec{y}_i; B_2\vec{\beta}_{2_i}, \sigma_i^2 I)}. \quad (3.31)$$

The Maximum Likelihood (ML) estimates of the numerator are

$$[\hat{\vec{\beta}}_{1_i} \ \hat{\vec{\beta}}_{2_i}] = \hat{\vec{\beta}}_i = \arg \max_{\vec{\beta}_i} \log \mathcal{N}(\vec{y}_i; B\vec{\beta}_i, \sigma_i^2 I) = (B^T B)^{-1} B^T \vec{y}_i \quad (3.32)$$

and

$$\hat{\sigma}_i^2 = \arg \max_{\sigma_i^2} \log \mathcal{N}(\vec{y}_i; B\hat{\vec{\beta}}_i, \sigma_i^2 I) = \frac{\|\vec{y}_i - B\hat{\vec{\beta}}_i\|^2}{N_T}. \quad (3.33)$$

Similarly, the ML estimates for the denominator are  $\hat{\vec{\beta}}_{2_i} = (B_2^T B_2)^{-1} B_2^T \vec{y}_i$  and  $\hat{\sigma}_i^2 = \frac{\|\vec{y}_i - B_2\hat{\vec{\beta}}_{2_i}\|^2}{N_T}$ .

By substituting the ML estimates into Eq. (3.30), we get

$$z_i = \log \frac{(\|\vec{y}_i - B_2 \hat{\beta}_{2_i}\|)^{N_T}}{(\|\vec{y}_i - B \hat{\beta}_i\|)^{N_T}} \quad (3.34)$$

$$= \frac{N_T}{2} \log \frac{\vec{y}_i^T (B(B^T B)^{-1} B^T - B_2(B_2^T B_2)^{-1} B_2^T) \vec{y}_i}{\vec{y}_i^T (I - B(B^T B)^{-1} B^T) \vec{y}_i} \quad (3.35)$$

where,  $z_i$  is a monotonically increasing function of the  $F_i$  (Eq. (3.28)) in the conventional GLM detector (see Cosman [7] for detailed derivations). We can also consider  $z_i$  as a noisy statistic indicating a voxel's activation state.

## 3.2 Experiments and Results

We compared the detectors on the fMRI scans obtained during an auditory memory experiment. Each experiment consisted of five rest epochs and four task epochs, each epoch 30 seconds long. In the rest condition, the subjects were instructed to concentrate on the noise of the scanner and not to move. In the task condition, the subjects were presented with a series of pre-recorded single-digit numbers, three seconds per number. The subjects were asked to tap their index finger to their thumb when hearing a number that was the same as the one spoken two numbers before. Experiments were repeated ten times for each subject. The original study contained nine subjects, but for the purpose of voxel-by-voxel comparison of the detectors, we present the results for one subject across all detectors. T1-weighted structural anatomical images were acquired for each subject on a 1.5 Tesla GE signa clinical MR scanner using a 3D-SPGR sequence. Axial EPI images were acquired on the same scanner with imaging parameters as follows: TR/TE=2500/50 msec, flip angle=90, 64x64 acquisition matrix, 24 slices, slice thickness=6mm. More details on this study can be found in [40], which also presents the results of GLM-based detection performed in SPM [13], one of the standard software packages for fMRI analysis.

### 3.2.1 T-test vs. MI

While T-test compares the means of the conditional intensity distributions for the two conditions, implicitly assuming two Gaussian distributions, the MI compares the

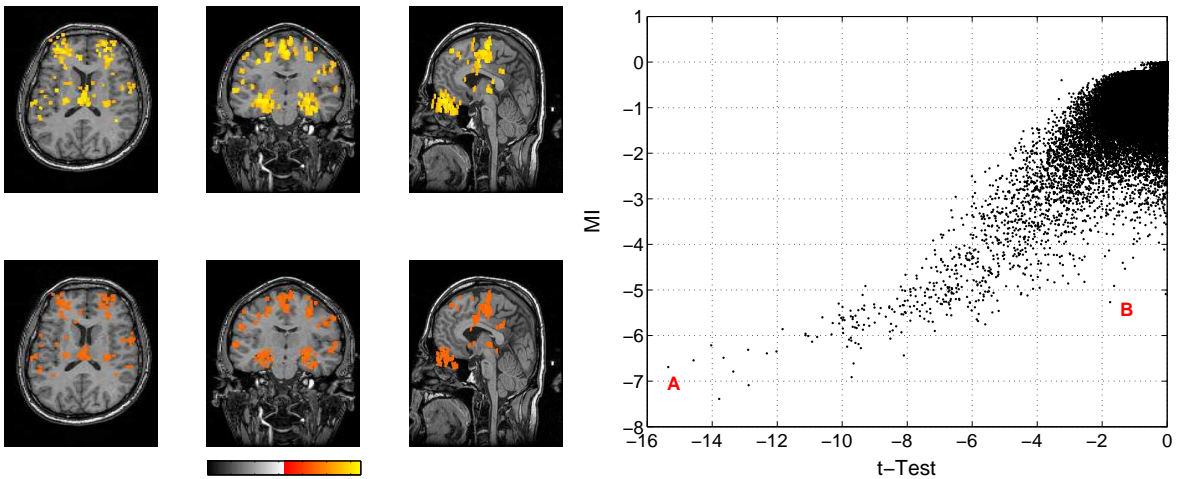


Figure 3-2: *Left: maximum intensity projections (axial, coronal and sagittal) of the activation maps detected by T-test (top) and MI (bottom). The colormap is used to show the negative logarithm of P-values. Right: joint scatter of P-values (log-scale) produced by T-test and MI.*

shapes of the distributions. Thus we would expect T-test to yield a higher statistical power (lower P-values) for voxels with a substantial difference between the two conditions, and MI to be able to detect subtle dependencies between the voxel intensities and the protocol that are typically missed by T-test.

Fig. 3-2 illustrates the detection results for the two detectors. The images on the left show the activation maps obtained by thresholding the P-values. The plot on the right displays the joint scatter of the (log) P-values estimated by the two methods. As expected, strongly activated voxels are assigned much lower P-values by T-test than by MI. We also observe that the mapping between the P-values is not linear and some voxels are detected by MI that would not pass a threshold based on T-test. Fig. 3-3 illustrates this point for the two voxels, A and B, marked on the scatter plot of Fig. 3-2. We observe that voxel A is detected as highly significant by both methods as its time course and the corresponding conditional intensity distributions fit the Gaussian model fairly well. Voxel B corresponds to what we call a *partial activation*. Its time course seems to have a global trend of declining activation, but in the first several blocks, its intensity is definitely synchronized with the protocol. As a result, the MI detector will label this voxel as active for a substantially more conservative threshold relative to the most active voxels in the volume. This leads us to conclude

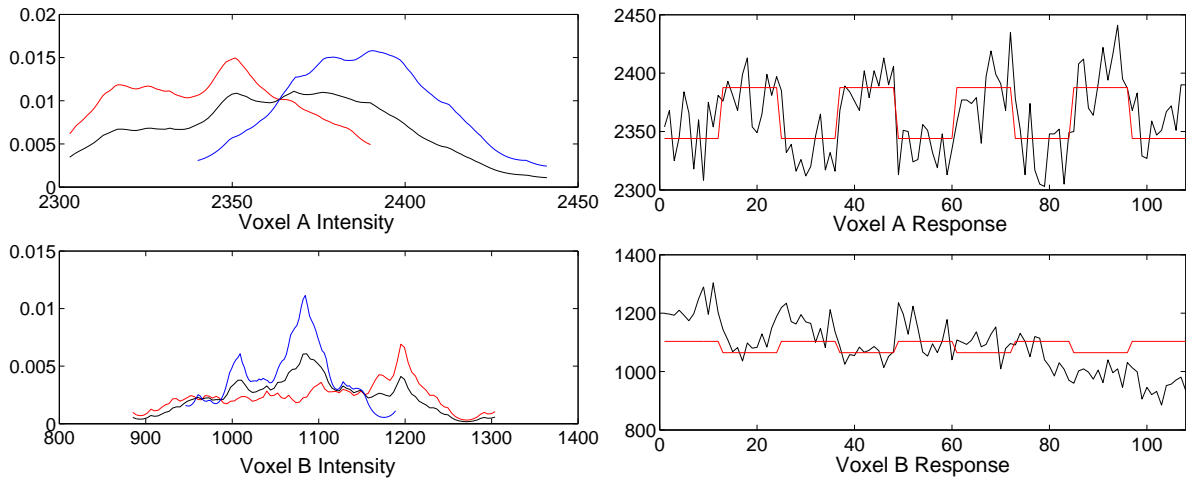


Figure 3-3: *Left: intensity histograms during task ( $p_{Y|U=1}(y)$ , blue), during rest ( $p_{Y|U=0}(y)$ , red), and over the entire run ( $p_Y(y)$ , black) for two voxels from the right plot in Fig. 3-2. Voxel A:  $P_{T-test} = 10^{-15}$ ,  $P_{MI} = 10^{-7}$ ; Voxel B:  $P_{T-test} = 10^{-2}$ ,  $P_{MI} = 10^{-5}$ . Right: time course (black) and scaled protocol timeline (red) for the same voxels.*

that although the statistical significance of the most active voxels is reduced in MI-based detection, the method can successfully detect the areas that are only partially correlated with the protocol.

### 3.2.2 T-Test vs. GLM

GLM extends the statistical framework of T-test to include a model for the HRF, yielding improved statistical significance (lower P-values) for the voxels whose temporal response fits the specified model. Appendix A.4 shows that we can derive T-test formula from GLM with certain assumptions. This section empirically illustrates the difference of these two detectors.

Fig. 3-4 illustrates the detection results for both methods. Examining the activation maps, we note that incorporating a temporal response model allows GLM to discriminate between the cortical activation and the artifacts caused by the eye motion. It also improves detection in the cortical area, leading to more contiguous activation maps. Fig. 3-5 illustrates the detailed differences between the two methods on the four voxels highlighted on the joint scatter plot of Fig. 3-4. We first note that the regressor in the GLM estimation (a convolution of the protocol with the HRF, shown in blue), while similar to the protocol, models the delay of the response and the

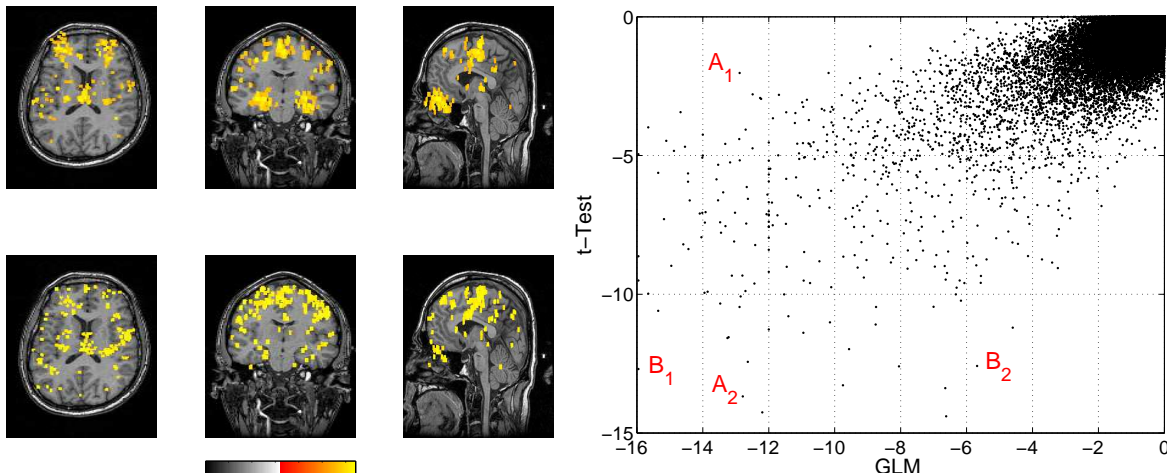


Figure 3-4: *Left: maximum intensity projection (axial, coronal and sagittal) of the activation maps detected by T-test (top) and GLM (bottom). The colormap is used to show the negative logarithm of P-values. Right: joint scatter of P-values (log-scale) produced by T-test and GLM.*

slight over- and under-shoot in the intensity at the boundaries between the blocks. The temporal responses of voxels  $A_1$  and  $A_2$  have a similar quality fit to the proposed HRF model, leading to similar P-values in GLM detection. However, the response of  $A_1$  is delayed relative to the protocol and is deemed much less significant under the T-test. This points to a problem in the T-test detection that is often mitigated by omitting several time points on each side of the block boundaries. However, this practise discards valuable measurements and typically reduces the detector's sensitivity. On the other hand, voxels  $B_1$  and  $B_2$  are assigned similar P-values by T-test, but  $B_1$  fits the HRF model better, and is declared more significant under the GLM. This example highlights the importance of choosing the correct shape of the HRF to be fitted to the voxel time courses. The response function might actually vary spatially in the brain, which implies that we might get better detection by using less restrictive HRF models, even if it leads to potentially reduced statistical power at the most active voxels.

Modelling the temporal correlation of the noise in fMRI signals is out of the scope of this project. Several different approaches have been proposed in [5, 11].

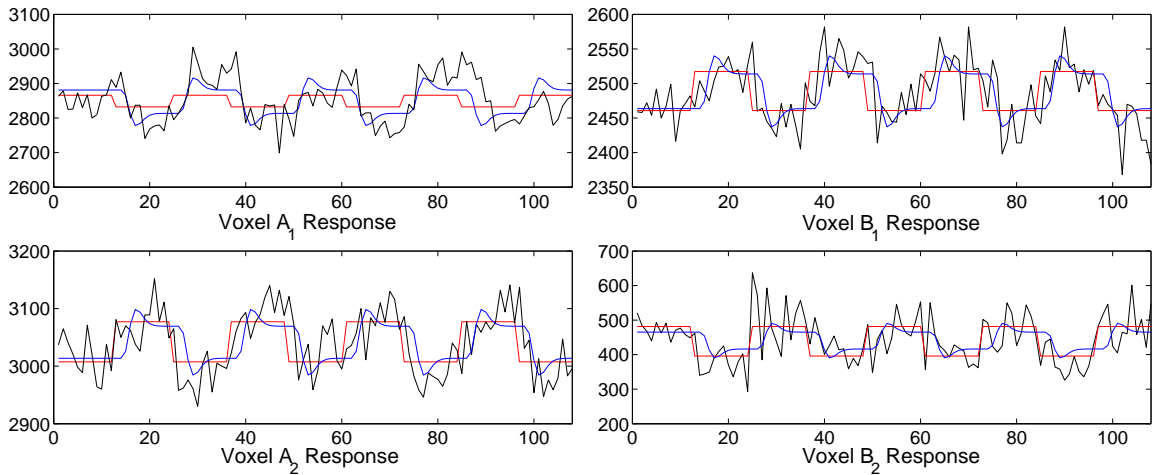


Figure 3-5: *Time course (black), T-test protocol (red), and the GLM regressor (blue) for four voxels from Fig. 3-4. Voxel A<sub>1</sub>:  $P_{T-test} = 10^{-2}$ ,  $P_{GLM} = 10^{-13}$ ; Voxel A<sub>2</sub>:  $P_{T-test} = 10^{-14}$ ,  $P_{GLM} = 10^{-13}$ ; Voxel B<sub>1</sub>:  $P_{T-test} = 10^{-13}$ ,  $P_{GLM} = 10^{-16}$ ; Voxel B<sub>2</sub>:  $P_{T-test} = 10^{-13}$ ,  $P_{GLM} = 10^{-6}$ .*

### 3.3 Summary

In this chapter, we discussed the implicit modelling assumptions made by several fMRI detectors used in the field and compared them empirically on a real fMRI example. The experiments illustrate some of the theoretical points for specific voxels in the images. In general, the trade-off between the statistical power gained by stronger modelling assumptions and the ability to detect activations that violate these assumptions was very pronounced throughout the experiments. GLM improves detection power by explicitly modelling the system response, which causes it to miss the areas whose responses do not fit the pre-specified shape of the response function. MI, as a non-parametric technique, necessitates giving up some of the statistical power for the ability to detect activations that are weakly coupled with the protocol. Such methods can be useful in exploratory studies where the nature of the dependency between the protocol and the activation is either not known or is suspected to be nonlinear, such as in some of the auditory cortex activations.

This study points to several promising directions for research in this area. Incorporating time response into MI-based estimation and detection can potentially improve the quality of the detector, similar to the improvement we saw in the GLM over the T-test. Another direction is to employ spatial modelling to combat noise in

the activation. We expect that the GLM- and MI-based detections can be further strengthened by incorporating more sophisticated models of spatial structure of the signal into the detectors. The next two chapters discuss such spatial models and their benefits for fMRI detection.



## Chapter 4

# Spatial Regularization for fMRI Detection

As discussed in Chapter 3, the voxel-by-voxel fMRI detectors compare the time course of each voxel with the experimental protocol, labelling those voxels whose time courses correlate significantly with the protocol as “active”. The analysis of each voxel with these detectors results in the statistical parametric map (SPM). A threshold or a pair of thresholds applied to the SPM produces a binary or a trinary map of active areas, and this binary or trinary map is also called the activation map. It has been proven that some parts of the brain are de-active associated with certain tasks that a subject performs. The trinary activation maps can encode this information. As we saw in the previous chapter, because of a low signal-to-noise ratio (SNR) in fMRI signal, the activation map typically contains many small false positive islands.

To reduce these false positive islands, most researchers take advantage of biological models that imply contiguous activation areas, choosing to smooth the signal spatially using a Gaussian filter prior to detection. Unfortunately, Gaussian smoothing, though intended to combat low SNR, leads to overly smoothed SPMs and a loss of detail in the resulting activation maps. A number of alternative approaches have explicitly incorporated spatial and temporal correlations into the estimation procedure. Examples include autoregressive spatio-temporal models [41, 5], Markov random fields (MRFs) [10, 9, 7], Bayesian models inferring hidden psychological states [19], and

adaptive thresholding methods that adjust the significance of active regions according to their size, based on the Gaussian random field theory [12]. In this chapter, we focus on MRFs for modelling spatial coherency and study their performance. We assume that, given the activation state of each voxel, the time courses of different voxels are independent and can be reduced to a sufficient statistic. Therefore we concentrate on spatial regularization of the activation map. Temporal regularization models can be easily incorporated into our framework by changing the activation statistic.

For MRFs with binary states, an exact solution can be obtained in polynomial time [17]. Kim [39] introduced a fMRI detector by augmenting MI statistics with a binary MRF prior; Cosman [7] introduced a similar detector by augmenting GLM statistics with a binary MRF prior. However, if one wants to go beyond binary states (e.g. treating positively and negatively activated voxels differently), the problem of estimating the optimal activation states becomes intractable and approximation algorithms must be used. Prior work in MRF-based fMRI detection employed simulated annealing [10, 31] or iterated conditional modes [33]. We adopt the Mean Field solver, introduced in statistical physics [27] and widely used for image segmentation [21, 30, 42, 23]. In our experiments with binary MRFs, Mean Field algorithm produced results comparable to those of the exact solver while reducing computation time by one to two orders of magnitude. In trinary MRFs, its computation time is on the same order of magnitude as that of the binary case.

We also experimented extensively with the Belief Propagation algorithm [28], which often produces a better approximation, but did not find it to be more accurate in this application. Treating messages from different neighbors independently, Belief Propagation does not provide accurate approximation for graphs with many small loops. We therefore primarily present the experimental results using the Mean Field algorithm.

The next two sections explain how the GLM and MI detectors can be augmented with an MRF prior. In the experiments, we focus on the results generated from the GLM detector because it is the most popular fMRI detector in neuroscience. We then review the Mean Field solution [27] and explain how to set the model parameters.

Section 4.5 presents the empirical evaluation of the method using synthetic data. In Chapter 5, we will combine the evaluations of the current detectors with the refined detectors introduced in Chapter 5 on a real fMRI data set.

## 4.1 Background Statistics

In this section, we revisit two voxel-by-voxel detectors, GLM and MI, which are discussed in Chapter 3. We omit the T-test here, because it can be derived as a special case of GLM (see Append A.4 for details). We will show how to incorporate an MRF prior into the ML estimates of these two detectors.

### 4.1.1 Sufficient Statistic for the General Linear Model

In Section 3.1.3, we derived GLM through the maximum likelihood approach. We write the two hypotheses as the following:

$$\begin{aligned} H_0: \vec{y}_i &= B_2 \vec{\beta}_{2_i} + \vec{\epsilon}_i \text{ (or } \vec{\beta}_{1_i} = \vec{0}) \\ H_1: \vec{y}_i &= B_1 \vec{\beta}_{1_i} + B_2 \vec{\beta}_{2_i} + \vec{\epsilon}_i \text{ (or } \vec{\beta}_{1_i} \neq \vec{0}) \end{aligned}$$

where  $\vec{\epsilon}_i \sim \mathcal{N}(\vec{0}, \sigma^2 I)$ . The subscript  $i$  specifies voxel  $i$ .  $B_1$  represents the protocol-dependent component, and  $B_2$  presents the protocol-independent component. Their corresponding regression coefficient vectors are  $\vec{\beta}_{1_i}$  and  $\vec{\beta}_{2_i}$ . We obtain the maximum log likelihood ratio as

$$z_i = \log \frac{\max_{\vec{\beta}_{1_i}, \vec{\beta}_{2_i}, \sigma_i^2} p(\vec{y}_i | H_1)}{\max_{\vec{\beta}_{2_i}, \sigma_i^2} p(\vec{y}_i | H_0)} \quad (4.1)$$

$$= \log \frac{\max_{\vec{\beta}_{1_i}, \vec{\beta}_{2_i}, \sigma_i^2} \mathcal{N}(\vec{y}_i; B_1 \vec{\beta}_{1_i} + B_2 \vec{\beta}_{2_i}, \sigma_i^2 I)}{\max_{\vec{\beta}_{2_i}, \sigma_i^2} \mathcal{N}(\vec{y}_i; B_2 \vec{\beta}_{2_i}, \sigma_i^2 I)} \quad (4.2)$$

$$= \log \frac{(\|\vec{y}_i - B_2 \hat{\vec{\beta}}_{2_i}\|)^{N_T}}{(\|\vec{y}_i - B \hat{\vec{\beta}}_i\|)^{N_T}} \quad (4.3)$$

$$= \frac{N_T}{2} \log \frac{\vec{y}_i^T (B(B^T B)^{-1} B^T - B_2 (B_2^T B_2)^{-1} B_2^T) \vec{y}_i}{\vec{y}_i^T (I - B(B^T B)^{-1} B^T) \vec{y}_i} \quad (4.4)$$

where  $z_i$  is a monotonically increasing function of the F statistic (Eq. (3.28)) in the conventional GLM detector (see Section 3.1.3 for detailed derivation). We can also consider  $z_i$  as a noisy statistic indicating voxel  $i$ 's activation state.

### 4.1.2 Sufficient Statistic for the Mutual Information Detector

As shown in Section 4.1.2, we declare a voxel active if there is significant dependency between its signal and the experimental protocol. Assuming each sample in a time sequence as an observation of the voxel's signal, we can form the following hypotheses:

$$H_0(\text{voxel } i \text{ is not active}) : \vec{Y}_i, \vec{\Omega} \sim p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega}) = p_{\vec{Y}_i}(\vec{y}_i) \cdot p_{\vec{\Omega}}(\vec{\omega})$$

$$H_1(\text{voxel } i \text{ is active}) : \vec{Y}_i, \vec{\Omega} \sim p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega})$$

where,  $\Omega$  is the value of the protocol (0 - rest, 1 - task), and  $\omega_t$  is the protocol signal at time  $t$ . This leads to a sufficient statistic that is proportional to MI:

$$z_i = \log \frac{p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega} | H_1)}{p_{\vec{Y}_i, \vec{\Omega}}(\vec{y}_i, \vec{\omega} | H_0)} \quad (4.5)$$

$$= \log \frac{\prod_{t=1}^{N_T} p_{Y_i, \Omega}(y_{it}, \omega_t)}{\prod_{t=1}^{N_T} p_{Y_i}(y_{it}) \cdot p_{\Omega}(\omega_t)} \quad (4.6)$$

$$= N_T I(Y_i; \Omega) \quad (4.7)$$

where,  $I(\cdot; \cdot)$  is the mutual information between two random variables.

## 4.2 Markov Priors

We let  $\vec{X} = [X_1, \dots, X_{N_V}]$  be the activation configuration of all voxels. Random variable  $X_i$  denotes voxel  $i$ 's discrete activation state. In this work, we focus on  $X_i$  being either a binary or a trinary random variable. The binary  $X_i$  indicates active or not active states; the trinary  $X_i$  indicates positively active, negatively active, and

not active states.  $\vec{x} = [x_1, \dots, x_{N_V}]$  is one possible configuration. Here, we assume an MRF prior on  $\vec{X}$ . According to Hammersley-Clifford theorem [4],  $\vec{X}$  follows the Gibbs distribution:

$$P_{\vec{X}}(\vec{x}) = \frac{1}{\lambda} \prod_{\langle i,j \rangle} \Psi_{ij}(x_i, x_j) \prod_i \Psi_i(x_i). \quad (4.8)$$

Eq. (4.8) is defined in terms of the singleton potentials  $\Psi_i(x_i)$  that provide bias over state values  $x_i$  for voxel  $i$ , and the pairwise potentials  $\Psi_{ij}(x_i, x_j)$  (often referred to as the compatibility matrices) that evaluate the compatibility of voxel  $i$  being in state  $x_i$  and voxel  $j$  being in state  $x_j$  for each pair  $\langle i, j \rangle$  of neighboring voxels.  $\lambda$  is a normalization constant, also called the partition function. Since there are  $2^{N_V}$  and  $3^{N_V}$  possible configurations in binary and trinary MRFs, it is an NP-hard problem to compute  $\lambda$ . Given the activation statistic values  $\vec{z}$ , we seek the maximum *a posteriori* (MAP) estimate of the activation configuration:

$$\vec{x}^* = \arg \max_{\vec{x}} P_{\vec{X}|\vec{Z}}(\vec{x}|\vec{z}) \quad (4.9)$$

$$= \arg \max_{\vec{x}} P_{\vec{X}, \vec{Z}}(\vec{x}, \vec{z}) \quad (4.10)$$

$$= \arg \max_{\vec{x}} P_{\vec{X}}(\vec{x}) P_{\vec{Z}|\vec{X}}(\vec{z}|\vec{x}) \quad (4.11)$$

$$= \arg \max_{\vec{x}} \frac{1}{\lambda} \prod_{\langle i,j \rangle} \Psi_{ij}(x_i, x_j) \prod_i \Psi_i(x_i) P_{Z_i|X_i}(z_i|x_i). \quad (4.12)$$

We consider the statistic from either the GLM detector or the MI detector as a random vector  $\vec{Z} = [Z_1, \dots, Z_N]$ .  $z_i$  is the statistic we obtain from voxel  $i$  using either of the two detectors (Eq. (4.1) and Eq. (4.5)). Eq. (4.12) is obtained based on the assumption that the noise statistics at different voxels are independent given the activation state of each voxel, and the likelihood over the volume can therefore be written as a product of the individual likelihood terms for each voxel.

Fig. 4-1 depicts the corresponding graphical model, using a two-dimensional grid for illustration purposes only. The estimation is performed fully in 3D in all experiments reported here. We assume a spatially stationary generative model, i.e.,  $P_{Z_i|X_i}$ ,  $\Psi_i$ , and  $\Psi_{ij}$  are identical for all voxels in the volume. The observations (the fMRI

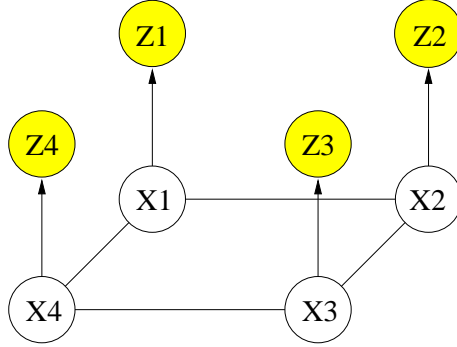


Figure 4-1: *Graphical model for  $P_{\bar{x}, \bar{z}}$ .*

signal, and in Chapter 5, the anatomical information) move the MAP estimate away from the spatially stationary configurations.

Given either the GLM or the MI statistic, we aim to find the optimum activation configuration. This is equivalent to the MAP solution (Eq. (4.9)). However, a direct search for the optimal activation configuration is intractable because the search space is an exponential function of the number of voxels in a volume. Under a *binary* MRF assumption, Greig [17] showed that the search can be reduced to the Minimum-Cut-Maximum-Flow network problem, which can be solved in polynomial time via the Ford-Fulkerson algorithm. We refer to this exact solver as Min-Max throughout this thesis. Cosman [7] and Kim [39] showed how to apply Min-Max to the GLM and MI statistics. However, Min-Max is still computationally intensive when applied to the volumetric data: in our experiments, it took 1-3 hours, depending on the pairwise potential settings and the initial threshold applied to the GLM or MI statistic. On the other hand, the Mean Field approximation for MRFs is fast (ten to hundred times faster than Min-Max on the 3D grids we consider in this thesis) and reasonably accurate, as our results in the remainder of this section indicate. In the next section, we derive the Mean Field algorithm through a variation approximation approach.

### 4.3 A Variational Algorithm: Mean Field

A number of variational approaches have been successfully used for inference and estimation in large densely connected graphical probability models for which exact

probabilistic calculations are no longer feasible. The idea of variational algorithms is to approximate  $P_{\vec{X}|\vec{Z}}(\vec{X}|\vec{z})$  by a relatively simple probability mass function,  $Q(\vec{X})$ , through minimization of KL-Divergence between the two distributions:

$$D(Q(\vec{X})||P_{\vec{X}|\vec{Z}}(\vec{X}|\vec{z})) = \sum_{\vec{x}} Q(\vec{x}) \log(Q(\vec{x})) - \sum_{\vec{x}} Q(\vec{x}) \log(P_{\vec{X}|\vec{Z}}(\vec{x}|\vec{z})). \quad (4.13)$$

Since the extremum problem in Eq. (4.13) involves an unknown function,  $P_{\vec{X}|\vec{Z}}(\vec{X}|\vec{z})$ , the variational algorithms solve this problem by restricting the space of admissible functions  $Q(\vec{X})$ .  $Q(\vec{X})$  is a valid probability mass function for all possible activation configurations such that  $\sum_{\vec{x}} Q(\vec{x}) = 1$ . KL-Divergence measures how well  $Q(\vec{X})$  approximates  $P_{\vec{X}|\vec{Z}}(\vec{X}|\vec{z})$ . It is non-negative and is zero if and only if the variational distribution over the hidden variables equals to the true posterior probability,  $Q^*(\vec{X}) = P_{\vec{X}|\vec{Z}}(\vec{X}|\vec{z})$ . Using KL-divergence, we aim to find the best approximation to  $P_{\vec{X}|\vec{Z}}(\vec{x}|\vec{z})$ .

It is easy to see that the minimum of  $D(\cdot)$  is achieved for the same state configuration  $\vec{x}$  that minimizes the so called *free energy*,  $F_{MF}$ :

$$F_{MF} = D(Q(\vec{X})||P_{\vec{X}|\vec{Z}}(\vec{X}|\vec{z})) - \log(P_{\vec{Z}}(\vec{z})) \quad (4.14)$$

$$= \sum_{\vec{x}} Q(\vec{x}) \log(Q(\vec{x})) - \sum_{\vec{x}} Q(\vec{x}) \log(P_{\vec{X}|\vec{Z}}(\vec{x}|\vec{z})) - \log(P_{\vec{Z}}(\vec{z})) \quad (4.15)$$

$$= \sum_{\vec{x}} Q(\vec{x}) \log(Q(\vec{x})) - \sum_{\vec{x}} Q(\vec{x}) \log(P_{\vec{X},\vec{Z}}(\vec{x}, \vec{z})) + \sum_{\vec{x}} Q(\vec{x}) \log(P_{\vec{Z}}(\vec{z})) - \log(P_{\vec{Z}}(\vec{z})) \quad (4.16)$$

$$= \sum_{\vec{x}} Q(\vec{x}) \log(Q(\vec{x})) - \sum_{\vec{x}} Q(\vec{x}) \log(P_{\vec{X},\vec{Z}}(\vec{x}, \vec{z})) \quad (4.17)$$

$$= -H(Q) - E_Q[\log(P_{\vec{X},\vec{Z}}(\vec{x}, \vec{z}))] \quad (4.18)$$

$$= -H(Q) - E_Q[\log((P_{\vec{Z}|\vec{X}}(\vec{z}|\vec{x}))P_{\vec{X}}(\vec{x}))]. \quad (4.19)$$

The likelihood of observations given a hidden activation configuration,  $P_{\vec{Z}|\vec{X}}(\vec{z}|\vec{x})$ ,

is formulated according to the modelling assumptions. So far, we have not discussed how to select  $Q(\vec{x})$ . The first term in Eq. (4.19) suggests that  $Q(\vec{x})$  should describe as much variation as possible to achieve large entropy. In other words,  $Q(\vec{x})$  should be close to an uniform distribution. The second term implies we should choose  $Q(\vec{x})$  that maximize  $E_Q[\log(P_{\vec{Z}|\vec{X}}(\vec{z}|\vec{x})P_{\vec{X}}(\vec{x}))]$ . A variational algorithm usually restricts the space of  $Q(\vec{x})$  by selecting a certain parametric family of  $Q(\vec{x})$ . It then employs an iterative algorithm to minimize  $F_{MF}$  in terms of  $Q(\vec{x})$ 's parameters. The solution  $Q^*(\vec{x})$  may set  $F_{MF}$  to a local minimum rather than a global one. Therefore,  $Q^*(\vec{x})$  may not be the optimum choice among all  $Q(\vec{x})$  in our search space.

In the case of probability distributions, the appropriate simplification comes from properties of independence. The simplest family of variational distribution is one where all the variables,  $\{X_i : i = 1, \dots, N_V\}$ , are independent of each other. More precisely, the Mean Field approximation assumes

$$Q(\vec{X}) = \prod_i b_i(x_i) \quad (4.20)$$

where  $b_i(x_i)$  is a valid probability mass function so that  $\sum_{x_i=1}^M b_i(x_i) = 1$ .  $M$  is the number of possible states of  $X_i$ .

Substituting Eq. (4.20) and Eq. (4.8) into Eq. (4.19), we get

$$\begin{aligned} F_{MF}(b_i) &= \sum_{\{x_1, \dots, x_N\}} \prod_l b_l(x_l) \log(\prod_i b_i(x_i)) \\ &\quad - \sum_{\{x_1, \dots, x_N\}} \prod_l b_l(x_l) \log(\prod_{\langle i, j \rangle} \Psi_{ij}(x_i, x_j) \prod_k \Psi_k(x_k) P_{Z_k|X_k}(z_k|x_k)) \\ &\quad + \log(\lambda) \end{aligned} \quad (4.21)$$

$$\begin{aligned} &= - \sum_{\langle i, j \rangle} \sum_{x_i=1}^M \sum_{x_j=1}^M b_i(x_i) b_j(x_j) \log(\Psi_{ij}(x_i, x_j)) \\ &\quad + \sum_i \sum_{x_i=1}^M b_i(x_i) [\log(b_i(x_i)) - \log(\Psi_i(x_i) P_{Z_i|X_i}(z_i|x_i))] + \log(\lambda). \end{aligned} \quad (4.22)$$



We can rewrite Eq. (4.22) as the following:

$$\begin{aligned}
F_{MF}(b_i) = & - \sum_i \sum_{j \in \mathcal{N}(i)} \sum_{x_i=1}^M \sum_{x_j=1}^M b_i(x_i) b_j(x_j) \log(\Psi_{ij}(x_i, x_j)) + \\
& \sum_i \sum_{x_i=1}^M b_i(x_i) [\log(b_i(x_i)) - \log(P_{Z_i|X_i}(z_i|x_i) \Psi_i(x_i))] + \log(\lambda) \quad (4.23)
\end{aligned}$$

where  $\mathcal{N}(i)$  is the set of neighboring voxels for voxel  $i$ . The approximation is a solution of the following optimization problem:

$$\min F_{MF}(b_i) \quad (4.24)$$

$$\text{s.t. } \sum_{x_i=1}^M b_i(x_i) = 1 \quad \forall i. \quad (4.25)$$

It is not necessary to solve this constrained extremum problem using Lagrange multipliers because we can re-scale the solution to satisfy the constrain. Taking derivative of  $F_{MF}(b_i)$  with respect to  $b_i$ , we obtain

$$\begin{aligned}
\frac{\partial F_{MF}(b_i)}{\partial b_i} = & - \sum_{j \in \mathcal{N}(i)} \sum_{x_j=1}^M b_j(x_j) \log \Psi_{ij}(x_i, x_j) + \log b_i(x_i) + 1 \\
& - \log P_{Z_i|X_i}(z_i|x_i) \Psi_i(x_i). \quad (4.26)
\end{aligned}$$

Setting  $\frac{\partial F_{MF}(b_i)}{\partial b_i} = 0$  under the constrains  $\sum_{x_i=1}^M b_i(x_i) = 1 \quad \forall i$ , we obtain the following update rule for voxel  $i$  that can be used in an iterative procedure:

$$\boxed{b_i^{t+1}(x_i) \leftarrow \gamma P_{Z_i|X_i}(z_i|x_i) \Psi_i(x_i) e^{\sum_{j \in \mathcal{N}(i)} \sum_{x_j=1}^M b_j^t(x_j) \log \Psi_{ij}(x_i, x_j)}} \quad (4.27)$$

The normalization constant  $\gamma$ , ensures that the solution is a valid probability mass function.  $P_{Z_i|X_i}(z_i|x_i)$  is the likelihood of observing  $z_i$  given state  $x_i$ .  $\Psi_i(x_i)$  is the single node potential, and  $\Psi_{ij}(x_i, x_j)$  is the compatibility matrix.  $\mathcal{N}(i)$  is the set of voxel  $i$ 's neighbors. In each iteration of the Mean Field algorithm, the voxel's belief is updated according to the linear combination of its neighbors' beliefs in the previous iteration. The probability model (i.e.,  $P_{Z_i|X_i}$ ,  $\Psi_i$ , and  $\Psi_{ij}$ ) determines the exact form

of the update rule. Each voxel is assigned the state value with the highest belief at the end of the procedure (e.g., for binary MRFs, the voxel is set active if  $b_i(1) > b_i(0)$ ). When the iterative procedure converges, we can obtain an approximation to the MAP solution of the original MRF problem (Eq. 4.9).

$$\hat{\vec{x}}^* = \arg \max Q(\vec{x}) \quad (4.28)$$

The parameter setting for  $\Psi_i(x_i)$  and  $\Psi_{ij}(x_i, x_j)$  will be discussed in the next section. This iterated approximation algorithm is ten to one hundred times faster than the Min-Max solver on the 3D grids we consider in our experiments.

## 4.4 Estimating Model Parameters

The potential functions  $\Psi_i$ , and  $\Psi_{ij}$  and the observation likelihood  $P_{Z_i|X_i}$  must correspond to our notions of the appropriate bias toward desired solutions. In this work, we follow a common practice of setting the potential functions (same for all voxels) to the corresponding marginal probability distributions estimated from the data:  $\Psi_i(x_i)$  is set to the expected percentage of voxels in state  $x_i$ ,  $\Psi_{ij}(x_i, x_j)$  is set to the joint frequency of the states  $x_i$  and  $x_j$ , and  $P_{Z_i|X_i}$  is approximated by a smoother version of a class-conditional histogram. Other forms of potential functions have also been explored in different applications [9, 14, 15].

Lack of training data or ground truth necessary for estimating the marginal frequencies is a more serious problem. Unlike the segmentation application, where manual segmentations by experts can be used to construct priors on the frequencies and co-occurrences of tissue types, in most fMRI experiments even the experts cannot provide such information. Model parameters in the currently used detectors are either set using researcher’s intuition on the underlying activation properties (e.g., the threshold in GLM or the kernel width in Gaussian smoothing) or estimated from the input images (e.g., the noise variance in GLM). We take a similar approach of first running the GLM detector without smoothing and using the resulting SPM at

a user-chosen threshold to estimate the probability model. To study the sensitivity of the method to the parameter settings, we ran experiments where the values of the singleton potentials and the compatibility matrices varied substantially (0.1 to 2 times the estimated parameters). The variability in the detection accuracy (3-7%) was within the variability across different data sets as reported below.

## 4.5 Empirical Evaluation

In this section, we study the performance of the GLM detectors each augmented with Gaussian smoothing or MRF smoothing in synthetic fMRI data sets at different SNR levels. The evaluation consists of two parts: binary and trinary activation configurations. We start with the binary activation setting in order to evaluate the performance of the Mean Field approximation compared to Min-Max solver. Then we extended the evaluation to the three-state activation configuration. We would like to examine whether smoothing is a necessary procedure in fMRI detection in both activation settings. Besides a quantitative comparison presented in this section, we also generated activation maps using the detectors. To reduce redundancy, we will present the activation maps generated by these detectors along with their anatomically guided extensions, which are investigated in the next chapter, in Section 5.2.1. Validation of these detectors using real fMRI data is shown in Section 5.2.2.

### 4.5.1 Synthetic Data Sets with Binary Activation States

To quantitatively evaluate the performance of the detectors, we generated realistic phantom data by applying the EM segmentation [29] to an anatomical MRI scan and placing activation areas of variable size (average diameter of 15mm) randomly in the gray matter. We then downsampled the scan to an fMRI resolution volume. The gray matter voxels represent 7 – 10% of the total number of voxels in a volume, and the active voxels represent about 6 – 10% of the gray matter voxels in these images. We then created simulated fMRI scans based on the two-gamma hemodynamic response function, on an event-related experimental protocol, and on different levels of noise.

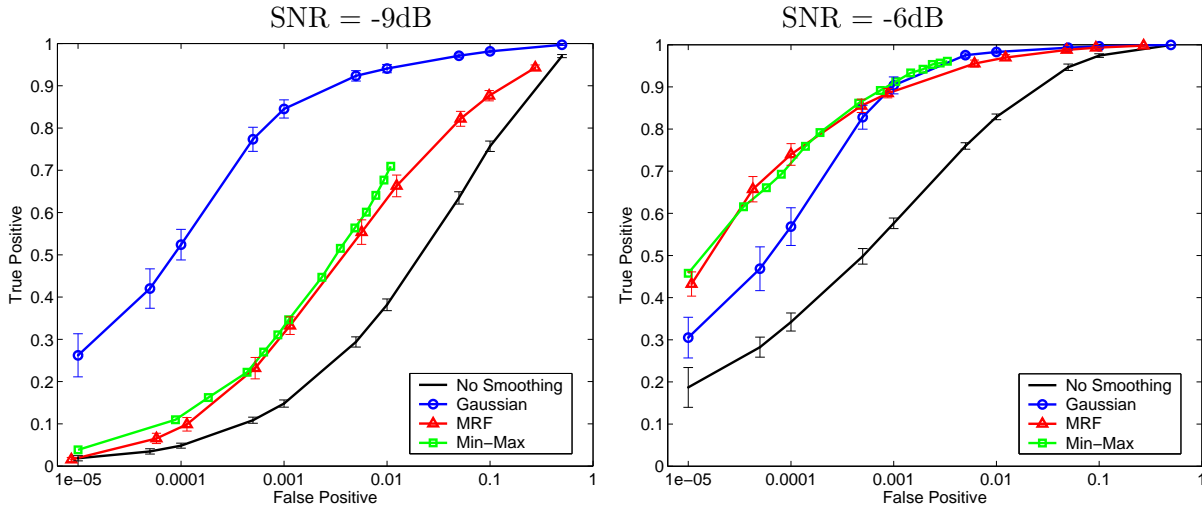


Figure 4-2: ROC curves for different smoothing techniques, at two noise levels. False positive rate is shown on the log scale.

The ground truth is a binary activation configuration in these data sets. We will discuss trinary activation configurations in the next section.

We used the estimated SNR,  $\widehat{\text{SNR}} = -10 \log_{10}(|B_1 \hat{\beta}_1|^2) / \hat{\sigma}^2$ , to determine a realistic level of the simulated noise as the true SNR is inaccessible for real fMRI scans. Since the signal and the noise overlap in some frequency bands, part of the noise energy is assigned to the estimated signal during detection. The estimated SNR is therefore an optimistic approximation of the true SNR, which can still be used as a monotonic upper bound of the true SNR. In our real fMRI studies, the estimated SNR is about -5dB. Here, we illustrate the results for two levels of true SNR, -6dB and -9dB, which correspond to estimated SNR of -4.3dB and -6.2dB respectively.

In all the experiments with binary activation data sets, we used the same GLM detector based on a 10-bin FIR function. To create a baseline for comparison, we first ran the GLM detector without any smoothing. For the Gaussian-smoothing-based detector, we first smoothed the signal spatially using a Gaussian filter, and then applied a GLM detector. To evaluate the Markov priors, we ran GLM coupled with the exact Min-Max solver and with the approximate Mean Field solver on the same raw images. Fig. 4-2 shows the ROC curves created for the four methods by varying the threshold applied to the GLM statistics. Due to the large number of voxels in the volume and the relatively small number of active voxels, only the very low false

positive rates are of interest (we focus on the false positive rates below 0.1%, which corresponds to about 10% of the total number of the active voxels, or approximately 250 voxels). The error bars indicate the standard deviation of the true detection rate over 15 different, independently created and processed, data sets. The Min-Max ROC curve does not have the error bars, as the estimation takes too long (1 to 3 hours for a single run). Moreover, the Min-Max ROC curve is incomplete because extreme threshold values cause it to run even longer (we stopped the runs after 3 hours).

The Mean Field detection accuracy is very close to the exact Min-Max solution, providing a reasonable approximation to the exact solution that also takes much less time to compute (most Mean Field runs finished in a few minutes). The Min-Max accuracy is sometimes lower than the Mean Field accuracy, which appears to contradict the optimality of the Min-Max solver. However, we note that both algorithms solve a particular estimation problem that does not necessarily describe the ground truth precisely but rather approximates it using a Markov model. Thus, the lowest energy state under this model might not be the best detector in practice. It is still reassuring to see that the approximate solver performs as well as the exact algorithm. It also suggests that even stronger spatial priors could improve detection accuracy.

As expected, the accuracy of all methods improves with increasing SNR. At high noise levels (low SNR), the Gaussian-smoothing-based detector outperforms the MRF-based detector. As the simplest smoothing technique, Gaussian smoothing is more robust to noise. We also believe that our current way of constructing the likelihood term in the MRF model over-emphasizes the data evidence over the prior. We are investigating ways to compensate for this in the estimation of the model. As the SNR increases, MRFs provide better regularization of the activation state (for example, at SNR=-6dB, at the false positive rate of 0.01%, the MRF-based detector outperforms the Gaussian-smoothing-based detector by about 15% in true detection accuracy; at 70% true detection, the MRF-based detector approximately halves the false detections compared to the Gaussian-smoothing-based detector). With the improving scanning technology, we believe MRFs will become even more helpful in reducing spurious false detection islands.

## 4.5.2 Synthetic Data Sets with Trinary Activation States

According to neuroscience literature [3, 35, 24], some regions in the brain show negative activations when a subject performs certain tasks. For example, Schwartz [34] found that the pre-frontal cortex regions usually show negative activation during rapid eye movement sleep. Negative activation is caused by decrease in glucose metabolism, with a detailed interpretation of this phenomenon still an active research area.

To examine the performance of our detectors in a more realistic setting, we generated synthetic data sets with three activation states. The data generating process is very similar to the one explained in Section 4.5.1. The activation areas were placed in the gray matter with average diameter 15mm. While all of the activation areas were discussed are positively active in the previous section, we randomly assigned the activation regions being either positively or negatively active in this experiment. In other words, all voxels in an activation region have an identical activation state; voxel belonging to different activation regions, though they are very close, may be in different activation states. Approximately, 7% and 3% of the gray matter voxels are positively and negatively active voxels respectively. We then created simulated fMRI scans based on the same parametric hemodynamic response function, event-related protocol, and noise levels as those in the binary simulated fMRI scans.

In the experiments with trinary activation states, we use a slightly modified GLM detector based on the same 10-bin FIR function. The F statistic in this GLM model is distributed as  $F_{10, N_T-10}$ . Since F statistic is one sided, it cannot distinguish between positive activation and negative activations. In the modified GLM detector, a voxel is labelled as a positively active voxel if its positive regression coefficients have a higher sum of energy than the sum of energy among the negative regressor coefficients:  $\sum_j (\beta_{1_i})_j^2 > \sum_k (\beta_{1_i})_k^2$ , where  $(\beta_{1_i})_j > 0$ ,  $(\beta_{1_i})_k < 0$ , and  $\vec{\beta}_{1_i} = [(\beta_{1_i})_1, (\beta_{1_i})_2, \dots, (\beta_{1_i})_{N_{\beta_1}}]$ . We augmented the negatively active voxel with a minus sign to its corresponding maximum log likelihood ratio derived in Eq. (4.1). Therefore, the statistics of the true negative activation voxels are usually below zero, and the statistics of the true positive activation voxels are usually above zero. The

non-active voxels' statistics usually have small magnitudes.

The estimation procedure using the GLM detector, the Gaussian-smoothing-based detector, and the MRF-based detector is the same as that discussed in Section 4.5.1, except that we cannot solve the MRF problem using the Min-Max solver because of the trinary activation configuration. Hence, we only report the results from the MRF-based detector solved by the Mean Field algorithm, and the result is the MAP solution of the marginal probability returned from Mean Field:  $\hat{x}^* = \arg \max_{\vec{x}} Q(\vec{x})$ .  $\hat{x}^*$  is the approximate MAP solution for the original MRF problem.

We adjust the definitions of false positive and true positive rate for the trinary activation state case. False positive rate is defined as the percentage of non-active voxels being classified as either positively or negatively active voxels. We define true positive rate as the percentage of the active voxels (including positively and negatively active voxels) that are classified correctly. As we can see, the misclassifications between the two activation types are ignored. In addition to the ROC analysis, we also report the results in a form of a confusion matrix.

Fig. 4-3 shows the ROC curves for the GLM detector, the Gaussian-smoothing-based detector, and the MRF-based detector by varying the threshold applied to the GLM statistics. Similar as Section 4.5.1, the error bars are computed over ten data sets, independently created and processed. Since we do not include misclassifications between the positively and the negatively active voxels, the true positive rate cannot reach 100% if there are any misclassifications between the two activation types.

Since the difference in the statistic value is larger between voxels in the positive and negative activation states than the one between the positive and the non-active states or between the negative and the non-active states, there are few negatively active voxels falsely labelled as positively active voxels or vice versa. Misclassifications between the positively active and non-active voxels or between the negatively active and non-active voxels are more common.

Tables 4.1 shows the average confusion matrices of the three detectors over the same ten data sets when false positive rate is fixed at 0.05%. Confusion matrices list the percentage of voxels in state  $S_i$  that are classified as state  $S_j$ . For instance, the

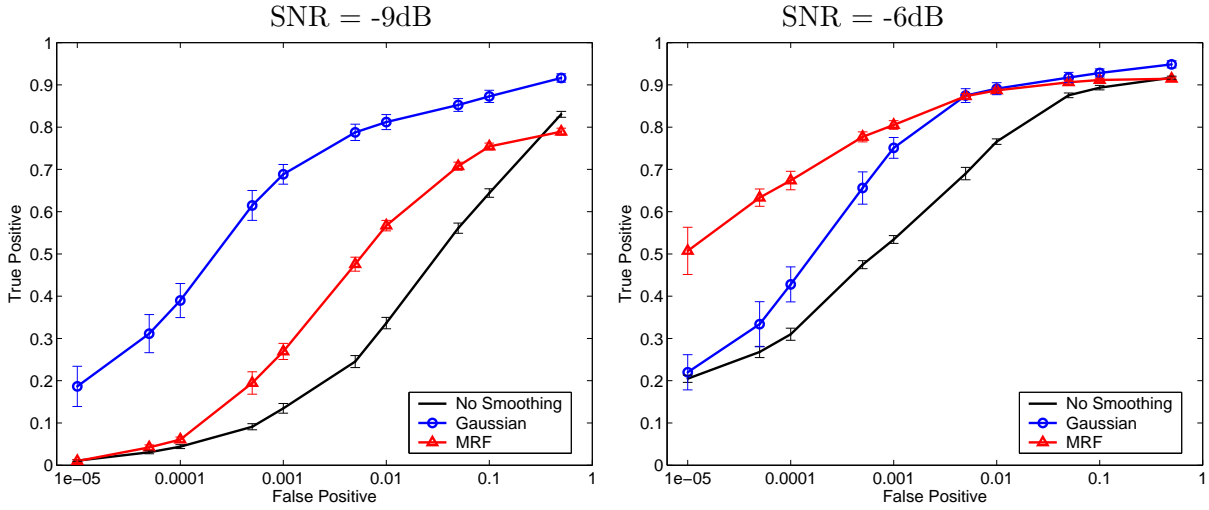


Figure 4-3: ROC curves for different smoothing techniques, at two noise levels. False positive rate is shown on the log scale. It is impossible to achieve 100% true detection rate, since false detections between positively and negatively active states are ignored.

(a) GLM detector, SNR = -9 dB

Ground Truth	Detection		
	Negative Active	Not Active	Positive Active
Negative Active	9.03	89.73	1.24
Not Active	0.03	99.9	0.02
Positive Active	0.88	89.51	9.61

(b) GLM detector, SNR = -6 dB

Ground Truth	Detection		
	Negative Active	Not Active	Positive Active
Negative Active	46.6	50.02	3.34
Not Active	0.03	99.9	0.02
Positive Active	3.53	49.60	46.87

(c) Gaussian-smoothing-based detector, SNR = -9 dB

Ground Truth	Detection		
	Negative Active	Not Active	Positive Active
Negative Active	54.59	44.43	0.98
Not Active	0.02	99.95	0.03
Positive Active	1.14	32.03	66.83

(d) Gaussian-smoothing-based detector, SNR = -6 dB

Ground Truth	Detection		
	Negative Active	Not Active	Positive Active
Negative Active	57.45	42.35	0.20
Not Active	0.02	99.95	0.03
Positive Active	0.14	28.23	71.63

(e) MRF-based detector, SNR = -9 dB

Ground Truth	Detection		
	Negative Active	Not Active	Positive Active
Negative Active	15.46	82.74	1.80
Not Active	0.03	99.94	0.03
Positive Active	1.01	79.44	19.55

(f) MRF-based detector, SNR = -6 dB

Ground Truth	Detection		
	Negative Active	Not Active	Positive Active
Negative Active	74.85	20.69	4.45
Not Active	0.02	99.95	0.02
Positive Active	4.45	16.28	79.26

Table 4.1: Detection performance of three selected detectors at two SNR levels.



first and the second entries in Table 4.1(a) mean that 9% of the true negatively active voxels are detected correctly and that 89.73% of the true negatively active voxels are detected as not active voxels when we control the false positive rate at 0.05%. The standard deviations of all detectors are 1-5%.

Comparing the ROC curves (Fig. 4-3) or the confusion matrices (Table. 4.1) with the ROCs generated in the binary activation data sets at SNR= -6dB (Fig. 4-2), the MRF-based detector outperforms the Gaussian-smoothing-based detector by a larger margin in the trinary activation data set. In the binary case, the detection difference is approximately 10% between the MRF-based detector and the Gaussian-smoothing-based detector; in the trinary activation case, there is approximately 20% (in both negatively and positively active states) between these two detectors. The Gaussian-smoothing-based detector outperforms the MRF-based detector by a smaller margin in the low SNR setting. The Gaussian-smoothing-based detector performs worse in the trinary case than in the binary case, because spatial averaging suppresses the signals of adjacent voxels that are in positive and negative activation states. As a result, both of the two voxels are classified as non-active voxels. In addition, due to the fact that there are twice as many positively active voxels as negatively active voxels in our synthetic data, the Gaussian-smoothing-based detector shows worse detection performance for the negatively active voxels. Another interesting observation is that the Gaussian-smoothing-based detector has similar performance in both high and low SNR data sets. We again see the robustness of the Gaussian smoothing method.

## 4.6 Summary

We investigated improvements in fMRI detection provided by Gaussian smoothing and Markov priors spatial modelling, and our experiments confirmed the importance of spatial regularization in reducing fragmentation of the activation maps. We evaluated the methods on phantom data by performing the ROC analysis and the confusion matrix analysis. In the high-noise setting, the Gaussian-smoothing-based detector outperformed other methods; as the SNR in the images increased, the Markov pri-

ors offered a substantial improvement in detection accuracy. At high SNR levels, there is a larger margin, in terms of true detection, between the MRF-based detector and the Gaussian-smoothing-based detector in the trinary activation configuration than that in the binary activation configuration. This difference occurs because the Gaussian-smoothing-based detector suppresses the signal when it averages positively active signals with negatively active signals. We will further validate the detectors using real fMRI data sets in the next chapter.

In this chapter, we introduced a fast approximate solver, Mean Field, in application to MRF-based fMRI detection and showed that it provides reasonably accurate approximations to the exact solution while taking substantially less time to evaluate. We note that since the Markov model itself is an approximation of the real geometry of the activation regions, we should not dwell on the small differences in the activation maps introduced by the approximate solvers but rather focus on their performance relative to the ground truth. An MRF provides a spatial prior that refines the structure of the resulting activation map over the Gaussian smoothing, as demonstrated by our experiments on the phantom data sets. The general nature of the Mean Field algorithm allows straightforward extension of the probabilistic model to include the tissue type for each voxel. We will present this refined detector in the next chapter.

# Chapter 5

## Spatial Regularization with Anatomical Structure

Chapter 4 presented MRFs for modelling spatial coherency of the activation maps. We adopted the Mean Field algorithm to approximate the MAP solution for the MRF problem. We also compared our detectors with the Gaussian-smoothing-based detector on binary and trinary activation configurations, as well as at different SNR levels.

We further refine the activation priors by incorporating anatomical information. Similarly to the atlas-based segmentation, where a probabilistic atlas serves as a spatially varying prior on the tissue types, the anatomical information can provide a prior on the activation map. Intuitively speaking, we want the prior to reflect the fact that activation is much more likely to occur in gray matter than in white matter, and not at all in cerebrospinal fluid (CSF) or bone. In addition, the spatial coherency of activation is strong within each tissue and not across tissue boundaries. To incorporate the anatomical information, we augment the MRF model introduced in Chapter 4 with tissue type information. In this refined model, each hidden node encodes both the tissue type and the activation state. The segmentation from the anatomical scans provides an additional, potentially noisy, observation at each node. We derive the detection algorithm for this model and evaluate it on simulated and real data. The anatomically guided MRF-based detector is able to achieve high

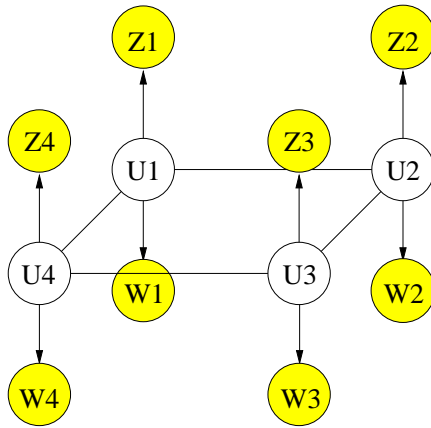


Figure 5-1: *Graphical model for  $P_{\vec{U}, \vec{Z}, \vec{W}}$ .*

detection accuracy with significantly shorter time courses compared to the standard GLM detector.

Anatomical scans have certainly been used in fMRI analysis and visualization before. Hartvig [18] used the anatomical information in his marked point process spatial prior. In many systems (e.g., BrainVoyager [1]), the subject’s anatomical image is transformed into a standard coordinate frame (such as Talairach) and the functional activation map is displayed on the surface that corresponds to the cortical sheet in that coordinate frame. Other systems (e.g., FSL [2]) rely on sophisticated segmentation algorithms to extract a topologically correct representation of the cortical surface from the anatomical scan [8]. Performing Gaussian smoothing on the surface eliminates irrelevant voxels from the weighted average for the cortical locations. In contrast, our approach does not require a surface extraction algorithm, but instead utilizes anatomical information to inject the anatomically based coherency bias into the detection algorithm while performing the computation directly on the volumetric data. The inspiration for this work comes from the success enjoyed by MRFs in providing spatial smoothing priors for image segmentation [21, 23, 30, 42].

## 5.1 Anatomical Priors for Spatial Regularization

The general nature of the Mean Field algorithm allows straightforward extension of the MRF model to include the tissue type for each voxel. We define  $\vec{V} = [V_1, \dots, V_{N_V}]$  to be the tissue types of all voxels, and  $\vec{W} = [W_1, \dots, W_{N_V}]$  the tissue type observations,

such as a result of an automatic segmentation procedure.  $W_i$ 's are noisy observations due to imperfect registration between the fMRI image and the anatomical scan, the mismatch in their resolutions, and the noise in the segmentation itself. Now each voxel has two hidden attributes: the activation state  $X_i$  and the tissue type  $V_i$ . We combine these attributes into a single hidden node  $U_i$ , as illustrated in Fig. 5-1. For example, for binary activation states (active or not active) and three tissue types (gray matter, white matter, or other) case,  $U_i$  has six possible states. In our experiments with synthetic and real data sets, we combined binary or trinary activation states with three tissue types.

The MRF prior on  $\vec{U}$  is similar to the one introduced in Eq. (4.8):

$$P_{\vec{U}}(\vec{u}) = \frac{1}{\lambda} \prod_{\langle i,j \rangle} \Psi_{ij}(u_i, u_j) \prod_i \Psi_i(u_i) \quad (5.1)$$

where,  $\Psi_{ij}(u_i, u_j)$  and  $\Psi_i(u_i)$  are pairwise and singleton potentials. Similarly to the derivations in the previous section, the MAP estimate in this case is as follows:

$$\vec{u}^* = \arg \max_{\vec{u}} P_{\vec{U}|\vec{Z},\vec{W}}(\vec{u}|\vec{z}, \vec{w}) \quad (5.2)$$

$$= \arg \max_{\vec{u}} P_{\vec{U}}(\vec{u}) P_{\vec{Z}|\vec{U}}(\vec{z}|\vec{u}) P_{\vec{W}|\vec{U}}(\vec{w}|\vec{u}) \quad (5.3)$$

$$= \arg \max_{\vec{u}} \frac{1}{\lambda} \prod_{\langle i,j \rangle} \Psi_{ij}(u_i, u_j) \prod_i \Psi_i(u_i) P_{Z_i|U_i}(z_i|u_i) P_{W_i|U_i}(w_i|u_i) \quad (5.4)$$

where,  $P_{Z_i|U_i}(z_i|u_i)$  is the likelihood of obtaining statistic  $z_i$  given that voxel  $i$  is at state  $u_i$ , and  $P_{W_i|U_i}(w_i|u_i)$  is the likelihood of obtaining segmentation observation  $w_i$  given that voxel  $i$  is at state  $u_i$ . We assume that the segmentation observation ( $\vec{W}$ ) and the fMRI observation ( $\vec{Z}$ ) are conditionally independent given the state of the voxel since they are obtained from two different images.

We again use the Mean Field algorithm to approximate  $P_{\vec{U}|\vec{Z},\vec{W}}(\vec{u}|\vec{z}, \vec{w})$  by selecting a distribution  $Q^*(\vec{u})$ , among the family of voxel-independent distribution,  $Q(\vec{u}) = \prod_i b_i(u_i)$ , in terms of minimizing the KL-divergence.

It is straightforward to derive the iterative update rule following the procedure

presented from Eq. (4.21) to Eq. (4.27), resulting the following update rule:

$$\boxed{b_i^{t+1}(u_i) \leftarrow \gamma P_{W_i|U_i}(w_i|u_i) P_{Z_i|U_i}(z_i|u_i) \Psi_i(u_i) e^{\sum_{j \in \mathcal{N}(i)} \sum_{u_j=1}^M b_j^t(u_j) \log \Psi_{ij}(u_i, u_j)}} \quad (5.5)$$

This update rule is similar to Eq. (4.27), with the exception of the extra likelihood term  $P_{W_i|U_i}(w_i|u_i)$  for the tissue type observation. The compatibility matrix  $\Psi_{ij}(u_i, u_j)$  is  $M \times M$ , where  $M$  is the number of states in  $U_i$ .  $\Psi_{ij}(u_i, u_j)$  is a larger compatibility matrix compared with  $\Psi_{ij}(x_i, x_j)$  (Eq. 4.27), because it includes tissue type information. We follow the same parameter estimation procedure as described in Section 4.

Since we are interested in the MAP estimate of the activation state, we first marginalize  $Q(\vec{u}) = Q(\{\vec{x}, \vec{v}\})$  with respect to the tissue type  $\vec{v}$ , and then obtain the approximation MAP solution.

$$\hat{\vec{x}}^* = \arg \max_{\vec{v}} \sum_{\vec{v}} Q(\{\vec{x}, \vec{v}\}) \quad (5.6)$$

## 5.2 Empirical Evaluation

In this section, we use both synthetic and real data sets to study the detection improvement introduced by incorporating anatomical information. We compare different ways of incorporating anatomical information into the GLM detectors<sup>1</sup>: the anatomically guided GLM detector, the anatomically guided Gaussian-smoothing-based detector, and the anatomically guided MRF-based detector that was introduced in Section 5.1. The anatomically guided GLM detector suppresses activations outside of the gray matter using segmentation as a map (“soft” masking could also account for mis-registration and errors in segmentation). To incorporate anatomical information into the Gaussian filter, we adjust the weights of the filter based on the tissue types of the voxel’s neighbors: while evaluating the filter at voxel  $i$ , we assign higher weights to the neighbors sharing the same segmentation results as voxel  $i$ .

---

<sup>1</sup>It is straightforward to augment the MI detector with anatomical information, but we focus on the GLM detector in this thesis.

## 5.2.1 Synthetic Data Sets

In this section, we use synthetic data to compare the performance of the anatomically guided detectors by performing the ROC analysis and the confusion matrix analysis.

### Synthetic Data Sets with Binary Activation States

We use the same phantom data sets described in Section 4.5.1 to evaluate the performance of the anatomically guided detectors. Fig. 5-2 illustrates the ROC analysis for the three regularization methods investigated in Section 4.5.1 (solid lines) and their anatomically based variants (dashed lines). We omit the Min-Max solver for the MRF model, as it cannot handle multi-valued states.

In addition to the trends observed before, we note that the anatomical information significantly boosts the performance of all detectors at all noise levels. At high noise levels (SNR = -9dB) and false positive rates between 0.01% and 0.1%, all methods gain at least 10% in true detection rate when using anatomical information. The MRF model benefits more than the Gaussian-smoothing model, but its detection accuracy is still lower. At a lower noise level (SNR = -6dB), the anatomically guided GLM detector approaches the performance of the Gaussian-smoothing-detector. At 0.01% false positive rate, the anatomically guided MRF-based detector outperforms the anatomically guided Gaussian-smoothing-based detector by about 15% in true detection rate, achieving over 90% detection accuracy. The large boost experienced by the basic GLM detector when augmented with anatomical information is easy to understand: since false detections occur relatively uniformly throughout the volume, masking the gray matter improves the performance substantially. It is important to note that applying the gray-matter mask ignores the false detections inside the gray matter.

Besides the quantitative analysis presented above, we find it useful to visually inspect the resulting activation maps. Fig. 5-4 at the end of this section illustrates the detection results by showing one axial slice of the estimated activation map. The top image shows the phantom activation areas that were placed in the volume and

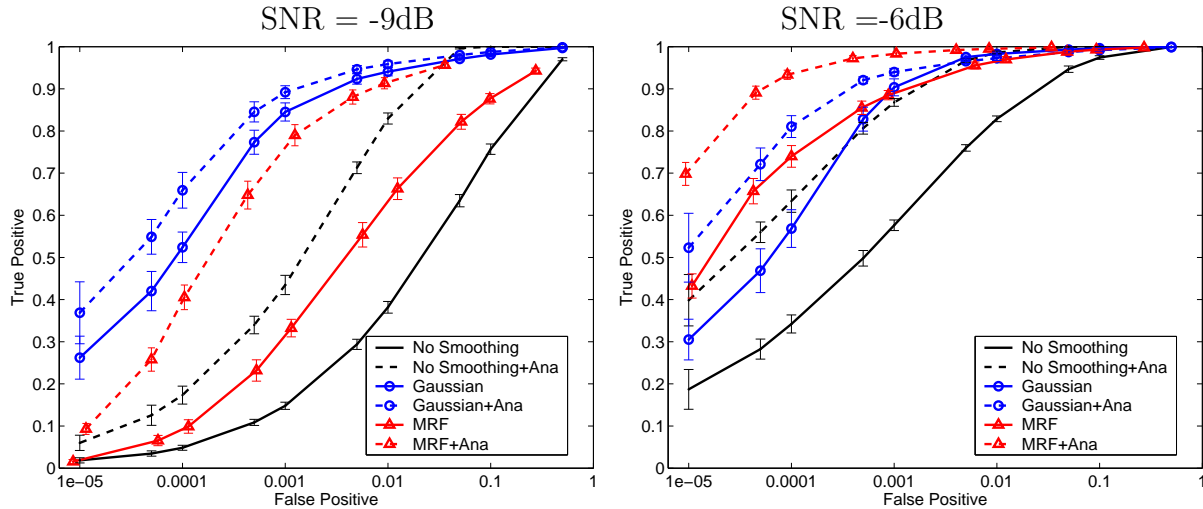


Figure 5-2: *ROC curves for different smoothing techniques augmented with the anatomical information, at two noise levels. False positive rate is shown on the log scale.*

used to generate the simulated fMRI scan. The bottom four rows show the same slice in the reconstructed volume at two different noise levels. All the reconstructions were performed at 0.05% false positive rate. In other words, each image in Fig. 5-4 shows one slice in the reconstructed volume that corresponds to a point on the ROC curve of the respective detector at 0.05% false positive rate.

The basic GLM detector produces a fragmented activation map that contains a number of false detection islands at high SNR (Fig. 5-4g) and shows very little of the original activation at low SNR (Fig. 5-4b). Given either of these maps, the users would have troubles inferring the true activation areas and disambiguating them from spurious false detections. The Gaussian-smoothing-based detector (Fig. 5-4c,j) leads to a reasonable estimate of the ground truth, and its result largely tends to be “spherical”, which may change the shape of the detected activations. The smoothing effectively overestimates the extent of the regions. Consequently, many false positive voxels in the Gaussian-smoothing-based detector’s activation map occur at the boundaries of the activation regions. Injecting anatomical information reduces this over-smoothing effect for some of the areas. At low SNR (-9dB), the MRF model (Fig. 5-4d) fills in many of the active pixels that were missed by the GLM detector, but as we saw before, its results are not as good as that produced by the Gaussian-smoothing-based



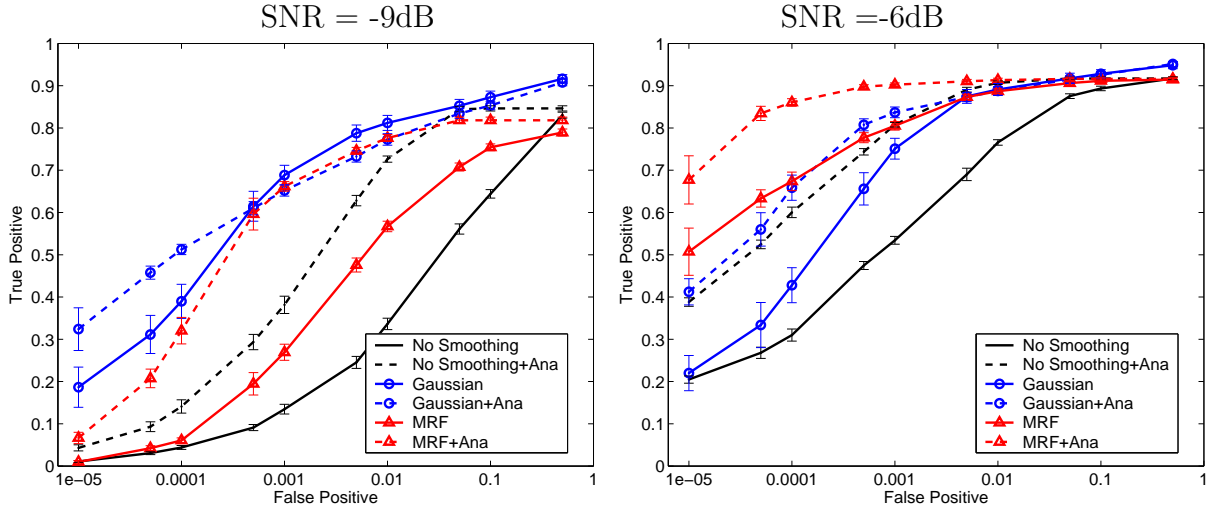


Figure 5-3: *ROC curves for different smoothing techniques augmented with the anatomical information, at two noise levels. False positive rate is shown on the log scale. It is impossible to achieve 100% true detection rate, since false detections between positively and negatively active states are ignored.*

detector. At higher SNR (-6dB), MRF produces a relatively accurate result (Fig. 5-4j). Not all of the scatter activation islands are removed through regularization, but the activation map looks more similar to the ground truth. The activation map is further improved when anatomical information is incorporated into the model.

### Synthetic Data Sets with Three Activation States

To examine the performance of the GLM detector, the Gaussian-smoothing-based detector, and the MRF-based detector with their extended anatomically guided versions in a trinary activation configuration, we use the data sets generated in Section 4.5.2.

Fig. 5-3 is the average ROC statistics of the six detectors over ten data sets at two different noise levels. Tables 5.1 and 5.2 show the corresponding average confusion matrices when false positive rate is fixed at 0.05%. Similar as the rest of the experiment in the synthetic data sets, we used the same GLM detector based on the 10-bin FIR function. The 10-bin FIR GLM detector introduces difficulty of differentiating the positive activations from the negative activations. We adopt the convention mentioned in Section 4.5.2: a positively active voxel is a voxel whose positive regression coefficients has a higher sum of energy than sum of energy among its negative re-

(a) No Smoothing

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	9.03	89.73	1.24
Not Active	0.03	99.95	0.02
Positive Active	0.88	89.51	9.61

(b) No Smoothing+Ana

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	29.16	66.43	4.41
Not Active	0.03	99.95	0.02
Positive Active	4.13	66.34	29.54

(a) Gaussian

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	54.59	44.43	0.98
Not Active	0.02	99.95	0.03
Positive Active	1.14	32.03	66.83

(b) Gaussian+Ana

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	55.15	42.73	2.12
Not Active	0.02	99.95	0.03
Positive Active	2.04	32.51	65.45

(a) MRF

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	15.46	82.74	1.80
Not Active	0.03	99.94	0.03
Positive Active	1.01	79.44	19.55

(b) MRF+Ana

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	56.34	36.69	6.69
Not Active	0.03	99.94	0.03
Positive Active	6.18	32.04	61.78

Table 5.1: *Detection performance of three selected detectors and their corresponding anatomically guided extensions at  $-9$  dB SNR.*

gression coefficients. We reuse the definitions of the false positive and true positive rates defined in the last chapter. False positive rate is defined as the percentage of non-active voxels being classified as either positively or negatively active voxels; true positive rate is defined as the percentage of the active voxels (including positively and negatively active voxels) being correctly classified. Because of our definitions, the true detection rate cannot reach 100% in the ROC curves, because misclassifications between positive and negative activations are not counted.

Since we are interested in examining the detection improvement provided by incorporating the additional spatial structure into the detectors in this chapter, we display the confusion matrices of the detectors and their corresponding anatomically guided versions side-by-side.

Fig. 5-3, Table. 5.1, and Table. 5.2 illustrate that incorporating the anatomical structure into the detectors improves the detection accuracy over all detectors except the Gaussian-smoothing-based detector in the high noise setting. Comparing the de-

(a) No Smoothing

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	46.6	50.02	3.34
Not Active	0.03	99.95	0.02
Positive Active	3.53	49.60	46.87

(b) No Smoothing+Ana

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	74.09	19.67	6.24
Not Active	0.03	99.95	0.02
Positive Active	6.46	18.81	74.72

(c) Gaussian

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	57.45	42.35	0.20
Not Active	0.02	99.95	0.03
Positive Active	0.14	28.23	71.63

(d) Gaussian+Ana

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	74.56	24.21	1.24
Not Active	0.02	99.95	0.03
Positive Active	1.04	13.97	84.99

(a) MRF

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	74.85	20.69	4.45
Not Active	0.02	99.95	0.02
Positive Active	4.45	16.28	79.26

(b) MRF+Ana

<i>Ground Truth</i>	<i>Detection</i>		
	Negative Active	Not Active	Positive Active
Negative Active	88.97	3.81	7.23
Not Active	0.03	99.95	0.02
Positive Active	7.33	23.5	90.32

Table 5.2: *Detection performance of three selected detectors and their corresponding anatomically guided extensions at  $-6$  dB SNR.*

tection accuracy between negative and positive activation over all confusion matrices, we observe that the Gaussian-smoothing-based detectors (with and without anatomical information) show the highest favor for positive activation due to the fact that there is at least twice as many positively active voxels as the negatively active voxels in our synthetic data sets. Therefore, we should not spatially smooth the signals using a Gaussian filter if we are interested in locating the negative activation areas. On the other hand, Gaussian smoothing virtually considers non-activation as the middle state between the positive activation and negative activation states. As a result, the false detection between positively active and non-active voxels or between negatively active and non-active voxels is high, but the false detection between positively active and negatively active voxels is low.

In the high noise environment (SNR=  $-9$ dB), there is no significant improvement introduced by the anatomical information in the Gaussian-smoothing-based detector. Even though the anatomical structure enables different weights, it cannot prevent

equal-weighted averaging of the signals of different activation types because both positive and negative activation areas lie in the gray matter. In Section 5.2.1, we concluded that the Gaussian-smoothing-based detectors (with and without anatomical information) has a better performance compared with the MRF-based detectors (with and without anatomical information) in the binary activation configuration and at a low SNR level. However, the anatomically guided MRF-based detector provides similar detection accuracy compared with the anatomically guided Gaussian-smoothing-based detector in the trinary activation configuration. It is again caused by the Gaussian smoothing technique’s over-smoothing effect over different activation types.

In the low noise setting (SNR=−6dB), anatomical information boosts at least 15% in terms of true detection for the detectors. The anatomically guided MRF-based detector achieves close to 90% detection accuracy for both negative and positive activations. Comparing Table 5.2(b) with Table 5.2(d), the anatomically guided GLM detector and the anatomically guided Gaussian-smoothing-based detector show similar detection accuracy for negative activation. It again shows the undesired performance of the Gaussian-smoothing-based detector for detecting negative activations.

Similarly to the last section, we present one axial slice of the resulting activation maps, generated by each detector, for visual inspection (Fig. 5-5). The top image shows the phantom activation areas that were placed in the volume and used to generate the simulated fMRI scans. The bottom four rows show the same slice in the reconstructed volume at two different noise levels. All the reconstructions were performed at 0.05% false positive rate. We use yellow and blue to represent correctly detected positively active voxels and correctly detected negatively active voxels, respectively. Red represents the voxels that are either falsely classified as active voxels, including positive and negative activations (corresponding to the statistics in the first and third columns of the second row in the tables), or falsely classified between positively and negatively active voxels (corresponding to the statistics in the third column of the first row and the first column of the third row in the tables).

The basic GLM produces a fragmented activation map that contains a number

of false detection islands at high SNR (Fig. 5-5h) and shows very little of the original activation at low SNR (Fig. 5-5b). In contrast to the results obtained in Section 5.2.1, the Gaussian-smoothing-based detector (Fig. 5-5c,j) cannot produce a reasonable estimate of the ground truth at either SNR level because it washes out signals of different activation types during averaging. The Gaussian-smoothing-based detector works better for the activation areas whose shapes are spherical or square. Incorporating anatomical information cannot improve the Gaussian smoothing result much at the high noise setting. At low SNR (-9dB), the MRF model fills in many of the active pixels that were missed by the GLM detector, but it does not produce as good result as Gaussian smoothing. At higher SNR (-6dB), MRF (Fig. 5-5j) produces a relatively accurate result. It is even comparable to the result provided by the anatomically guided Gaussian-smoothing-based detector. When the anatomical information is incorporated into the MRF model, the activation map looks more similar to the ground truth.

## 5.2.2 Real fMRI Experiments

In the real fMRI experiments, since the ground truth is unavailable, constructing ROC analysis or confusion matrix analysis is not possible. Instead, we visually compare the resulting activation maps. In addition, we compare different detectors by evaluating their performance on reduced-length time courses in both binary and trinary activation configurations. This approach effectively evaluates the ability of each detector to reconstruct the true activation areas with less evidence on the strength of the signal.

We used the same fMRI data sets described in Section 3.2 to examine the GLM detectors augmented with different spatial priors. The fMRI scans were obtained during an auditory “two-back” word block design experiment for studying working memory [40]. The estimated SNR when averaging over all voxels in the brain was -4.7dB (-2.3dB when averaging voxels in selected ROIs relevant to the task).

Fig. 5-6 and 5-7 depict the resulting activation maps for the same axial slice in our fMRI data sets, corresponding to the binary and trinary activation assumptions,

respectively. Both of them are thresholded SPMs, and voxels with 0.8% highest statistics are labelled as active voxels. While we employed the 10-bin FIR GLM detector in the experiments on the synthetic data sets, we used GLM with the two-gamma hemodynamic response function in the real data sets. We chose this strong regression model for the real fMRI data to overcome many variations in the hemodynamic response in the brain. The  $B_1$  matrix (Eq. (3.21) or (4.1)) is the convolution of the experimental protocol and the two-gamma function.  $B_1$  is a column vector, and  $\beta_1$  is a scalar in this case. Similarly to the experiment on the synthetic data sets, we used the maximum log likelihood as the activation statistic  $z_i$ . In the trinary activation case, and we used the maximum log likelihood, augmented with the sign of  $\beta_1$ , as the activation statistic.

Fig. 5-6a and 5-7a show the reconstructed activation maps using GLM without any spatial smoothing on the full-length fMRI signals (all 9 epochs). The ground truth for this scan is unknown, but we can use these maps as visual references when evaluating the performance of the detectors on the time course of reduced length. Fig. 5-6b-d and 5-7b-d are the activation maps obtained by applying the GLM detector, the Gaussian-smoothing-based detector, and the MRF-based detector to the first 5 epochs of the time course. Fig. 5-6e-g and 5-7e-g are the extended anatomically guided versions of Fig. 5-6b-d and 5-7b-d, respectively.

The activation maps that are constructed by applying the GLM detector to the first five epochs of the time course (Fig. 5-6b and 5-7b) are more fragmented due to loss in SNR from reducing the length of the signal. Although Gaussian smoothing removes most of the single voxel activation islands, its activation map (Fig. 5-6c and 5-7c) is an overestimate compared with Fig. 5-6a and 5-7a. Injecting anatomical information slightly reduces the overestimate in the Gaussian smoothing. In the trinary case, the Gaussian-smoothing-based detector (with and without anatomical information) underestimates negative activation due to averaging signals of different activation types and a relatively small amount of weak negative activation areas.

MRF regularization (Fig. 5-6e-f) yields reasonable reconstruction results that are close to the activation map estimated from the full-length signal, but do not look

overly smoothed. This highlights the potential benefit of using the Markov priors in fMRI detection. Similarly to what we have observed in the experiments using the synthetic data sets, the MRF-based detectors (Fig. 5-7e-f) show a better detection accuracy in detecting the negative activation. The MRF model also benefits from using anatomical information to remove spurious activations.

### 5.3 Summary

In Chapter 4, our experiments demonstrated that spatial regularization significantly reduces fragmentation of the activation maps. The insight of this chapter is that we can use anatomical information to bias the fMRI detectors. The Gaussian-smoothing-based detector can be straightforwardly augmented with the anatomical prior by rescaling the coefficients of the smoothing kernel. Moreover, we derived an algorithm for the anatomically guided MRF-based detector. We used synthetic data to validate the detectors' performance by presenting the ROC analysis and the confusion matrices analysis. We also evaluated them by studying their ability to recover activation from significantly shorter time courses using real fMRI data. Both binary and trinary activation configurations were employed in the evaluations. The anatomically guided detectors enable us to shorten fMRI scan length by half while retaining detection power comparable with full-length fMRI scans. We found that Gaussian-smoothing-based detectors (with and without anatomical structure) are not suitable for analysis of trinary activation states. Among all the anatomically guided detectors, the anatomically guided MRF-detector showed the best performance.

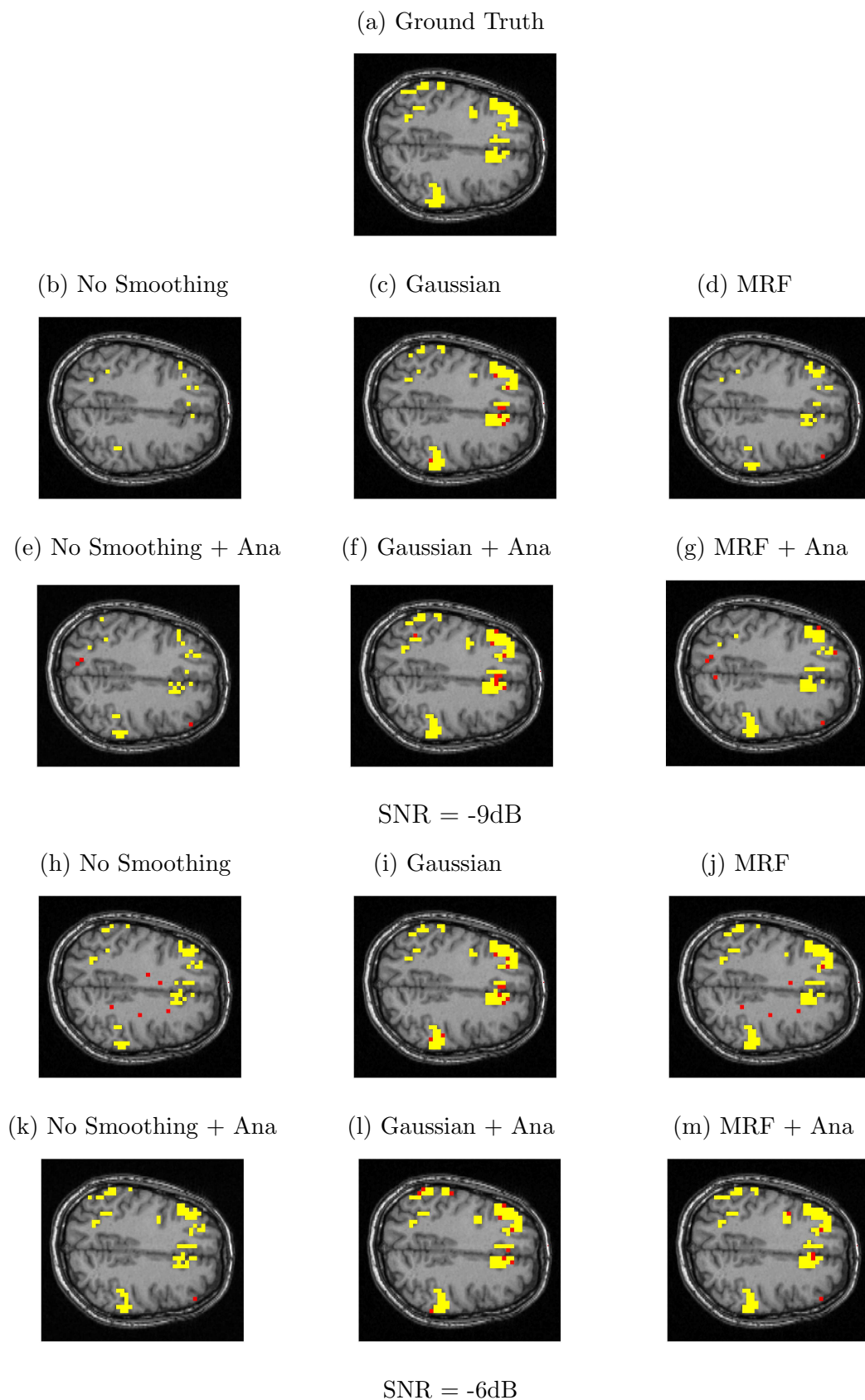


Figure 5-4: One slice from estimated activation maps for the same ground truth at 0.05% false positive rate. True detections are shown in yellow; false positive voxels are shown in red.



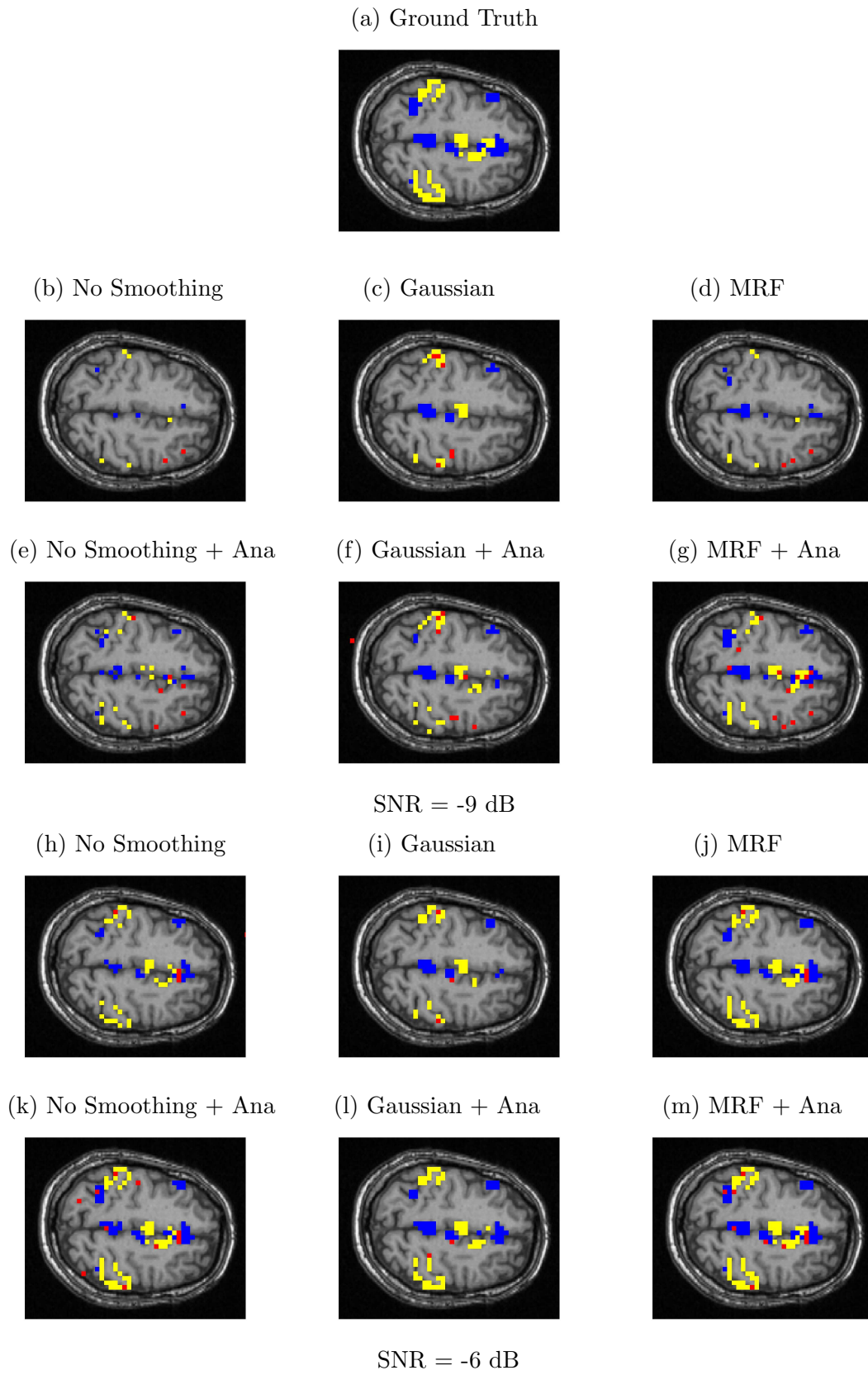
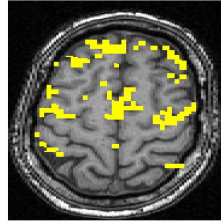
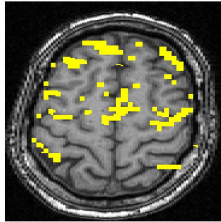


Figure 5-5: One slice from estimated activation maps for the same ground truth at 0.05% false positive rate. True positive and negative detections are shown in yellow and blue; false detections are shown in red.

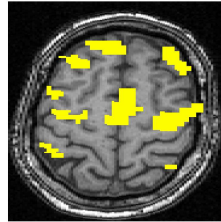
(a) No smoothing (long)



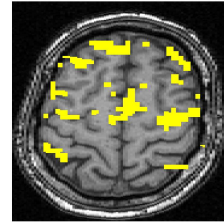
(b) No Smoothing



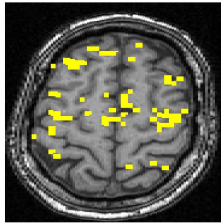
(c) Gaussian



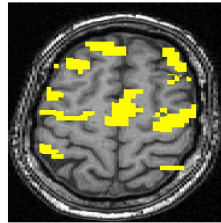
(d) MRF



(e) No Smoothing + Ana



(f) Gaussian + Ana



(g) MRF + Ana

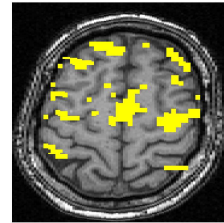
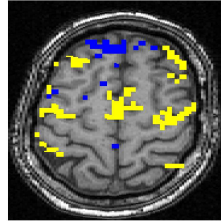
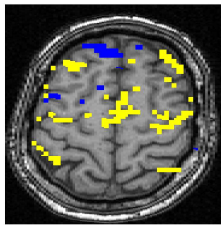


Figure 5-6: *Real fMRI study. One slice in the estimated activation map. (a) No spatial smoothing, using the entire time course. (b)-(f) Estimation based on the first five epochs of the time course using different spatial smoothing methods.*

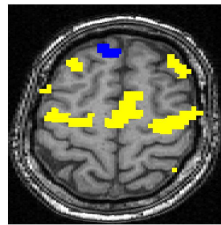
(a) No smoothing (long)



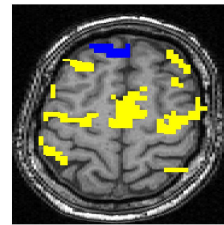
(b) No Smoothing



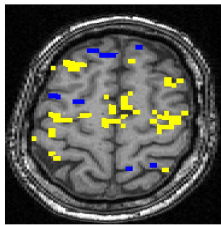
(c) Gaussian



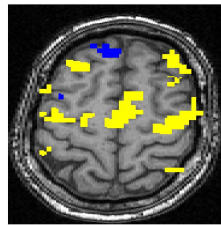
(d) MRF



(e) No Smoothing + Ana



(f) Gaussian + Ana



(g) MRF + Ana

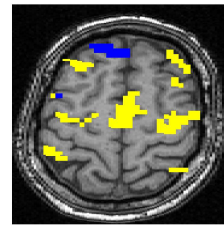


Figure 5-7: *Real fMRI study. One slice in the estimated activation map. (a) No spatial smoothing, using the entire time course. (b)-(f) Estimation based on the first five epochs of the time course using different spatial smoothing methods. Positive and negative detections are shown in yellow and blue, respectively.*



# Chapter 6

## Discussion and Conclusions

This thesis discusses fundamental fMRI analysis techniques, compares different spatial regularization methods, and introduces a novel approach for spatial regularization: incorporating anatomical structure into the regularization process.

In Chapter 3, we discussed the implicit assumptions of the selected voxel-by-voxel detectors: T-test, GLM, and the MI detector. We performed empirical evaluations to verify the trade-offs between the statistical power gained by stronger modelling assumptions and the ability to detect activations that violate these assumptions. In general, incorporating a specific form of HRF into GLM improves its detection power, but this approach misses the areas whose responses do not fit the pre-specified shape of the response function. Using the FIR functions or the Fourier functions to model HRF can reduce this effect. With special assumptions, we also demonstrated that T-test can be derived from the GLM framework. On the other hand, the non-parametric MI-based detector is useful in exploratory studies where the nature of the dependencies between the protocol and the activation is either not known or is suspected to be nonlinear.

Our empirical study of the voxel-by-voxel detectors points to a promising direction in an fMRI research area: incorporating spatial regularization into the voxel-by-voxel detectors. As described in Chapter 4, spatial regularization takes advantage of the spatial coherency of activation. We compared the conventional Gaussian filter with the more sophisticated MRF spatial model. We demonstrated that, in the synthetic

and real data sets, with both binary and trinary activation configurations, GLM-based activation can be further strengthened by incorporating the MRF model. To efficiently solve the MRF problem, we explored the Mean Field approximation, a variation inference algorithm. It provides reasonably accurate approximations to the exact solution in the binary configuration setting, while taking substantially less time to evaluate.

We further refined the spatial regularization process by incorporating anatomical information in Chapter 5. It is straightforward to incorporate anatomical information into the Gaussian filter and the MRF prior. We noticed that the Gaussian-smoothing-based detectors (with and without anatomical information) are not suitable for analysis of fMRI data if we are interested in both positively active and negatively active regions. The anatomically guided MRF-based detector only slightly increases the computation time, while offering significant improvement in detection accuracy.

One of the problems that should be investigated in the future is the partial voluming effects. Images that provide anatomical information have higher resolutions than fMRI scans. In this thesis, we downsampled anatomical scans to match the resolutions of the functional scans. Future studies can improve on our method by using high-resolution anatomical scans to better classify the activation in the functional voxels that are on the boundary of the gray matter, leading to a “super-resolution” detector. Another problem is the registration issue between functional and anatomical images, which causes the anatomically guided MRF-based detector not to perform as well in the real data sets as it does in the synthetic data sets.

We validated the detectors by performing ROC analysis or confusion matrix analysis using phantom data, and we also validated them by studying their ability to recover activation from significantly shorter time courses using real data. In the high noise settings, the Gaussian-smoothing-based detector outperformed other detectors; as the SNR in the images increased, the Markov priors offered a substantial improvement in detection accuracy. Using MRF smoothing prior enabled us to shorten fMRI scan length by half while retaining detection power comparable with full-length fMRI scans. We expect a similar effect to occur with respect to spatial resolution when

we extend the method to utilize the anatomical information at the original scan resolution. As the quality of the scanning equipment improves, sophisticated spatial models, such as MRFs, will become even more important in recovering the details of activation regions.

We hope that researchers in the field will find our empirical results helpful in understanding the fundamental differences among the detectors and in selecting appropriate detection techniques for future studies. We also hope that this research will lead to more sophisticated spatial models for fMRI activation detection.





# Appendix A

## Detailed Derivations

### A.1 The Independence of the Sample Mean and Sample Variance

**Theorem A.1:** Suppose that the random variables,  $X_1, \dots, X_N$  are i.i.d., and each has a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Then, its sample mean,  $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N X_i$ , and sample variance,  $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \hat{\mu})^2$ , are independent random variables. Furthermore,  $\hat{\mu}$  has a normal distribution with mean  $\mu$  and variance  $\sigma^2/N$ , and  $(N-1)\frac{\hat{\sigma}^2}{\sigma^2}$  has a  $\chi^2$ -distribution with  $N-1$  degrees of freedom.

**Proof:** We standardize  $X_i$  by letting  $X'_i = \frac{X_i - \mu}{\sigma}$  for  $i = 1, 2, \dots, N$ . The random variables,  $X'_1, \dots, X'_N$  are i.i.d., and each has a standard normal distribution,  $\mathcal{N}(0, 1)$ .  $X'_i$ 's sample mean and sample variance are  $\hat{\mu}' = \frac{1}{\sigma}(\hat{\mu} - \mu)$  and  $\hat{\sigma}'^2 = \frac{1}{N-1} \sum_{i=1}^N (X'_i - \hat{\mu}')^2$ , respectively.

To prove Theorem A.1, we first prove  $\hat{\mu}'$  and  $\hat{\sigma}'^2$  are independent variables. Then, we will prove that  $\hat{\mu}'$  has a zero-mean normal distribution and that  $\hat{\sigma}'^2$  has a  $\chi^2$ -distribution with  $N-1$  degrees of freedom.

It is straightforward to see that

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N X_i \sim \mathcal{N}(\mu, \sigma^2/N). \quad (\text{A.1})$$

As a result,

$$\hat{\mu}' = \frac{1}{N} \sum_{i=1}^N X'_i = \frac{1}{\sigma} (\hat{\mu} - \mu) \sim \mathcal{N}(0, 1/N). \quad (\text{A.2})$$

We introduce an  $N \times N$  orthonormal matrix,  $A$ , with the first row,  $\vec{u} = [\frac{1}{\sqrt{N}}, \dots, \frac{1}{\sqrt{N}}]$ . We can obtain the rest of the rows by the *Gram-Schmidt process* while maintaining  $A$ 's orthogonality. Since  $A$  is orthonormal,  $A^T A = I$  and  $\det(A) = 1$ .

We define  $\vec{X}' = [X'_1, \dots, X'_N]$  and  $\vec{Y} = [Y_1, \dots, Y_N]$ . Then, we let

$$\vec{Y} = A \vec{X}'. \quad (\text{A.3})$$

Multiplication of  $\vec{X}'$  by an orthonormal matrix,  $A$ , corresponds to a rotation of  $\vec{X}'$  in an  $N$ -dimensional space possibly followed by changing the signs of some coordinates. Since the components in  $\vec{X}'$  are i.i.d. standard normal distributed, the resulting random variables  $Y_1, \dots, Y_n$  are also i.i.d. standard normal distributed. We would like to show that  $\sum_{i=1}^N X'^2_i = \sum_{i=1}^N Y_i^2$ .

$$\sum_{i=1}^N X'^2_i = \vec{X}'^T \vec{X}' = \vec{X}'^T (A^T A) \vec{X}' = \vec{Y}^T \vec{Y} = \sum_{i=1}^n Y_i^2 \quad (\text{A.4})$$

$Y_1$ , the first element in  $\vec{Y}$ , can be expressed as the sample mean of  $X'$ :

$$Y_1 = \vec{u} \vec{X}' = \sum_{i=1}^n \frac{1}{\sqrt{N}} X'_i = \sqrt{N} \hat{\mu}'. \quad (\text{A.5})$$

Combining Eq. (A.4) and Eq. (A.5), we get

$$\sum_{i=2}^n Y_i^2 = \sum_{i=1}^N Y_i^2 - Y_1^2 = \sum_{i=1}^N X'^2_i - N \hat{\mu}'^2 = \sum_{i=1}^N (X'_i - \hat{\mu}')^2. \quad (\text{A.6})$$

Since random variables  $Y_1, \dots, Y_N$  are independent, the two random variables,  $Y_1$  and  $\sum_{i=2}^N Y_i^2$ , are independent which leads to the independent relationship between  $\hat{\mu}'$  and  $\sum_{i=1}^N (X'_i - \hat{\mu}')^2 = (N-1) \hat{\sigma}'^2$ . That means  $\hat{\mu}'$  and  $\hat{\sigma}'^2$  are independent. Moreover, the  $N-1$  random variables  $Y_2, \dots, Y_N$  are i.i.d., and each of them has a standard normal

distribution. Hence,  $\sum_{i=2}^N Y_i^2$  has a  $\chi^2$ -distribution with  $N - 1$  degrees of freedom. Therefore,  $(N - 1)\hat{\sigma}'^2 = \sum_{i=1}^N (X'_i - \hat{\mu}')^2$  also has a  $\chi^2$ -distribution with  $N - 1$  degrees of freedom.

Next, we are going to show the independent relationship between  $\hat{\mu}$  and  $\hat{\sigma}^2$  based on the independency proven above. From Eq.(A.2), we get  $\hat{\mu} = \mu + \sigma\hat{\mu}'$ . On the other hand,

$$\hat{\sigma}'^2 = \frac{1}{N-1} \sum_{i=1}^N (X'_i - \hat{\mu}')^2 \quad (\text{A.7})$$

$$= \frac{1}{N-1} \sum_{i=1}^N (X_i'^2 - \hat{\mu}'^2) \quad (\text{A.8})$$

$$= \frac{1}{(N-1)\sigma^2} \sum_{i=1}^N ((X_i - \mu)^2 - (\hat{\mu} - \mu)^2) \quad (\text{A.9})$$

$$= \frac{1}{(N-1)\sigma^2} \sum_{i=1}^N (X_i^2 - 2N\mu\hat{\mu} + 2N\mu\hat{\mu} - \hat{\mu}^2) \quad (\text{A.10})$$

$$= \frac{1}{(N-1)\sigma^2} \sum_{i=1}^N (X_i^2 - \hat{\mu}^2) \quad (\text{A.11})$$

$$= \frac{\hat{\sigma}^2}{\sigma^2}. \quad (\text{A.12})$$

Hence,  $\hat{\sigma}^2$  can be expressed as  $\hat{\sigma}^2 = \sigma^2\hat{\sigma}'^2$ .  $(N - 1)\hat{\sigma}'^2 \sim \chi_{N-1}^2$  leads to  $(N - 1)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_{N-1}^2$ . Since  $\hat{\mu}'$  and  $\hat{\sigma}'^2$  are independent, random variables,  $\hat{\mu} = \mu + \sigma\hat{\mu}'$  and  $\hat{\sigma}^2 = \sigma^2\hat{\sigma}'^2$ , are independent as well.

An interesting notice to the above proof is that even though the definition of the sample variance is a function of the sample mean, they are independent when the random variables have i.i.d. normal distributions.

## A.2 T-Distribution

The T-distribution is also known as the Student's distribution. This section proves that the ratio of sample mean and sample variance of an i.i.d. normal distribution follows a T-distribution.

**Definition:** Consider two independent random variables  $W$  and  $Z$ , such that  $W$  has a  $\chi^2$  distribution with  $n$  degrees of freedom and  $Z$  has a standard normal distribution. Suppose that a random variable  $T$  is defined by the following equation:

$$T = \frac{Z}{\sqrt{W/n}}. \quad (\text{A.13})$$

Then, the distribution of  $T$  is called that the T-distribution with  $n$  degrees of freedom.

In general, if a random variable  $T$  has a T-distribution with  $n$  degrees of freedom ( $n > 1$ ), then it can be shown that  $E(|T|^k) < \infty$  for  $k < n$  and that  $E(|T|^k) = \infty$  for  $k \geq n$ . In other words, the first  $n - 1$  moments of  $T$  exist, but no moments of higher order exist. Therefore, the moment generating function for a T-distribution does not exist.

In the Appendix A.1, we introduced random samples,  $X_1, \dots, X_N$ , from a normal distribution with mean  $\mu$  and variance  $\sigma^2$ . Again,  $\hat{\mu}$  denotes the sample mean,  $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N X_i$ ;  $\sigma^2$  denotes the sample variance,  $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \hat{\mu})^2$ .

If we define  $Z = \frac{\hat{\mu} - \mu}{\sqrt{\sigma^2/N}}$  and  $W = (N - 1)\hat{\sigma}^2/\sigma^2$ , it follows the proof from the previous section that  $W$  and  $Z$  are independent,  $W$  has a  $\chi^2$ -distribution with  $N - 1$  degrees of freedom, and  $Z$  has standard normal distribution. We let

$$T = \frac{Z}{\sqrt{\frac{W}{N-1}}} = \frac{\hat{\mu} - \mu}{\sqrt{\hat{\sigma}^2/N}}. \quad (\text{A.14})$$

According to the definition,  $T$  has a T-distribution with  $N - 1$  degrees of freedom. An important aspect of Eq. (A.14) is that neither the value  $T$  nor the distribution of  $T$  depends on  $\sigma^2$ .

We can get the following conclusions. Let  $X_1, \dots, X_N$  be random samples from a normal distribution having mean  $\mu$  and variance  $\sigma^2$ . Let  $\hat{\mu} = \frac{1}{N} \sum_{i=1}^N X_i$  and  $\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \hat{\mu})^2$ . Then, the distribution of random variable,  $\frac{\hat{\mu} - \mu}{\sqrt{\hat{\sigma}^2/N}}$ , is a T-distribution with  $N - 1$  degrees of freedom.

## A.3 Confidence Intervals

Confidence intervals provide an alternative to using an estimator  $\hat{\theta}$  when we wish to estimate an unknown parameter  $\theta$ . We can find an interval  $(a, b)$  that we think has high probability of containing  $\theta$ . The length of such an interval gives us an idea of how closely we can estimate  $\theta$ .

If we know  $\hat{\theta}$  has a distribution  $p_{\hat{\theta}}(\theta)$  and we select a confidence level,  $\gamma$ , the confidence interval can be formulated as the following:

$$P(a < \hat{\theta} < b) = \int_a^b p_{\hat{\theta}}(\theta) d\theta = \gamma. \quad (\text{A.15})$$

For a given  $\gamma$ , we can find  $a$  and  $b$  from  $\hat{\theta}$ 's distribution table. Interval  $(a, b)$  is called the confidence interval for  $\hat{\theta}$  with confidence coefficient  $\gamma$ . It is common to make the statement that the unknown value  $\hat{\theta}$  lies in the interval  $(a, b)$  with confidence  $\gamma$ .

### Example: F-distribution

As described in Eq. (3.28),

$$F_i = \frac{(C\hat{\beta}_{1_i})^T (Cov(C\hat{\beta}_{1_i}))^{-1} (C\hat{\beta}_{1_i})}{N_{\beta_1}}.$$

Under  $H_0$ ,  $F_i$  is F-distributed,  $F_i|H_0 \sim F_{N_{\beta_1}, N_T - N_{\beta}}$ .  $N_T$  is the number of samples in a time course.  $N_{\beta}$  and  $N_{\beta_1}$  are the number of regressors in  $\vec{\beta}_i$  and  $\vec{\beta}_{1_i}$ , respectively. A certain confidence coefficient  $\gamma = \gamma^*$  is selected; however, the choice is entirely arbitrary. We can obtain  $(a, b)$  from the F-distribution table. In practice,  $a$  is fixed at zero for F-distribution. If  $F_i \in (b, \infty)$ , we reject the null hypothesis with  $100\gamma^*\%$  confidence.

## A.4 Paired T-test is a special case of GLM

In this section, we are going to show that Paired T-test (T-test) is a special case of GLM. That means, under certain assumptions, we can derive the T-test formula from the GLM framework. A subject undergoes fMRI while alternating between task and rest periods. In most block design experiments, the protocols are in

$\{\text{rest}\} \{\text{task}\} \{\text{rest}\} \dots \{\text{rest}\}$  format. The following derivation is independent of the setting of an experimental protocol.

Let  $\{y_t : t \in \{\text{rest}\}\}$  and  $\{y_t : t \in \{\text{task}\}\}$ . Since both T-test and GLM are voxel-by-voxel detectors, and we are interested in their relationship in detecting the signal of a single voxel, we drop the script,  $i$ , in this section to avoid complicate notation. We denote  $\vec{y}_{rest}$  and  $\vec{y}_{task}$  as the samples taken under the rest and task conditions, respectively. To simplify the later derivation, we first sort the signal samples based on their corresponding experimental conditions.  $\vec{y}$  denotes the sorted signal samples,  $\vec{y} = [\vec{y}_{task} \ \vec{y}_{rest}]$ . To show that T-test is a special case of GLM, we need the following assumptions:

1. We consider the brain as an LTI system, whose impulse response under either task or rest condition is an impulse function. The magnitude of these two impulse functions may be different.
2. Noise is white under either task or rest condition.

We set two hypotheses while deriving GLM through the maximum likelihood ratio approach in Chapter. We rewrite these hypotheses according to the above assumptions as the following:

$$H_0: \vec{y} = B_2\beta_2 + \vec{\epsilon} \text{ (or } \beta_1 = 0)$$

$$H_1: \vec{y} = B_1\beta_1 + B_2\beta_2 + \vec{\epsilon} \text{ (or } \beta_1 \neq 0)$$

where,  $\vec{\epsilon} \sim \mathcal{N}(\vec{0}, \sigma^2 I)$ . According to the first assumption,  $B_1 = [1 \dots 1 \ 0 \dots 0]^T$ ,  $B_2 = [0 \dots 0 \ 1 \dots 1]^T$ , and  $\beta_1$  and  $\beta_2$  are scalars. There are  $N_{task}$  1's following by  $N_{rest}$  0's in  $B_1$  and  $N_{rest}$  1's following by  $N_{task}$  0's in  $B_2$ , where  $N_{task}$  and  $N_{rest}$  are the number of samples taken in the task and rest conditions, respectively.

Let  $\eta_1 = \frac{\beta_1 + \beta_2}{2}$  and  $\eta_2 = \frac{\beta_1 - \beta_2}{2}$ . Combining  $\eta_1$  and  $\eta_2$  into a vector,  $\vec{\eta} = [\eta_1 \ \eta_2]$ , we can express the relation between  $\vec{\eta}$  and  $\vec{\beta}$  as

$$\vec{\eta} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \vec{\beta} \text{ and } \vec{\beta} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \vec{\eta}.$$

We can also rewrite the design matrix as the following:

$$B' = B \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = [B_1 \ B_2] \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & \cdots 1 & 1 & \cdots & 1 \\ 1 & \cdots 1 & -1 & \cdots & -1 \end{pmatrix}^T \quad (\text{A.16})$$

and

$$B'_2 = \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}^T \quad (\text{A.17})$$

where, the first column of  $B'$  are  $N_T$  1's and its second column is  $N_{task}$  1's followed by  $N_{rest}$  0's.  $B'_2$  is a vector of  $N_T$  1's. We can rewrite the two hypotheses as the following:

$$H_0: \vec{y} = B'_2 \eta_1 + \vec{\epsilon} \text{ (or } \eta_2 = 0)$$

$$H_1: \vec{y} = B'_1 \eta_2 + B'_2 \eta_1 + \vec{\epsilon} \text{ (or } \eta_2 \neq 0)$$

where,  $\vec{\epsilon} \sim \mathcal{N}(\vec{0}, \sigma^2 I)$ . From Eq. (3.34), we get

$$F = \frac{\vec{y}^T (B' (B'^T B')^{-1} B'^T - B'_2 (B'_2{}^T B'_2)^{-1} B'_2{}^T) \vec{y} / (2 - 1)}{\vec{y}^T (I - B' (B'^T B')^{-1} B'^T) \vec{y} / (N_T - 2)}. \quad (\text{A.18})$$

Under  $H_0$ ,  $F \sim F_{1, N_T - 2}$ . We can further substitute and expand  $F$  to show that  $F$  equals the squared T-test formula.

$$\begin{aligned} F &= \frac{\vec{y}^T (B' (B'^T B')^{-1} B'^T - B'_2 (B'_2{}^T B'_2)^{-1} B'_2{}^T) \vec{y} / (2 - 1)}{\vec{y}^T (I - B' (B'^T B')^{-1} B'^T) \vec{y} / (N_T - 2)} \\ &= \frac{\bar{y}_t^2 N_{task} + \bar{y}_{rest}^2 N_{rest} - N_T \bar{y}^2}{(\bar{y}_{task}^T \bar{y}_{task} - \bar{y}_{task}^2 N_{task} + \bar{y}_{rest}^T \bar{y}_{rest} - \bar{y}_{rest}^2 N_{rest}) / (N_T - 2)} \\ &= \frac{\bar{y}_{task}^2 N_{task} + \bar{y}_{rest}^2 N_{rest} - (1/N_T) (N_{task} \bar{y}_{task} + N_{rest} \bar{y}_{rest})^2}{(|\bar{y}_{task} - \bar{y}_{rest} \vec{1}|^2 + |\bar{y}_{rest} - \bar{y}_{task} \vec{1}|^2) / (N_T - 2)} \\ &= \frac{\bar{y}_{task}^2 N_{task} (1 - N_{task}/N_T) + \bar{y}_{rest}^2 N_{rest} (1 - N_{rest}/N_T) - 2 N_{task} N_{rest} \bar{y}_{task} \bar{y}_{rest} / N_T}{S^2} \\ &= \frac{(\bar{y}_{task} - \bar{y}_{rest})^2}{S^2 (1/N_{task} + 1/N_{rest})} \end{aligned} \quad (\text{A.19})$$

where,  $\bar{y}_{task}$  and  $\bar{y}_{rest}$  are the sample means taken in the task and rest condition;  $\bar{y}$  is the sample mean taken in the entire experiment, including task and rest periods. Comparing Eq. (A.19) with Eq. (3.2), we get  $F = T_i^2$ . It is possible to compute  $T_i$

from  $F$  if we include the calculation of  $\text{sign}(\bar{y}_{task} - \bar{y}_{rest})$ <sup>1</sup>. To summarize, T-test can be derived as a special case of GLM.

---

<sup>1</sup> $\text{sign}(x) = 1$  if  $x \geq 0$ ;  $\text{sign}(x) = -1$  if  $x < 0$ .







# Bibliography

- [1] BrainVoyager software package. <http://www.brainvoyager.de>.
- [2] FMRI software library. <http://www.fmrib.ox.ac.uk/fsl>.
- [3] Aron, A.R., *et al.* Human Midbrain Sensitivity to Cognitive Feedback and Uncertainty During Classification Learning. *Journal of Neurophysiol*, 92:1144-1152, 2004.
- [4] Besag, J. Spatial interaction and statistical analysis of lattice systems. *Acad. R. Statistical Soc. Series B*, 36:721–741, 1974.
- [5] Burock, M.A., and Dale, A.M. Estimation and detection of event-related fmri signals with temporally correlated noise: A statistically efficient and unbiased approach. *Human Brain Mapping*, 11:249–260, 2000.
- [6] Casella, G. and Berger, R.L. Statistical Inference. *Duxbury*, 1990.
- [7] Cosman, E.R., Fisher, J. and Wells, W.M. Exact MAP activity detection in fMRI using a GLM with an spatial. *In Proc. MICCAI'04*, 2:703–710, 2004.
- [8] Dale, A.M., *et al.* Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. *NeuroImage*, 9:179-194, 1999.
- [9] Descombes, X., Kruggel, F., and Von Cramon, D.Y. fMRI signal restoration using a spatio-temporal Markov random field preserving transitions. *NeuroImage*, 8:340–349, 1998.
- [10] Descombes, X., Kruggel, F. and Von Cramon, D.Y. Spatio-temporal fMRI analysis using Markov random fields. *IEEE TMI*, 17(6):1028–1039, 1998.

- [11] Friman, O. and Westin, C.F. Bias in Resampling-Based Thresholding of Statistical Maps in fMRI. *Proc. MICCAI'04*, 711-718, 2004.
- [12] Friston, K.J., *et al.* Assessing the significance of local activations using their spatial extent. *Human Brain Mapping*, 1:210–220, 1994.
- [13] Friston K.J. Statistical Parametric Maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189-210, 1995.
- [14] Geman, S. and McClure, D. Statistical methods for tomographic image reconstruction. *Proc. 46th Session of ISI*, 51:22–26, 1987.
- [15] Geman, S. and Reynolds, G. Constrained restoration and recovery of discontinuities. *IEEE Trans. PAMI*, 14:367–383, 1992.
- [16] Good, P. Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypothesis. *Springer Verlag*, 1994.
- [17] Greig, D.M., Porteous, B.T. and Gramon, D.Y. Exact maximum a posteriori estimation for binary images. *J. R. Statistical Society*, 51:271–279, 1989.
- [18] Hartvig, N.V. A stochastic Geometry model for functional Magnetic resonance images *Scandinavian Journal of Statistics*, 29:333–253, 2002.
- [19] Hojen-Sorensen, F., Hansen, L.K. and Rasmussen, C.E. Bayesian modeling of fMRI time series. *Adv. Neuroinform. Processing Syst.*, Vol.12:754–760, 2000.
- [20] Jezzard, F., *et al.* Functional MIR – An Introduction to Methods. *OXFORD*, 2002.
- [21] Kapur, T., *et al.* Enhanced spatial priors for segmentation of magnetic resonance imagery. *Proc. MICCAI'98*, 148-157, 1998.
- [22] Kim, J. Spatio-Temporal fMRI Signal Analysis Using Information Theory. *Masters thesis*, Massachusetts Institute of Technology, 2000

- [23] Langan, D.A., *et al.* Use of the mean-field approximation in an EM-based approach to unsupervised stochastic model-based image segmentation. *Proc. ICASSP*, 3:57–60, 1992.
- [24] Liotti, M., *et al.* Brain responses associated with consciousness of breathlessness (air hunger). *PNAS*, 98:2035-2040, 2001
- [25] Mayberg, H.S., *et al.* Regional Metabolic Effects of Fluoxetine in Major Depression: Serial Changes in Relationship to Clinical Response. *Biological Psychiatry*, 48:830-843, 2000.
- [26] Miller, E.G. and Fisher, J.W. ICA using spacings estimates of entropy. *Proc. 4th Symposium of Independent Component Analysis and Blind Signal Separation*, 2003.
- [27] Parisi, G. *Statistical Field Theory*. Addison-Wesley, 1998.
- [28] Pearl, J. Probabilistic reasoning in intelligent systems: networks of plausible inference. *Morgan Kaufmann*, 1988.
- [29] Pohl, K.M., *et al.* Anatomical guided segmentation with non-stationary tissue class distributions in an expectation-maximization framework. *Proc. IEEE ISBI*, 81–84, 2004.
- [30] Pohl, K.M., *et al.* Incorporating non-rigid registration into expectation maximization algorithm to segment MR images. *Proc. MICCAI'02*, 508-515, 2002.
- [31] Rajapakse, J.C. and Piyaratna, J. Bayesian modeling of fMRI time series. *IEEE Transactions on Biomedical Engineering*, 48:1186–1194, 2001.
- [32] Sachs, L. Applied Statistics: A Handbook of Techniques. *Springer Verlag*, 1984.
- [33] Salli, E., *et al.* Contextual clustering for analysis of functional MRI data. *IEEE TMI*, 20:403–413, 2001.
- [34] Schwartz, S. and Maquet P. Sleep imaging and the neuro-psychological assessment of dreams. *Trends in Cognitive Science*, 6:23-30, 2002.

- [35] Seghier, M.L., *et al.* Bayesian modeling of fMRI time series. *NeuroImage*, 21:463–472, 2004.
- [36] Stein, E.A., *et al.* Nicotine-induced limbic cortical activation in the human brain: a functional MRI study. *Am. J. Psychiatry*, 8:1009–15, 1998.
- [37] Student The probable error of a mean. *Biometrika*, 6:1–25, 1908.
- [38] Temple, E., *et al.* Disruption of the Neural Response to Rapid Acoustic Stimuli in Dyslexia. Evidence from Functional MRI". *Proc. Natl. Acad. Sci. USA*, 97:13907–12, 2000.
- [39] Tsai, A., *et al.* Analysis of fMRI data using mutual information. *Proc. 2nd Symposium on Med. Image Computing and Computer-Assited Intervention*, 473–480, 1999.
- [40] Wei, X., *et al.* Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *NeuroImage*, 21:1000–1008, 2004.
- [41] Woolrich, M.W., *et al.* Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE TMI*, 23(2):213–231, 2004.
- [42] Zhang, J. The mean-field theory in EM procedures for markov random field. *IEEE Trans. on Signal Processing*, 40:2570–2583, 1992.