

Experimental Design Issues in Functional Brain Imaging with High Temporal  
Resolution

by

Reem Alghamdi

A Dissertation Presented in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

Approved April 2019 by the  
Graduate Supervisory Committee:

Ming-Hung Kao, Chair  
John Fricks  
Rong Pan  
Mark Reiser  
John Stufken


ARIZONA STATE UNIVERSITY

August 2019

## ABSTRACT

Functional brain imaging experiments are widely conducted in many fields for studying the underlying brain activity in response to mental stimuli. For such experiments, it is crucial to select a good sequence of mental stimuli that allow researchers to collect informative data for making precise and valid statistical inferences at minimum cost. In contrast to most existing studies, the aim of this study is to obtain optimal designs for brain mapping technology with an ultra-high temporal resolution with respect to some common statistical optimality criteria. The first topic of this work is on finding optimal designs when the primary interest is in estimating the Hemodynamic Response Function (HRF), a function of time describing the effect of a mental stimulus to the brain. A major challenge here is that the design matrix of the statistical model is greatly enlarged. As a result, it is very difficult, if not infeasible, to compute and compare the statistical efficiencies of competing designs. For tackling this issue, an efficient approach is built on subsampling the design matrix and the use of an efficient computer algorithm is proposed. It is demonstrated through the analytical and simulation results that the proposed approach can outperform the existing methods in terms of computing time, and the quality of the obtained designs. The second topic of this work is to find optimal designs when another set of popularly used basis functions is considered for modeling the HRF, e.g., to detect brain activations. Although the statistical model for analyzing the data remains linear, the parametric functions of interest under this setting are often nonlinear. The quality of the design will then depend on the true value of some unknown parameters. To address this issue, the maximin approach is considered to identify designs that maximize the relative efficiencies over the parameter space. As shown in the case studies, these maximin designs yield high performance for detecting brain activation compared to the traditional designs that are widely used in practice.

## DEDICATION

*To my mother*  
*Mrs. Sundalah Alghamdi* 

## ACKNOWLEDGMENTS

My graduate study has been a long journey full of joys and difficulties. I sincerely thank God for being with me and for giving me the strength and encouragement especially during all of the challenging moments in completing this dissertation.

First and foremost, I would like to express my sincere gratitude and appreciation to my chair Dr. Ming-Hung Kao for his supervision, guidance and encouragement. His enthusiasm, depth of knowledge, constructive criticism and insightful suggestions are the backbone of this research.

I would also like to extend my appreciation to my committee members: Dr. John Fricks, Dr. Rong Pan, Dr. Mark Reiser and Dr. John Stufken for their concise comments and valuable suggestions during my prospectus and defense.

A special thanks to Dr. Mark Reiser for his continued support. He has been a great mentor since I got admitted to the Master and PhD programs. Also, a special thanks to Dr. Jeffrey Wilson for his supervision and guidance on my master degree project.

I am deeply grateful to every faculty member who taught me here at ASU. I thank all the members of staff at the School of Mathematical and Statistical Science at ASU for making this school inspirational as always, especially IT support, Renate Mittelmann.

I would also like to thank every friend who stood beside me during the difficult moments and helped me to achieve this goal.

I gratefully acknowledge King Abdullah scholarship and King Saud University scholarship for funding me to complete my graduate studies.

More personally, I can not express my thanks to the most important people in my life. I am forever thankful to my beloved parents, Mr. Abdullah Alghamdi and Mrs. Sundalah Alghamdi, as well as my brothers and sisters, for their motivation and

inspiration. Your prayers for me are what sustained me thus far.

Lastly, and most importantly, I would like to express my heartfelt thanks to my wonderful husband, Ahmed Alghamdi, for his tremendous love and all of the sacrifices that he has made on my behalf. Words can not express how grateful and thankful I am to my sweetheart-daughter, Majd, my baby-son, Mohammed, and my butterfly-daughter, Marah, for being such good kids and making my life full of happiness and always cheering me up.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
CHAPTER	
1 INTRODUCTION .....	1
2 BACKGROUND .....	6
2.1 Experimental Design Settings .....	6
2.2 General Linear Models .....	7
2.3 Temporal Autocorrelations in fNIRS and fMRI Time Series .....	10
2.4 Optimality Criteria .....	11
2.5 Maximin Approach .....	12
2.6 A Genetic Algorithm for Obtaining fMRI/fNIRS Designs .....	13
3 A GENERAL LINEAR MODEL FOR ESTIMATION .....	17
3.1 Methodology .....	18
3.2 Stimulation Results with $\Delta T = \tau_{TR}$ .....	35
3.2.1 Estimation of the HRF Under the <i>A</i> -optimality Criterion ...	36
3.3 Stimulation Results with $\Delta T \neq \tau_{TR}$ .....	48
3.3.1 Estimation of the HRF Under the <i>A</i> -optimality Criterion ...	49
4 A GENERAL LINEAR MODEL FOR DETECTION .....	60
4.1 Temporal Derivative .....	60
4.2 Methodology .....	61
4.2.1 Estimation of the Amplitude .....	62
4.3 <i>D</i> -optimality Criterion .....	65
4.4 The Applied Maximin Approach .....	67
4.5 Simulation Results .....	67

CHAPTER	Page
5 CONCLUSION .....	77
REFERENCES .....	79

## LIST OF TABLES

Table	Page
3.1 $A$ -optimality for Estimation When $Q = 1$ ; $\Delta T = \tau_{TR}$ .....	37
3.2 $A$ -optimality for Estimation When $Q = 2$ ; $\Delta T = \tau_{TR}$ .....	40
3.3 Relative Efficiency and CPU for Estimation When $Q = 1, 2, 3, 4,$ and 6; $\Delta T = \tau_{TR}$ .....	42
3.4 $A$ -optimality for Estimation of Method II Versus Method III; $\Delta T = \tau_{TR}$	43
3.5 The Optimal Designs $\mathbf{d}_s^*$ Versus Traditional Designs; $\Delta T = \tau_{TR}$ .....	44
3.6 The Optimal Designs $\mathbf{d}_s^*$ Evaluated at the Traditional Designs Setting; $\Delta T = \tau_{TR}$ .....	45
3.7 $A$ -optimality for Estimation When $Q = 1$ ; $\Delta T \neq \tau_{TR}$ .....	50
3.8 $A$ -optimality for Estimation When $Q = 2$ ; $\Delta T \neq \tau_{TR}$ .....	52
3.9 Relative Efficiency and CPU for Estimation When $Q = 1, 2, 3, 4,$ and 6; $\Delta T \neq \tau_{TR}$ .....	54
3.10 $A$ -optimality for Estimation for Method II Versus Method III; $\Delta T \neq \tau_{TR}$	56
3.11 The Optimal Designs $\mathbf{d}_s^*$ Versus Traditional Designs; $\Delta T \neq \tau_{TR}$ .....	59
4.1 Locally Optimal Designs; Method IA .....	69
4.2 $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_\eta^*, \mathbf{d}_{\theta_0}^*)$ of $\mathbf{d}_{Mm}^*$ Method IA Versus $\mathbf{d}_{\theta_c}^*$ and $\mathbf{d}_{\theta_c \theta_t}^*$ .....	70
4.3 $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_\eta^*, \mathbf{d}_{\theta_0}^*)$ of $\mathbf{d}_{Mm}^*$ From Method IA Versus Some Tradi- tional Designs .....	71
4.4 Locally Optimal Designs; Method IB .....	73
4.5 $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_\eta^*)$ of $\mathbf{d}_{Mm}^*$ Versus $\mathbf{d}_{\theta_c}^*$ and $\mathbf{d}_{\theta_c \theta_t}^*$ ; Method IB .....	74
4.6 $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_\eta^*)$ of $\mathbf{d}_{Mm}^*$ Versus Some Traditional Designs; Method IB	74
4.7 Evaluation of $\mathbf{d}_{Mm}^*$ Method IB Versus $\mathbf{d}_{Mm}^*$ Method IA .....	76
4.8 $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_{\theta_0}^*)$ of $\mathbf{d}_{Mm}^*$ Method IB Versus $\mathbf{d}_{\theta_c}^*$ and $\mathbf{d}_{\theta_c \theta_t}^*$ .....	76



## LIST OF FIGURES

Figure	Page
3.1	Relative Design Efficiencies and Relative CPU Times for Estimation When $Q = 1$ and $\tau_{ISI} = 2, 3, 4,$ and $5$ Seconds . . . . .
	38
3.2	Relative Design Efficiencies and Relative CPU Times for Estimation When $Q = 2$ and $\tau_{ISI} = 2, 3, 4,$ and $5$ Seconds . . . . .
	39
3.3	The Optimal $\mathbf{d}_s^*$ of Method II When $Q = 1$ . . . . .
	46
3.4	The Optimal $\mathbf{d}_s^*$ of Method II When $Q = 2$ . . . . .
	47
3.5	Relative Design Efficiencies and Relative CPU Times for Estimation When $Q = 1; \Delta T \neq \tau_{TR}$ . . . . .
	51
3.6	Relative Design Efficiencies and Relative CPU Times for Estimation When $Q = 2; \Delta T \neq \tau_{TR}$ . . . . .
	53
3.7	The Optimal $\mathbf{d}_s^*$ of Method II When $Q = 1; \Delta T \neq \tau_{TR}$ . . . . .
	57
3.8	The Optimal $\mathbf{d}_s^*$ of Method II When $Q = 2; \Delta T \neq \tau_{TR}$ . . . . .
	58
4.1	The Maximin Designs $\mathbf{d}_{Mm}^*$ Obtained From Method IA . . . . .
	72
4.2	The Maximin Designs $\mathbf{d}_{Mm}^*$ Obtained From Method IB . . . . .
	75

## Chapter 1

### INTRODUCTION

Functional brain imaging is commonly used in research fields such as cognitive neuroscience, medical science, and psychology for studying functions of the brain and human behavior. Many neuro-imaging experiments have been conducted by researchers to understand the inner workings of the brain and many pioneering brain mapping technologies hold great promise for, among others, diagnosing brain disorder such as Alzheimer disease. Currently, there exists a number of different imaging techniques that allow researchers to study some physiological reactions in the brain accompanying brain activation. Each of these techniques provides a unique perspective on brain function, although they differ in what they attempt to measure, as well as in their temporal and spatial resolutions.

Functional Magnetic Resonance Imaging (fMRI) is one of the leading brain mapping techniques that is used for studying brain activity in response to mental stimuli or during a resting state. In typical fMRI experiments for studying the effects of some mental stimuli (e.g., images or sounds) to the brain, each experimental subject is exposed to a sequence of mental stimuli while lying inside an fMRI scanner which repeatedly scans the subject's brain to collect thousands of signals over time. These signals are measured using the different magnetic properties of deoxygenated and oxygenated blood. Changes in the fMRI signal intensity mainly follows from the changes in the ratio of concentrations of deoxygenated and oxygenated blood in activated brain regions (Bandettini and Cox, 2000). This Blood Oxygenation Level Dependent (BOLD) signal is an indirect measure of underlying neuronal activity (Logothetis and Wandell, 2004). The BOLD signals are repeatedly collected by an fMRI

scanner every  $\tau_{TR}$  seconds (e.g., 2) from each brain voxel (three-dimensional imaging element) to form fMRI time series. At the end of the experiment, the fMRI scanner may collect a time series from each of the  $64 \times 64 \times 30$  brain voxels, each having a size of approximately  $3 \times 3 \times 5mm^3$  (Lazar, 2008, Section 2.1.1).

Recently, there has been interest in using a new technique called Ultra-Fast fMRI (UF-fMRI). Efromovich and Valdez-Jasso (2010) proposed a data-driven aggregation method of two wavelet estimators and applied it to analyze data collected from an UF-fMRI that scans the brain with a much faster speed (e.g.,  $\tau_{TR} = 50$  ms) than traditional fMRI. Proulx *et al.* (2014) considered an UF-fMRI called MR-Encephalography, with  $\tau_{TR} = 100$  ms. Lin *et al.* (2014) showed some advantages of an increased fMRI sampling rate such as having more robust and sensitive Granger causality estimates compared to conventional fMRI multi-slice acquisitions. It is also known that several other brain mapping techniques, such as Functional Near-Infrared Spectroscopy (fNIRS) to be introduced in the following paragraph, can attain a similar or an even higher temporal resolution; see also Table 1 of Scholkmann *et al.* (2014).

Functional Near-Infrared Spectroscopy (fNIRS) is a non-invasive brain imaging technique for measuring brain activity based on the absorption of near-infrared light between 650 and 950 nm through the intact skull (Villringer *et al.*, 1993; Villringer and Dirnagl, 1995). Like fMRI, fNIRS is a neurovascular coupling-based neuro-imaging technique that relates functional brain activity to the hemodynamic response (Pinti *et al.*, 2017). fNIRS is able to measure the distinct concentration changes of oxy-hemoglobin (HbO), deoxy-hemoglobin (HbR) and total hemoglobin (HbT) from diffusely scattered light measurements (Tak and Ye, 2014). The measurements of near infrared light reflected from the brain follow the Lambert-Beer law which indicates that HbO and HbR concentration changes can be estimated using optical density measurements at two wavelengths (Villringer and Dirnagl, 1995; Ye *et al.*, 2009; Tak

and Ye, 2014). In comparison to other neuro-imaging techniques, fNIRS is favorable for its high flexibility, portability, low cost, and higher temporal resolution that allows researchers to study the temporal behavior of the hemodynamic response to neural activation. Moreover, it is suitable and less distressing on children, elderly subjects and patients who cannot stand the confined environment of fMRI experiments (Schroeter *et al.*, 2004). However, fNIRS has poor spatial resolution and limited penetration depth due to the high level of light scattering within the tissue (Ye *et al.*, 2009).

For the previously described neuro-imaging studies, the quality of the collected data depends on the selected design (i.e. sequence of mental stimuli). Due to the complexity of these experiments, obtaining the best design is a challenging task. This design issue is normally mathematically intractable and computationally difficult. The large diversity of the experimental settings and protocols greatly contribute to the difficulty of design selection. Kao *et al.* (2009) proposed a genetic algorithm approach, which has been applied in some studies to address some important event-related fMRI (ER-fMRI) design issues. Genetic algorithms are metaheuristic algorithms which are widely considered in various optimization problems. However, this algorithm is not without flaws. It usually tends to require much CPU time, mainly due to a large degree of randomness is involved, and it may take several minutes to hours just to achieve one design as reported in Kao *et al.* (2009). Additionally, a further restriction of the previous works on the design of functional neuro-imaging studies is that they focused only on brain mapping technologies that have relatively low temporal resolutions with, e.g.,  $\tau_{TR} = 2$  seconds. With new advances in neuroscience, many recent studies now use brain mapping techniques that allow a relatively high temporal resolution. But unfortunately, obtaining a good design for these modern experiments can be more challenging. One of the major challenges is that the dimension of the design matrix  $\mathbf{X}$  in the statistical model is greatly enlarged. As to be explained later in this

dissertation, this makes it very difficult, if not infeasible, to compute and compare the statistical efficiencies of competing designs. A novel, efficient approach for obtaining good designs for these functional neuro-imaging studies is called for.

To date, a variety of functional neuro-imaging studies have been conducted for estimating the amplitude of evoked HRFs across different tasks. There is also a growing interest in studying the response latency and duration of activation. For studying these different aspects, choices of HRF models vary from a single canonical HRF to the nonlinear models. One of the popular basis functions is suggested by Glover (1999) to model the HRF by a double-gamma function, which is a linear combination of two Gamma probability density functions; see (2.1). This model assumes a fixed HRF shape for every brain region (voxel or channel) and for every subject. This assumption is not always valid as studies showed that the HRF shape may vary across brain regions and people, and that a misspecified shape can lead to wrong conclusions. To allow for uncertain HRF shapes, statistical analysis methods such as the use of nonlinear models have been studied; e.g., Lindquist and Wager (2007), Lindquist *et al.* (2009), Maus *et al.* (2010), and Kao *et al.* (2013). Another popular method suggested by Friston *et al.* (1998) is based on the double-gamma function and its partial derivative with respect to its delay. The use of the temporal derivative of the double-gamma function helps to capture differences in the onset times of the HRFs. To our knowledge, there currently exists no systematic research on the selection of optimal design for this popularly used model. Some of our findings in studying this design issues for fast neuro-imaging is also presented in this dissertation.

In this work, we are concerned with the experimental design problems for functional neuro-imaging studies when brain mapping technology with high temporal resolution is used. We introduce the relevant background information for this study in Chapter 2. In Chapter 3, we discuss the first topic on the subsampling approach

of the design matrix and illustrate some case studies to demonstrate the usefulness of this approach in obtaining optimal designs for estimating the HRF. In Chapter 4, we introduce the second topic on obtaining optimal designs when the HRF is modeled by a linear combination of the double-gamma function and its temporal derivative. A conclusion is provided in Chapter 5.

## Chapter 2

### BACKGROUND

#### 2.1 Experimental Design Settings

In a functional brain imaging experiment, there might be tens or hundreds of stimuli, possibly of different types, presented to the subject at different time points. Suppose that the experimenter is interested in studying  $Q$  different types of stimuli (e.g.,  $Q$  different images).  $Q$  is normally much smaller than the total number of stimuli, and the stimulus of each type will be presented at multiple time points during the experiment. We further assume that, starting at Time 0, a stimulus can possibly occur every  $\tau_{ISI}$  seconds (e.g.,  $\tau_{ISI} = 3$  seconds); each stimulus does not last longer than  $\tau_{ISI}$ , but stimuli of the same type have the same presentation duration (e.g., 1 seconds). During the time between the presentations of two consecutive stimuli, there will be a control period; e.g., a rest period or a visual fixation for studies with visual stimuli. For convenience, we also set the duration of the experiment to  $(N-1)\tau_{ISI}$  for an integer  $N$ . With this experimental setting, we use  $\mathbf{d} = \{d_1, \dots, d_N\}$  to represent a design (stimulus sequence), where  $d_n \in \{0, 1, \dots, Q\}$  determines the type of the stimulus to be presented at time  $(n-1)\tau_{ISI}$  where  $n = \{1, \dots, N\}$ . Specifically,  $d_n = 0$  indicates no stimulus presentation, whereas  $d_n = q > 0$  indicates an onset of the  $q^{\text{th}}$ -type stimulus at the  $n^{\text{th}}$  time point. While the stimulus sequence  $\mathbf{d}$  is being presented to the subject, a brain mapping instrument (e.g., fMRI or fNIRS) repeatedly collects the data,  $\mathbf{y}$ , from each region of the brain every  $\tau_{TR}$  seconds. The collected data can be used to make an inference about the effect of the stimuli to the brain. At an activated brain region, there will be a fluctuation in the intensity of the

BOLD or fNIRS signals following a stimulus onset. This signal change is typically described by a function of time called the Hemodynamic Response Function (HRF). Counting from the onset of the stimulus, the HRF may take  $\tau_{dur}$  to completely return to the baseline; e.g.  $\tau_{dur} = 32$  seconds for a brief ( $\sim 1$  seconds) stimulus. The HRF is the primary interest to neuroscientists and studying the HRF helps to understand the effect of each stimulus type to the brain. With  $Q$  stimulus types in the study, a study objective of interest is to estimate the corresponding  $Q$  HRFs; we note that stimuli of the same type are typically assumed to give rise to the same HRF throughout the study. Another study objective, which often is referred to as the detection problem, is to detect the regions of the brain that are activated by the stimuli. For the detection problem, the HRF is commonly approximated by the product of an assumed shape of the HRF basis function and an unknown coefficient. One of the popular basis functions used in fMRI/fNIRS studies (Glover, 1999; Ye *et al.*, 2009; Pinti *et al.*, 2017; Uga *et al.*, 2014) is the double gamma function which has the following form:

$$h^*(t) = \frac{t^5 e^{-t}}{5!} - \frac{1}{6} \cdot \frac{t^{15} e^{-t}}{15!}. \quad (2.1)$$

This double-gamma function is a linear combination of two Gamma probability density functions and  $h^*(t)$  is nearly zero after 32 seconds.

## 2.2 General Linear Models

Friston *et al.* (1995) were among the first to apply the general linear model, which assumes that data can be represented as a linear combination of several regressors plus a random error, in the analysis of fMRI time series (see also Worsley and Friston, 1995; Dale, 1999). On the other hand, Schroeter *et al.* (2004) and some other studies have applied general linear model in the analysis of fNIRS experiments (Plichta *et al.*, 2006, 2007; Koh *et al.*, 2007; Ye *et al.*, 2009; Jang *et al.*, 2009). Huppert



*et al.* (2009) introduced the first publicly available fNIRS analysis package (HomER), which uses a linear model and standard statistics to deal with fNIRS time-series and functional analysis techniques. Tak and Ye (2014) provided an extensive review of historical developments of fNIRS including both existing signal processing approaches and existing works of the statistical inference techniques.

Most analyses of functional neuro-imaging data often assume that the HRFs evoked by stimuli of the same type have the same shape and amplitude throughout the experiment. Depending on the selected design (i.e. stimulus sequence), it is often the case (and is preferred) that some HRFs overlap as a result of stimuli appearing in close succession; i.e., the next stimulus appears when the HRFs of the previous stimuli have not yet returned to the baseline. In many cases, the overlapping HRFs can be assumed to accumulate additively to contribute to the overall measured signal; this is especially true when  $\tau_{ISI}$  is not too brief. Most linear models used in neuro-imaging studies are built upon these assumptions. In these models, a basis function such as (2.1) is often used to model the shape of the HRF. It is also not uncommon to model the HRF by using the so-called Finite Impulse Response (FIR) basis functions. Specifically, the  $i$ th FIR basis function has a value of 1 in  $[(i-1)(\Delta T), i(\Delta T)]$  and is 0 otherwise, where  $\Delta T$  is the greatest value that divides both  $\tau_{TR}$  and  $\tau_{ISI}$ . Huppert *et al.* (2009) demonstrated the use of the FIR basis functions in HomER to allow researchers to visualize the shape of the HRF, and it allows HRFs of arbitrary shapes. Moreover, structured physiological noise and motion artifacts can affect fMRI and fNIRS results if they are not properly accounted for (Huppert, 2016). Furthermore, the autocorrelation of the collected data should be taken into account when analyzing functional neuro-imaging experiments. In what follows, we describe some commonly used general linear models for functional neuro-imaging studies.

The following linear model is commonly considered in many fMRI studies (Dale,

1999; Liu and Frank, 2004; Kao *et al.*, 2009), and similar models are also applied in many fNIRS studies (Huppert *et al.*, 2009; Barker *et al.*, 2013; Huppert, 2016).

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad (2.2)$$

where  $\mathbf{y}$  is a  $T \times 1$  vector of the signals collected by, e.g., fMRI or fNIRS from brain region and  $\mathbf{X}$  is the  $T \times (QK)$  design matrix for the parameter vector  $\mathbf{h}$ ; the matrix  $\mathbf{X}$  is determined by the selected design  $\mathbf{d}$ . The  $(QK) \times 1$  vector  $\mathbf{h} = (\mathbf{h}'_1, \dots, \mathbf{h}'_Q)'$  corresponds to the  $Q$  HRFs with  $\mathbf{h}_q = (h_{q1}, \dots, h_{qK})'$  representing the unknown heights of the HRF for the  $q^{\text{th}}$ -type stimulus at the stimulus onset ( $h_{q1}$ ) and the following  $K - 1$  (regularly spaced) time points. Specifically, the set of  $K$  FIR basis functions is used to model each HRF, and  $\mathbf{h}_q$  consists of the  $K$  corresponding coefficients for these basis functions. It is noteworthy that  $h_{qk}$  is the HRF height at  $(k - 1)\Delta T$  seconds after the stimulus onset;  $q = 1, \dots, Q$ ,  $k = 1, \dots, K$ , and  $K = \lfloor 1 + \tau_{dur}/\Delta T \rfloor$  with  $\lfloor a \rfloor$  being the integer part of  $a$ . With the previously defined  $\Delta T$ , the HRF parameters  $\mathbf{h}$  will contain all the HRF heights that can possibly contribute to  $\mathbf{y}$ ; see also Kao *et al.* (2009). The specified matrix  $\mathbf{S}$  and the unknown parameter vector  $\boldsymbol{\gamma}$  are included in Model (2.2) for modeling some nuisance effects such as the physiological noise and motion artifacts (which often appears as a trend or drift of  $\mathbf{y}$ ). The error  $\boldsymbol{\epsilon}$  is Normally distributed with expectation 0 and variance-covariance matrix  $\boldsymbol{\Sigma}$ . Model (2.2) is often used for estimating the (shape of) the HRFs,  $\mathbf{h}$ .

When a set of  $b$  basis functions is used to model the shape of the HRF, we may replace  $\mathbf{h}$  in Model (2.2) by  $\mathbf{H}\boldsymbol{\theta}$ . Here,  $\mathbf{H}$  is  $(\mathbf{I}_Q \otimes \mathbf{H}^*)$  where  $\mathbf{H}^*$  is  $K \times b$  matrix of the basis functions and  $\boldsymbol{\theta}$  consists of the  $Qb$  unknown coefficients. For example,  $\mathbf{H}^*$  can be a vector ( $b = 1$ ) formed by  $h^*(t)$  in (2.1) or a matrix ( $b = 3$ ) formed by  $h^*(t)$  and its first and second derivatives. In these cases, the statistical model considered

for analyzing  $\mathbf{y}$  becomes:

$$\mathbf{y} = \mathbf{X}\mathbf{H}\boldsymbol{\theta} + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad (2.3)$$

where a considered  $\boldsymbol{\theta} = [\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_Q]'$  represents the coefficients with  $\boldsymbol{\theta}_q = [\theta_{q1}, \dots, \theta_{qb}]'$ ,  $q = 1, \dots, Q$ . All the remaining terms of Model (2.3) are as in Model (2.2). It is noteworthy that the first term on the right hand side of (2.3) represents the convolution of the event onsets with the assumed basis functions of the HRF. Additionally, Model (2.2) can be viewed as a special case of Model (2.3) by setting  $\mathbf{H}^*$  to the identity matrix ( $\mathbf{I}_K$ ) of size  $K$ ; the  $\mathbf{H}$  matrix corresponds to the FIR basis functions.

### 2.3 Temporal Autocorrelations in fNIRS and fMRI Time Series

The autocorrelation of the collected BOLD signals is commonly modeled by the AR(1) process (Bullmore *et al.*, 1996). This is especially true for traditional fMRI settings when  $\tau_{TR}$  is not very small. Also, the use of the AR(2) process is not uncommon in the literature (Lindquist *et al.*, 2008). Some traditional fMRI studies have modeled noise in alternative ways such as using the  $m^{th}$  order autoregressive model, AR( $m$ ), (Worsley *et al.*, 2002) or first order autoregressive moving-average, ARMA(1, 1) model (Purdon *et al.*, 2001).

Some fNIRS studies also have applied AR(1) process for the correlated errors (Plichta *et al.*, 2006; Hofmann *et al.*, 2008; Jang *et al.*, 2009; Tak *et al.*, 2016). We thus assumed in our case studies that the error terms follow a stationary AR(1) process, although some other correlation structure such as AR( $m$ ) with  $m > 1$  may also be considered. It is noteworthy that Barker *et al.* (2013) proposed a general algorithm for solving the general linear model in the context of deconvolution (by using FIR basis functions) and canonical regression models for fNIRS that combines two established statistical methods: AR( $m$ ) models and iteratively re-weighted least squares. Their results demonstrated that the AR(1) model was insufficient for controlling type I

errors. With this, an AR( $m$ ) model with  $m > 1$  might be recommended for fNIRS studies or other functional neuro-imaging studies with a high temporal resolution. While we assume an AR(1) model for simplicity, our proposed method is not restricted to such a simple process and can be considered when other models are used.

## 2.4 Optimality Criteria

When the estimations of  $\mathbf{h}$  in Model (2.2) or  $\boldsymbol{\theta}$  in Model (2.3) are of interest, we would like an optimal design yielding the best least squares estimates,  $\hat{\mathbf{h}}$  or  $\hat{\boldsymbol{\theta}}$ ; respectively. Precisely, the Generalized Least Square Estimators (GLSE) are (Seber, 1977):

$$\begin{cases} \hat{\mathbf{h}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}, \text{ for estimation Model (2.2);} \\ \hat{\boldsymbol{\theta}} = (\mathbf{H}'\mathbf{X}'\mathbf{W}\mathbf{X}\mathbf{H})^{-1}\mathbf{H}'\mathbf{X}'\mathbf{W}\mathbf{y}, \text{ for detection Model (2.3).} \end{cases}$$

where  $\mathbf{W} = \mathbf{V}'(\mathbf{I}_T - \mathcal{P}_{\mathbf{VS}})\mathbf{V}$  and  $\mathcal{P}_{\mathbf{VS}} = \mathbf{V}\mathbf{S}((\mathbf{V}\mathbf{S})'\mathbf{V}\mathbf{S})^{-1}(\mathbf{V}\mathbf{S})'$  is the orthogonal projection matrix onto the column space of  $\mathbf{V}\mathbf{S}$ ,  $\mathbf{V}$  is a matrix satisfying  $\mathbf{V}\Sigma\mathbf{V}' \propto \mathbf{I}_T$ , and  $\mathbf{I}_T$  is the  $T$ -by- $T$  identity matrix. Following the theorem for the inverse of partitioned matrices (Harville, 1977; Seber, 1977; Rencher and Schaalje, 2008), the information matrices  $\mathbf{M}$  can be written as:

$$\begin{cases} \mathbf{M}(\mathbf{h}) = \mathbf{X}'\mathbf{W}\mathbf{X}; \\ \mathbf{M}(\boldsymbol{\theta}) = \mathbf{H}'\mathbf{X}'\mathbf{W}\mathbf{X}\mathbf{H}. \end{cases} \quad (2.4)$$

We target a design  $\mathbf{d}$  that makes this information matrix as 'large' as possible; note that  $\mathbf{M}(\mathbf{h})$  and  $\mathbf{M}(\boldsymbol{\theta})$  both depend on  $\mathbf{d}$  through the design matrix  $\mathbf{X}$ .

Dale (1999) and Friston *et al.* (1999) evaluated the quality of designs by considering the A-optimality criterion which can be formulated as the reciprocal of the trace of the inverse information matrix or the variance-covariance matrix. This criterion has then been applied in many fMRI design studies (Liu *et al.*, 2001; Wager and Nichols,

2003; Liu and Frank, 2004; Kao *et al.*, 2009; Maus *et al.*, 2010). In addition to the A-optimality criterion, the D-optimality criterion helps to find designs maximizing the determinant of  $\mathbf{M}(\mathbf{h})$  or  $\mathbf{M}(\boldsymbol{\theta})$ . This latter criterion also has been considered in, e.g., Wager and Nichols (2003). For a non-singular  $\mathbf{M}$ , the previously mentioned criteria can be defined as follows:

$$\phi(\mathbf{d}) = \begin{cases} \mathcal{R}/\text{tr}(\mathbf{M}^{-1}), & \text{for } A\text{-optimality;} \\ \frac{1}{\mathcal{R}} \log \det(\mathbf{M}), & \text{for } D\text{-optimality.} \end{cases} \quad (2.5)$$

Here  $\mathcal{R}$  is set to  $QK$  for Model (2.2) and  $Qb$  for Model (2.3) which corresponds to the number of the parameters of interest. For the designs which make  $\mathbf{M}$  singular, we set  $\phi(\mathbf{d})$  to 0. It also should be clear that, with the criteria in (2.5), the error variance  $\text{Var}(\boldsymbol{\epsilon}_t) = \sigma^2$  does not impact the selection of designs. Without loss of generality, we assume  $\sigma^2 = 1$  throughout this dissertation.

Finding an optimal design that maximizes the *A*-criterion or *D*-criterion normally requires efficient computer search algorithm. In cases where the information matrix depends on unknown parameters, it is useful to obtain a robust design that performs relatively well over the possible value of the unknown parameters. One approach to achieve such a design is the maximin approach. In the next section, we describe maximin approach and the genetic algorithm of Kao *et al.* (2009) for finding optimal designs for fMRI studies.

## 2.5 Maximin Approach

The main idea is to select an optimal design to efficiently achieve the study objectives (estimation and/or detection) by maximizing specific statistical criteria. However in some cases, these selected criteria may involve an unknown parameter vector, say,  $\boldsymbol{\varphi}$  whose value is normally uncertain at the design stage. One of the way to solve this problem is to consider a locally optimal design, which is the best design

for a given parameter vector value (Chernoff, 1953). Because a good guess for the parameter vector value is not always available, another possible approach we may consider is the maximin approach, which has been considered by Maus *et al.* (2010), Kao *et al.* (2013) and Kao and Mittelman (2014). The aim is at finding a design that is rather robust to a mis-specification of the  $\varphi$ -value. In particular, this is to obtain a maximin design  $\mathbf{d}_{Mm}^*$  that maximizes

$$\min_{\varphi \in \Omega} RE(\mathbf{d}; \mathbf{d}_{\varphi}^*) = \min_{\varphi \in \Omega} \frac{\phi(\mathbf{d}; \varphi)}{\phi(\mathbf{d}_{\varphi}^*; \varphi)}, \quad (2.6)$$

where  $\Omega$  is the parameter space of  $\varphi$ ,  $\phi(\mathbf{d}; \varphi)$  represents the value of the selected, the larger-the-better optimality criterion for a candidate design  $\mathbf{d}$  evaluated at a given  $\varphi$ -value,  $\mathbf{d}_{\varphi}^*$  is the locally optimal design maximizing  $\phi$  for the given  $\varphi$ . With these  $\mathbf{d}_{\varphi}^*$ s, we find  $\min_{\varphi \in \Omega} RE(\mathbf{d}; \mathbf{d}_{\varphi}^*)$  for each candidate design  $\mathbf{d}$ . To obtain  $\min_{\varphi \in \Omega} RE(\mathbf{d}; \mathbf{d}_{\varphi}^*)$ , we typically will need to find locally optimal designs  $\mathbf{d}_{\varphi}^*$  for all the  $\varphi$ -values over (a finite subset of)  $\Omega$ . Solving such a maximin design problem can be very time consuming. An efficient computational approach is often needed.

## 2.6 A Genetic Algorithm for Obtaining fMRI/fNIRS Designs

To search for a good design over the enormous space of all the candidate designs, Wager and Nichols (2003) advocated the use of the Genetic Algorithm (GA) technique. Later, Kao *et al.* (2009) proposed an improved GA which not only is faster but also can obtain fMRI designs that outperform those achieved by the algorithm of Wager and Nichols (2003).

In this study, we adapt Kao *et al.* (2009)'s GA to our first research topic. A brief description of the GA is in Algorithm 1.

A MATLAB program implementing this algorithm can be found in Kao (2009). When using this algorithm, we use the default values for the algorithmic parameters

---

**Algorithm 1**

---

- Step 1.** (Initial designs) Generate  $G$  (an even integer) initial designs as parents, which include blocked designs, random designs,  $m$ -sequence-based designs, and mixed designs. Obtain their design efficiencies based on the selected optimality criterion.
- Step 2.** (Crossover) With probability proportional to the design efficiency, randomly draw with replacement  $G/2$  pairs of distinct parent designs to crossover, i.e. to select a random cut-point and exchange the corresponding fractions before the cut-point of the paired designs.
- Step 3.** (Mutation) Randomly select a portion ( $q\%$ ) of the elements of all the  $G$  designs obtained in Step 2, and randomly replace these elements with randomly generated integers from the discrete uniform distribution over  $0, 1, 2, \dots, Q$ . The resulting  $G$  designs form the offspring designs.
- Step 4.** (Immigration) Randomly generate  $I$  immigrant designs from random designs, blocked designs, and their combinations (i.e. mixed designs).
- Step 5.** (Fitness) Obtain the design efficiencies of the offspring and immigrant designs obtained in Steps 3 and 4.
- Step 6.** (Natural selection) Select the best  $G$  designs from the pool of the parent, offspring, and immigrant designs to form the parents of the next generation and discard the others.
- Step 7.** (Stop) Repeat Steps 2 through 6 until a stopping rule is met. Then, keep track of the best design over generations.
-

presented in Kao (2009); i.e.,  $G = 20$ ,  $q\% = 1\%$ , and  $I = 4$ . The stopping rule that we use in the numerical result is the second stopping rule described in Kao (2009). That is, the algorithm will be stopped once there is no significant improvement in the achieved value of the optimality criterion. During the process, the accumulated improvement of the design efficiency is calculated every 200 generations of the GA, and is compared to the improvement achieved in the first 200 GA generations. The search is terminated once the improvement in the last 200 generations is no more than  $10^{-7}$  of that of the first 200 generations. In addition, we present below the initial designs that are included in the GA:

**Random designs.** Each element of a random design is generated randomly from a discrete uniform distribution over  $\{0, 1, \dots, Q\}$ .

***M*-sequences.** These designs are known as maximum-length shift register sequences, and are introduced to fMRI by Buracas and Boynton (2002). Specifically, an *m*-sequence is a Linear Recurring Sequence (LRS) over a finite field  $GF(Q + 1)$  of length  $N = (Q + 1)^p - 1$  for some integer  $p$  when  $Q + 1$  is a prime power; see Kao and Stufken (2015) for more details. These designs are known to outperform many other designs in estimating the HRFs, and can be easily generated by the MATLAB program of Liu (2004).

**Blocked designs.** We consider blocked designs having a 16 seconds-on-16 seconds-off pattern (e.g., rest period). For example, when  $Q = 1$ , the first 16 seconds is the off-period, and no stimulus is presented. During the next 16 seconds, stimuli of the same type is presented every  $\tau_{ISI}$ . This pattern is repeated for several cycles until the end of the experiment. These designs are known to perform well in the detection of activated brain regions. However, they do not perform well when the estimation of the HRF is of the primary interest, and



may rise the issue of confounding psychological effects (e.g., subject habituation or anticipation).

**Mixed designs.** These designs are usually obtained by concatenating a fraction of a blocked design with a fraction of an  $m$ -sequence ( or a random design). These designs are sometimes considered when both estimating the HRF and detecting brain activations are of interest.

## A GENERAL LINEAR MODEL FOR ESTIMATION

One of the study objectives in many functional imaging studies is to estimate the HRF. By considering a model that is the same or similar to (2.2), some high-quality designs for yielding a precise estimate of the HRF have been proposed in literature. For example, Dale (1999) indicated that random designs that have no perceivable pattern tend to yield a high statistical efficiency for estimating the HRF,  $\mathbf{h}$ . Buracas and Boynton (2002) advocated the use of  $m$ -sequences, whereas Aguirre *et al.* (2011) proposed the consideration of some de Bruijn sequences. Kao (2014, 2015) then put forward designs that can be generated by some Hadamard sequences and Paley-difference sets. Discussions on the optimality of the previously mentioned designs and their variants can be found in Kao (2013), Cheng *et al.* (2015) and Cheng *et al.* (2017). Lin *et al.* (2017) then extended some of the previous results to obtain optimal or highly efficient designs by using a certain type of orthogonal arrays, called circulant (almost-)orthogonal arrays.

However, the previous research has so far been focused only on traditional fMRI studies with a relatively low temporal resolution (e.g.,  $\tau_{TR} = 2$  seconds). To our knowledge, there is no guideline on the selection of optimal design for high temporal resolution neuro-imaging studies. One of our main goals is to address this void. In what follows, we first describe a major challenge on obtaining optimal designs for high temporal resolution neuro-imaging studies. We then propose an approach for tackling this challenge to efficiently obtain good designs for estimating the HRF in functional neuro-imaging experiments that utilize an ultra-fast or higher temporal resolution brain imaging technique.

### 3.1 Methodology

The linear model (2.2) is considered for modeling the data collected from experiments having high temporal resolution brain imaging. For this type of experiment, a brain mapping instrument (e.g. fMRI or fNIRS) collects the measured signals with high sampling rate (i.e. a small  $\tau_{TR}$ ). This high sampling rate will greatly enlarge the length of the HRF parameter vector  $\mathbf{h}$  and the size of the information matrix. This makes obtaining the optimal design very challenging. We now describe our proposed method for tackling this challenging design issue. To describe the main idea of our proposed method, we consider both cases whether  $\tau_{TR}$  divides  $\tau_{ISI}$  or not and rewrite Model (2.2) as:

$$\mathbf{y} = \mathbf{X}_s \mathbf{h}_s + \mathbf{X}_r \mathbf{h}_r + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \quad (3.1)$$

for some judiciously selected sub-design matrix  $\mathbf{X}_s$ . Specifically, possibly after some column permutations, the design matrix  $\mathbf{X}$  in (2.2) is partitioned as  $\mathbf{X} = [\mathbf{X}_s, \mathbf{X}_r]$ , and correspondingly  $\mathbf{h} = (\mathbf{h}'_s, \mathbf{h}'_r)'$ . All the remaining terms in (3.1) are defined as in (2.2). Instead of considering this full model, our main idea is to obtain an optimal design for  $\mathbf{h}_s$  by considering the following reduced model:

$$\mathbf{y} = \mathbf{X}_s \mathbf{h}_s + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\epsilon}. \quad (3.2)$$

An explanation for our selection of  $\mathbf{X}_s$  and  $\mathbf{h}_s$  will be provided below. With a given  $p$  and  $j$ , we obtain  $\mathbf{h}_s$  by sub-sampling each HRF parameter vector,  $\mathbf{h}_q$ . Specifically,  $\mathbf{h}_s = (\mathbf{h}'_{s1}, \dots, \mathbf{h}'_{sQ})'$ , where  $\mathbf{h}_{sq} = (h_{q,j}, h_{q,(p+j)}, \dots, h_{q,(\lfloor \frac{K-j}{p} \rfloor p + j)})'$  is a subvector of  $\mathbf{h}_q$  selected with specific  $p$  and  $j$ . With  $j = 1$ , each  $\mathbf{h}_{sq}$  consists of the unknown HRF heights of the  $q^{\text{th}}$ -type stimuli at (post-stimulus) time points  $0, p(\Delta T), 2p(\Delta T), \dots, \lfloor \frac{K-1}{p} \rfloor p(\Delta T)$ ; recall that  $\Delta T$  is the greatest value that makes both  $(\tau_{TR}/\Delta T)$  and  $(\tau_{ISI}/\Delta T)$  integers. With such  $\mathbf{h}_s$ ,  $\mathbf{X}_s$  can be formed accordingly. Specifically, we

may write  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_Q]$  in Model (3.1) with  $\mathbf{X}_q$  being the design matrix for the  $q^{\text{th}}$ -type stimuli, and write  $\mathbf{X}_s = [\mathbf{X}_{s1}, \dots, \mathbf{X}_{sQ}]$  in Model (3.2). The matrix  $\mathbf{X}_{sq}$  then consists of Columns  $j, (p+j), (2p+j), \dots, (\lfloor \frac{K-j}{p} \rfloor p + j)$  of  $\mathbf{X}_q$ . Note that  $\mathbf{X}_r$  in (3.1) is then the collection of all the remaining columns of  $\mathbf{X}$ . The selection of  $\mathbf{X}_s$  is further explained below.

To begin, we describe a general construction of the design matrix  $\mathbf{X}$  for a given design  $\mathbf{d} = \{d_1, \dots, d_N\}$ . In order to construct the design matrix for the  $q^{\text{th}}$ -type stimulus, we first define  $\boldsymbol{\delta}_q$  as the 0-1 indicator vector for the  $q^{\text{th}}$ -type stimulus; see also Kao and Stufken (2015); Kao *et al.* (2012). Specifically, the  $n^{\text{th}}$  element of  $\boldsymbol{\delta}_q$  is  $\delta_{q,n} = 1$  if the corresponding  $d_n$  in the design  $\mathbf{d} = \{d_1, \dots, d_N\}$  is the  $q^{\text{th}}$  stimulus type (i.e.  $d_n = q$ ); otherwise,  $\delta_{q,n} = 0$ . Define  $\boldsymbol{\omega}_q$  as the onset times of the  $q^{\text{th}}$ -type stimulus in the resolution of  $\Delta T$ . We have,

$$\boldsymbol{\omega}_q = \boldsymbol{\delta}_q \otimes [1, \mathbf{0}'_{m_{ISI}-1}]'$$

Here, we assume that the number of observations  $T$  is  $N(\tau_{ISI}/\tau_{TR})$  and  $m_{ISI} = \tau_{ISI}/\Delta T$ . By deleting rows from  $\boldsymbol{\omega}_q$  and keeping only rows  $1 + (i-1)m_{TR}$ , the resulting  $\boldsymbol{\omega}_q$  is then used to construct  $\mathbf{X}_q$  as follows:

$$\mathbf{X}_q = [\mathbf{I}_T \otimes (1, \mathbf{0}'_{m_{TR}-1})][\boldsymbol{\omega}_q, \mathbf{B}\boldsymbol{\omega}_q, \mathbf{B}^2\boldsymbol{\omega}_q, \dots, \mathbf{B}^{K-1}\boldsymbol{\omega}_q]; \quad q = 1, \dots, Q,$$

where  $m_{TR} = (\tau_{TR}/\Delta T)$ ,  $i = 1, 2, \dots, 1 + [(T-1)/m_{TR}]$ ,  $\mathbf{I}_T$  is  $T \times T$ ,  $K = \lfloor \tau_{dur}/\Delta T \rfloor + 1$ , and  $\mathbf{B} = \begin{bmatrix} \mathbf{0}' & 0 \\ \mathbf{I}_{T-1} & \mathbf{0} \end{bmatrix}$ .

The following three steps are considered for obtaining  $\mathbf{X}_s$ :

Step X-1. Partition  $\mathbf{X}$  as  $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_Q]$  where  $\mathbf{X}_q$  is the 0-1 design matrix for stimuli of the  $q^{\text{th}}$  type.

Step X-2. Set  $p = m_{ISI}$  (or one of its divisors) such that  $p > m_{TR}$  and keep columns  $j, (j + p), (j + 2p), \dots, (j + \lfloor \frac{K-j}{p} \rfloor p)$  of  $\mathbf{X}_q$  for  $j = 1, \dots, m_{TR}$  and leave out the other columns to form  $\mathbf{X}_{sq}$  for  $q = 1, \dots, Q$ .

Step X-3. Set  $\mathbf{X}_s = [\mathbf{X}_{s1}, \mathbf{X}_{s2}, \dots, \mathbf{X}_{sQ}]$ .

Each design matrix  $\mathbf{X}_{sq}$  can be written as  $\mathbf{X}_{sq} = [\mathbf{X}_{sq,1}, \mathbf{X}_{sq,2}, \dots, \mathbf{X}_{sq,m_{TR}}]$  and each design matrix  $\mathbf{X}_{sqj}$  can be written as:

$$\mathbf{X}_{sq,j} = [\mathbf{I}_T \otimes (1, \mathbf{0}'_{m_{TR}-1})][\mathbf{B}^{j-1}\boldsymbol{\omega}_q, \mathbf{B}^{p+j-1}\boldsymbol{\omega}_q, \mathbf{B}^{2p+j-1}\boldsymbol{\omega}_q, \dots, \mathbf{B}^{\lfloor K-j/p \rfloor p+j-1}\boldsymbol{\omega}_q],$$

where  $j = 1, \dots, m_{TR}$

As a toy example, we consider a design  $\mathbf{d} = \{1, 1, 1, 0, 0, 0\}$  with  $Q = 1$ ,  $\tau_{ISI} = 2$  seconds,  $\tau_{TR} = 0.8$  seconds, and  $\Delta T = 0.4$  seconds. We further assume  $\tau_{dur} = 4$  seconds for illustration purposes. In this case,  $K = \lfloor 4/\Delta T \rfloor + 1 = 11$ ,  $m_{ISI} = 5$ , and

$$\mathbf{X}\mathbf{h} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ \vdots \\ h_{11} \end{bmatrix}.$$

By using the previous approach,  $\mathbf{X}_s$  can be subsampled from  $\mathbf{X}$  for, e.g.,  $p = 4$  with cluster of  $m_{TR} = 2$  and  $\mathbf{X}_r$  contains the remaining columns of  $\mathbf{X}$ . Specifically, we

have

$$\mathbf{X}_s \mathbf{h}_s = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_5 \\ h_6 \\ h_9 \\ h_{10} \end{bmatrix}, \mathbf{X}_r \mathbf{h}_r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_3 \\ h_4 \\ h_7 \\ h_8 \\ h_{11} \end{bmatrix}.$$

But, if we set  $p = 5$  as an example of the divisor of  $m_{ISI}$ , we have:

$$\mathbf{X}_s \mathbf{h}_s = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_6 \\ h_7 \\ h_{11} \end{bmatrix}, \mathbf{X}_r \mathbf{h}_r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_3 \\ h_4 \\ h_5 \\ h_8 \\ h_9 \\ h_{10} \end{bmatrix}.$$

We note that the columns of  $\mathbf{X}_s$  are orthogonal to every column of  $\mathbf{X}_r$ ; i.e.  $\mathbf{X}_s' \mathbf{X}_r = \mathbf{0}$  when we set  $p = 5$  but this is not true when  $p = 4$ .

As a special case when  $\tau_{TR} = \Delta T$  (i.e.  $m_{TR} = 1$ ), we consider the same design  $\mathbf{d} = \{1, 1, 1, 0, 0, 0\}$  for simplicity with  $Q = 1$ ,  $\tau_{ISI} = 2$  seconds, but  $\tau_{TR} = \Delta T = 0.4$  seconds. Again, we further assume  $\tau_{dur} = 4$  seconds for illustration purposes. In this case,  $K = \lfloor 4/\Delta T \rfloor + 1 = 11$ ,  $m_{ISI} = 5$ , and

$$\mathbf{X}\mathbf{h} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \\ h_{10} \\ h_{11} \end{bmatrix}.$$

By using the previous approach,  $\mathbf{X}_s$  can be subsampled from  $\mathbf{X}$  for, e.g.,  $p = 2$  and  $\mathbf{X}_r$  contains the remaining columns of  $\mathbf{X}$ . Specifically, we have

$$\mathbf{X}_s\mathbf{h}_s = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_3 \\ h_5 \\ h_7 \\ h_9 \\ h_{11} \end{bmatrix}, \mathbf{X}_r\mathbf{h}_r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_2 \\ h_4 \\ h_6 \\ h_8 \\ h_{10} \end{bmatrix}.$$

For this toy example, the ordinary least square estimator (OLSE) of  $\mathbf{h}_s = (h_1, h_3, h_5, h_7, h_9, h_{11})'$  in Model (3.2) is  $\hat{\mathbf{h}}_s = (\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{X}'_s\mathbf{y}$ . By first assuming that  $\mathbf{S}\boldsymbol{\gamma} = 0$  and  $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_T$ , we see that the expectation of  $\hat{\mathbf{h}}_s$  under the full Model (3.1) is  $E(\hat{\mathbf{h}}_s) = \mathbf{h}_s + \mathbf{A}\mathbf{h}_r$  where  $\mathbf{A}$  is the alias matrix,

$$\mathbf{A} = (\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{X}'_s\mathbf{X}_r = \begin{bmatrix} 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{bmatrix}.$$

Clearly,  $\hat{\mathbf{h}}_s$  is biased by an amount that depends on the elements of both  $\mathbf{A}$  and  $\mathbf{h}_r$ , and here,

$$E(\hat{\mathbf{h}}_s) = \begin{bmatrix} \hat{h}_1 \\ \hat{h}_3 \\ \hat{h}_5 \\ \hat{h}_7 \\ \hat{h}_9 \\ \hat{h}_{11} \end{bmatrix} = \begin{bmatrix} h_1 + \frac{1}{3}h_6 \\ h_3 + \frac{1}{3}h_8 \\ h_5 + \frac{1}{3}h_{10} \\ h_7 + \frac{1}{3}h_2 \\ h_9 + \frac{1}{3}h_4 \\ h_{11} + \frac{1}{3}h_6 \end{bmatrix}.$$

But, if we set  $p = m_{ISI} = 5$ , we have:

$$\mathbf{X}_s \mathbf{h}_s = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_6 \\ h_{11} \end{bmatrix}, \quad \mathbf{X}_r \mathbf{h}_r = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_7 \\ h_8 \\ h_9 \\ h_{10} \end{bmatrix}.$$

The OLSE of  $\mathbf{h}_s = (h_1, h_6, h_{11})'$  is  $\hat{\mathbf{h}}_s = (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{X}'_s \mathbf{y}$ , and again by assuming  $\mathbf{S}\boldsymbol{\gamma} = \mathbf{0}$ , the expectation of  $\hat{\mathbf{h}}_s$  is now  $E(\hat{\mathbf{h}}_s) = \mathbf{h}_s$ . This follows directly from the fact that the columns of  $\mathbf{X}_s$  are orthogonal to every column of  $\mathbf{X}_r$ ; i.e.  $\mathbf{X}'_s \mathbf{X}_r = \mathbf{0}$ , and is generally true due to the following well-known result (the proof is omitted):

**Lemma 3.1.1** *The OLSE of  $\hat{\mathbf{h}}_s$  of  $\mathbf{y} = \mathbf{X}_s \mathbf{h}_s + \boldsymbol{\epsilon}$  is unbiased for  $\mathbf{h}_s$  under the full model  $\mathbf{y} = \mathbf{X}_s \mathbf{h}_s + \mathbf{X}_r \mathbf{h}_r + \boldsymbol{\epsilon}$  whenever  $\mathbf{X}'_s \mathbf{X}_r = \mathbf{0}$ .*

The key idea of our proposed design approach is to utilize the  $\phi$ -value (2.5) of the information matrix,  $\mathbf{M}_s = \mathbf{X}'_s \mathbf{W} \mathbf{X}_s$  of  $\mathbf{h}_s$  in Model (3.2) in lieu of that of



the ‘full’ information matrix  $\mathbf{M} = \mathbf{X}'\mathbf{W}\mathbf{X}$  of  $\mathbf{h}$  of Model (3.1). Here we denote the former  $\phi$ -value as  $\phi(\mathbf{d}|\mathbf{X}_s)$ , and refer the latter one to as  $\phi(\mathbf{d}|\mathbf{X})$ ; i.e.  $\phi(\mathbf{d}|\mathbf{X}_s)$  serves as a surrogate objective function for  $\phi(\mathbf{d}|\mathbf{X})$ . Under certain conditions, we show that a judiciously selected  $\mathbf{X}_s$  not only allows an unbiased estimate of  $\mathbf{h}_s$  as suggested by Lemma 3.1.1, but also can make  $\phi(\mathbf{d}|\mathbf{X}_s)$  a good surrogate of  $\phi(\mathbf{d}|\mathbf{X})$ . By moving away from some of these conditions, our simulation results further suggest that the use of  $\phi(\mathbf{d}|\mathbf{X}_s)$  allows us to efficiently identify very good designs for some realistic situations. To present our results, we first note that most optimality criteria  $\phi$ , including those in (2.5), are invariant to a simultaneous permutation of rows and columns of the information matrix. Consequently, a permutation of the columns of  $\mathbf{X}$  will not change the value of  $\phi(\mathbf{d}|\mathbf{X})$ . For simplicity, we thus will continue to write  $\mathbf{X}$  even at places where  $\mathbf{X}\mathbf{P}$  should be used with a permutation matrix  $\mathbf{P}$ ; i.e., we omitted such a  $\mathbf{P}$  matrix hereinafter.

As in the previous toy example, we continue to consider the following assumptions:

**Assumption 1.**  $\Sigma = \sigma^2\mathbf{I}_T$ , is the  $T$ -by- $T$  identity matrix.

**Assumption 2.** The nuisance term  $\mathbf{S}\boldsymbol{\gamma} = 0$ .

**Assumption 3.** The total scanning time will allow the HRF evoked by the last stimulus of any selected design to return to its baseline. Note that this assumption is not a strong assumption, and it is often achieved in practice.

**Assumption 4.** We consider cases with  $K \bmod p = 0$ .

With these assumptions, it can be easily seen that  $\mathbf{X}'\mathbf{W}\mathbf{X} = \mathbf{X}'\mathbf{X}$  whose diagonal elements are the same. With  $m_{TR} = 1$  (i.e.  $\tau_{TR}$  divides  $\tau_{ISI}$ ), we may write  $\mathbf{X}_q = [\mathbf{x}_{q,1}, \dots, \mathbf{x}_{q,K}]$  where  $\mathbf{x}_{q,k}$  is the  $k^{\text{th}}$  column. Then, under Assumption 3,  $\mathbf{B}^{k-1}\mathbf{x}_{q,1} = \mathbf{L}^{k-1}\mathbf{x}_{q,1}$ , where  $\mathbf{L}$  is the same as  $\mathbf{B}$ , except that the  $(1, T)^{\text{th}}$  element of  $\mathbf{L}$  is 1 in-

stead of 0; i.e.  $\mathbf{x}_{q,k}$  is a cyclic shift of  $\mathbf{x}_{q,k-1}$ . Therefore,  $\mathbf{x}'_{q,i}\mathbf{x}_{q,j} = \mathbf{x}'_{q,1}\mathbf{L}^{j-i}\mathbf{x}_{q,1}$ . For mathematical simplicity and under Assumption 4, we write  $\mathbf{X}_s = [\mathbf{X}_{s,1}, \dots, \mathbf{X}_{s,m_{TR}}]$ , where  $\mathbf{X}_{s,j} = [\mathbf{X}_{s,j1}, \dots, \mathbf{X}_{s,jQ}]$  and  $\mathbf{X}_{s,jq} = [x_{q,j}, \dots, x_{q, \lfloor \frac{qK-j}{p} \rfloor_{p+1}}]$ . Thus, we obtain the following matrices from (subsampling)  $\mathbf{X}$ :

$$\begin{aligned}
\mathbf{X}_s &= \begin{bmatrix} \mathbf{X}_{s,1} & \cdots & \mathbf{X}_{s,m_{TR}} \end{bmatrix} \\
&= \left[ (\mathbf{x}_{1,1}, \mathbf{x}_{1,p+1}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor_{p+1}}) \quad \cdots \quad (\mathbf{x}_{1,m_{TR}}, \mathbf{x}_{1,p+m_{TR}}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-m_{TR}}{p} \rfloor_{p+m_{TR}}}) \right], \\
\mathbf{X}_r^{(1)} &= \left[ (\mathbf{x}_{1,m_{TR}+1}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor_{p+m_{TR}+1}}) \right] = \mathbf{L}\mathbf{X}_{s,1}, \\
\mathbf{X}_r^{(2)} &= \left[ (\mathbf{x}_{1,m_{TR}+2}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-2}{p} \rfloor_{p+m_{TR}+2}}) \right] = \mathbf{L}\mathbf{X}_{s,2}, \\
&\vdots \\
\mathbf{X}_r^{(m_{TR})} &= \left[ (\mathbf{x}_{1,2m_{TR}}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-m_{TR}}{p} \rfloor_{p+2m_{TR}}}) \right] = \mathbf{L}\mathbf{X}_{s,m_{TR}}, \\
\mathbf{X}_r^{(m_{TR}+1)} &= \left[ (\mathbf{x}_{1,2m_{TR}+1}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor_{p+2m_{TR}+1}}) \right] = \mathbf{L}^2\mathbf{X}_{s,1}, \\
&\vdots \\
\mathbf{X}_r^{(2m_{TR})} &= \left[ (\mathbf{x}_{1,3m_{TR}}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-m_{TR}}{p} \rfloor_{p+3m_{TR}}}) \right] = \mathbf{L}^2\mathbf{X}_{s,m_{TR}}, \\
&\vdots \\
\mathbf{X}_r^{(p-1)} &= \left[ (\mathbf{x}_{1,p+m_{TR}-1}, \dots, \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor_{p+p+m_{TR}-1}}) \right] = \mathbf{L}^{\lceil \frac{p-m_{TR}}{m_{TR}} \rceil} \mathbf{X}_{s,1}, \\
&\vdots \\
\mathbf{X}_r^{(p-m_{TR})} &= \left[ \mathbf{x}_{1,p} \quad \mathbf{x}_{1,2p} \quad \mathbf{x}_{1,3p} \quad \cdots \quad \mathbf{x}_{Q, \lfloor \frac{QK-m_{TR}}{p} \rfloor_{p+p}} \right] \\
&= \mathbf{L}^{\lceil \frac{p-m_{TR}}{m_{TR}} \rceil} \mathbf{X}_{s,l}.
\end{aligned}$$

$$l = \begin{cases} 1 & , \text{ when } p \bmod m_{TR} = 0; \\ g & , \text{ when } p \bmod m_{TR} \neq 0, \end{cases} \quad (3.3)$$

where  $g \in \{1, \dots, m_{TR} - 1\}$  such that  $p = a \times m_{TR} + g$  for some integer  $a$ . For the first case of (3.3) when  $p \bmod m_{TR} = 0$ , we can obtain the following matrices from (subsampling)  $\mathbf{X}$  or it can be derived from the previous matrix when  $m_{TR} = 1$ :

$$\begin{aligned} \mathbf{X}_s &= \begin{bmatrix} \mathbf{x}_{1,1} & \mathbf{x}_{1,p+1} & \cdots & \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor p+1} \end{bmatrix}, \\ \mathbf{X}_r^{(1)} &= \begin{bmatrix} \mathbf{x}_{1,2} & \mathbf{x}_{1,p+2} & \cdots & \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor p+2} \end{bmatrix} = \mathbf{L} \mathbf{X}_s, \\ \mathbf{X}_r^{(2)} &= \begin{bmatrix} \mathbf{x}_{1,3} & \mathbf{x}_{1,p+3} & \cdots & \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor p+3} \end{bmatrix} = \mathbf{L}^2 \mathbf{X}_s, \\ &\vdots \\ \mathbf{X}_r^{(p-1)} &= \begin{bmatrix} \mathbf{x}_{1,p} & \mathbf{x}_{1,2p} & \cdots & \mathbf{x}_{Q, \lfloor \frac{QK-1}{p} \rfloor p+p} \end{bmatrix} = \mathbf{L}^{p-1} \mathbf{X}_s. \end{aligned}$$

The  $\mathbf{X}_r$  matrix in (3.1) is then defined as  $\mathbf{X}_r = [\mathbf{X}_r^{(1)}, \dots, \mathbf{X}_r^{(p-1)}]$ . It is noteworthy that when  $p$  is  $m_{ISI}$  or its divisors, we can have  $\mathbf{X}_s$  orthogonal to  $\mathbf{X}_r$  and  $\mathbf{X}_r^{(i)}$  orthogonal to  $\mathbf{X}_r^{(j)}$  for  $i \neq j$ .

**Proposition 3.1.2** *With the previously described assumptions 1-4 and when  $p$  is  $m_{ISI}$  or one of its divisors, and  $m_{TR} = 1$  s.t.  $p > m_{TR}$ , we have:  $\phi(\mathbf{d}|\mathbf{X}) = \phi(\mathbf{d}|\mathbf{X}_s)$ .*

Proof: We write  $\mathbf{X} \sim \mathbf{Z}$  when  $\mathbf{X}$  and  $\mathbf{Z}$  are similar matrices. With  $\mathbf{X}_s$ , and  $\mathbf{X}_r$

defined above, and  $p$  being  $m_{ISI}$  or its divisor, we have:

$$\mathbf{X}'\mathbf{X} \sim \begin{bmatrix} \mathbf{X}'_s\mathbf{X}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_r\mathbf{X}_r \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_s\mathbf{X}_s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_r^{(1)}\mathbf{X}_r^{(1)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}'_r^{(p-1)}\mathbf{X}_r^{(p-1)} \end{bmatrix} \quad (3.4)$$

$$= \begin{bmatrix} \mathbf{X}'_s\mathbf{X}_s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_s\mathbf{X}_s & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}'_s\mathbf{X}_s \end{bmatrix}. \quad (3.5)$$

The last equality follows from the orthogonality of  $\mathbf{X}_r^{(i)}$  and  $\mathbf{X}_r^{(j)}$  and that  $(\mathbf{X}_r^{(j)})'\mathbf{X}_r^{(j)} = (\mathbf{L}^{j-1}\mathbf{X}_s)'(\mathbf{L}^{j-1}\mathbf{X}_s) = \mathbf{X}'_s\mathbf{X}_s$ . For designs that makes  $\mathbf{h}$  (and thus  $\mathbf{h}_s$ ) estimable, we have:

$$\text{tr}[(\mathbf{X}'\mathbf{X})^{-1}] = p \text{tr}((\mathbf{X}'_s\mathbf{X}_s)^{-1}). \quad (3.6)$$

$$\log \det[(\mathbf{X}'\mathbf{X})] = p \log \det(\mathbf{X}'_s\mathbf{X}_s). \quad (3.7)$$

Then, we substitute (3.6) in (2.5) for  $A$ -optimality as follows:

$$\begin{aligned} \phi(\mathbf{d}|\mathbf{X}) &= \mathcal{R} / \text{tr}((\mathbf{X}'\mathbf{X})^{-1}) \\ &= \mathcal{R} / (p \text{tr}((\mathbf{X}'_s\mathbf{X}_s)^{-1})) \\ &= \mathcal{R}_s / \text{tr}((\mathbf{X}'_s\mathbf{X}_s)^{-1}) \\ &= \phi(\mathbf{d}|\mathbf{X}_s). \end{aligned}$$

The third equality,  $\mathcal{R}/p = \mathcal{R}_s$ , where  $\mathcal{R}_s$  corresponds to the number of columns of

$\mathbf{X}_s$ . Also, we substitute (3.7) in (2.5) for  $D$ -optimality as follows:

$$\begin{aligned}
\phi(\mathbf{d}|\mathbf{X}) &= \frac{1}{\mathcal{R}} \log \det(\mathbf{X}'\mathbf{X}) \\
&= \frac{p}{\mathcal{R}} \log \det(\mathbf{X}'_s\mathbf{X}_s) \\
&= \frac{1}{\mathcal{R}_s} \log \det(\mathbf{X}'_s\mathbf{X}_s) \\
&= \phi(\mathbf{d}|\mathbf{X}_s).
\end{aligned}$$

For the second case when  $p \bmod m_{TR} \neq 0$ , i.e.  $m_{TR} \neq 1$ , we have the following proposition.

**Proposition 3.1.3** *With the previously described assumptions 1-4 and when  $p$  is  $m_{ISI}$  or its divisors and  $m_{TR} \neq 1$ , we have  $\phi(\mathbf{d}|\mathbf{X}) < \phi(\mathbf{d}|\mathbf{X}_s)$  as:*

*i. for A-optimality:*

$$\begin{aligned}
\phi(\mathbf{d}|\mathbf{X}) &\propto \mathcal{R}_s / [\text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1}) + \Delta_1]. \\
\phi(\mathbf{d}|\mathbf{X}_s) &= \mathcal{R}_s / [\text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1})].
\end{aligned}$$

Here,  $\Delta_1 = \text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,l}\mathbf{X}_{s,l})^{-1})$ , where  $l = g$ .

*ii. for D-optimality:*

$$\begin{aligned}
\phi(\mathbf{d}|\mathbf{X}) &\propto \frac{1}{\mathcal{R}_s} [\log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})) + \Delta_2]. \\
\phi(\mathbf{d}|\mathbf{X}_s) &= \frac{1}{\mathcal{R}_s} [\log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}}))].
\end{aligned}$$

Here,  $\Delta_2 = \log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,l}\mathbf{X}_{s,l}))$ , where  $l = g$ .

Proof: We write  $\mathbf{X} \sim \mathbf{Z}$  when  $\mathbf{X}$  and  $\mathbf{Z}$  are similar matrices. With  $\mathbf{X}_s$ , and  $\mathbf{X}_r$

defined above, and  $p$  being  $m_{ISI}$  or its divisor, we have:

$$\begin{aligned}
\mathbf{X}'\mathbf{X} &\sim \begin{bmatrix} \mathbf{X}'_s\mathbf{X}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_r\mathbf{X}_r \end{bmatrix} = \begin{bmatrix} \mathbf{X}'_s\mathbf{X}_s & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_r^{(1)}\mathbf{X}_r^{(1)} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}'_r^{(p-m_{TR})}\mathbf{X}_r^{(p-m_{TR})} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{X}'_{s,1}\mathbf{X}_{s,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_{s,2}\mathbf{X}_{s,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}'_r^{(p-m_{TR})}\mathbf{X}_r^{(p-m_{TR})} \end{bmatrix}.
\end{aligned} \tag{3.8}$$

$$\text{Where } \mathbf{X}'_s\mathbf{X}_s = \begin{bmatrix} \mathbf{X}'_{s,1}\mathbf{X}_{s,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}'_{s,2}\mathbf{X}_{s,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}} \end{bmatrix}.$$

For designs that makes  $\mathbf{h}$  (and thus  $\mathbf{h}_s$ ) estimable, we have:

$$\begin{aligned}
\text{tr}[(\mathbf{X}'\mathbf{X})^{-1}] &= \text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1}) + \\
&\quad \text{tr}((\mathbf{X}'_r^{(1)}\mathbf{X}_r^{(1)})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_r^{(p-m_{TR})}\mathbf{X}_r^{(p-m_{TR})})^{-1}) \\
&= \left\lceil \frac{p-m_{TR}}{m_{TR}} \right\rceil [\text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1})] + \Delta_1.
\end{aligned} \tag{3.9}$$

$$\log \det[(\mathbf{X}'\mathbf{X})] = \left\lceil \frac{p-m_{TR}}{m_{TR}} \right\rceil \log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})) + \Delta_2. \tag{3.10}$$

where  $\Delta_1 = \text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,l}\mathbf{X}_{s,l})^{-1})$ , and  $\Delta_2 = \log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,l}))$ .

Then, we substitute (3.9) in (2.5) for  $A$ -optimality as follows:

$$\begin{aligned}
\phi(\mathbf{d}|\mathbf{X}) &= \mathcal{R} / \text{tr}((\mathbf{X}'\mathbf{X})^{-1}) \\
&= \mathcal{R} / \left( \left[ \frac{p - m_{TR}}{m_{TR}} \right] [\text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1})] + \Delta_1 \right) \\
&\propto \mathcal{R}_s / \left( \left[ \frac{p - m_{TR}}{pm_{TR}} \right] \text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1}) + \frac{1}{p}\Delta_1 \right) \\
&\propto \mathcal{R}_s / [\text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1}) + \Delta_1]. \\
\phi(\mathbf{d}|\mathbf{X}_s) &= \mathcal{R}_s / (\text{tr}((\mathbf{X}'_{s,1}\mathbf{X}_{s,1})^{-1}) + \cdots + \text{tr}((\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})^{-1})).
\end{aligned}$$

Here  $\mathcal{R} = \mathcal{R}_s \times p$  and  $\lceil \frac{p-m_{TR}}{pm_{TR}} \rceil \approx 1$ . Also, we substitute (3.10) in (2.5) for  $D$ -optimality as follows:

$$\begin{aligned}
\phi(\mathbf{d}|\mathbf{X}) &= \frac{1}{\mathcal{R}} \log \det(\mathbf{X}'\mathbf{X}) \\
&= \frac{1}{\mathcal{R}} \left[ \left[ \frac{p - m_{TR}}{m_{TR}} \right] \log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})) + \Delta_2 \right] \\
&\propto \frac{1}{\mathcal{R}_s} [\log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}})) + \Delta_2]. \\
\phi(\mathbf{d}|\mathbf{X}_s) &= \frac{1}{\mathcal{R}_s} [\log \det((\mathbf{X}'_{s,1}\mathbf{X}_{s,1}) + \cdots + (\mathbf{X}'_{s,m_{TR}}\mathbf{X}_{s,m_{TR}}))].
\end{aligned}$$

In what follows, we study  $\phi(\mathbf{d}|\mathbf{X})$  and  $\phi(\mathbf{d}|\mathbf{X}_s)$  by considering the same assumptions as in Proposition 3.1.2 and replacing assumption 2 with:

**Assumption 2\***.  $\mathbf{S} = \mathbf{j}_T$ , is the  $T$ -by-1 vector of ones.

The follow two results are useful, which can be found in Rencher and Schaalje (2008).

**Lemma 3.1.4** *If a square matrix of the form  $\mathbf{U} + \mathbf{v}\mathbf{v}'$  is nonsingular, where  $\mathbf{v}$  is a vector and  $\mathbf{U}$  is nonsingular matrix, then*

$$(\mathbf{U} + \mathbf{v}\mathbf{v}')^{-1} = \mathbf{U}^{-1} - \frac{\mathbf{U}^{-1}\mathbf{v}\mathbf{v}'\mathbf{U}^{-1}}{1 + \mathbf{v}'\mathbf{U}^{-1}\mathbf{v}}.$$

**Lemma 3.1.5** If  $\mathbf{Z}$  has the form  $\mathbf{Z} = \begin{pmatrix} \mathbf{U} & \mathbf{v} \\ \mathbf{v}' & 1 \end{pmatrix}$ , where  $\mathbf{v}$  is a vector and  $\mathbf{U}$  is nonsingular matrix, then

$$\det[\mathbf{U} + \mathbf{v}\mathbf{v}'] = (1 + \mathbf{v}'\mathbf{U}^{-1}\mathbf{v}) \det(\mathbf{U}).$$

**Proposition 3.1.6** Suppose the assumption 2\* and the other assumptions are the same as in Proposition 3.1.2, we have  $\phi(\mathbf{d}|\mathbf{X}) < \phi(\mathbf{d}|\mathbf{X}_s)$  as:

i. for the A-optimality criterion:  $\phi(\mathbf{d}|\mathbf{X}) = \mathcal{R}_s / [\text{tr}(\mathbf{X}'_s \mathbf{X}_s)^{-1} + \frac{\mathcal{B}}{1-p\mathcal{C}}]$ ,

$$\phi(\mathbf{d}|\mathbf{X}_s) = \mathcal{R}_s / [\text{tr}(\mathbf{X}'_s \mathbf{X}_s)^{-1} + \frac{\mathcal{B}}{1-\mathcal{C}}].$$

ii. for the D-optimality criterion:  $\phi(\mathbf{d}|\mathbf{X}) = \frac{1}{\mathcal{R}_s} [1 - p\mathcal{C}] \log \det(\mathbf{X}'_s \mathbf{X}_s)$ .

$$\phi(\mathbf{d}|\mathbf{X}_s) = \frac{1}{\mathcal{R}_s} [1 - \mathcal{C}] \log \det(\mathbf{X}'_s \mathbf{X}_s).$$

Here,  $\mathcal{B} = m^2 T^{-1} [\mathbf{j}'_{K_s} (\mathbf{X}'_s \mathbf{X}_s)^{-2} \mathbf{j}_{K_s}]$ ,  $\mathcal{C} = m^2 T^{-1} \mathbf{j}'_{K_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{K_s}$ ,  $\mathcal{R}_s = Q \times K_s$ ;  $K_s$  corresponds to the number of columns of  $\mathbf{X}_{sq}$  and  $m$  is the frequency of each stimulus.

Proof: With  $\mathbf{X}_s$ , and  $\mathbf{X}_r$  defined previously, and  $p$  being  $m_{ISI}$  or its divisor, we have:

$$\begin{aligned} \mathbf{X}'\mathbf{W}\mathbf{X} &\sim \begin{bmatrix} \mathbf{X}'_s \mathbf{W} \mathbf{X}_s & \mathbf{X}'_s \mathbf{W} \mathbf{X}_r \\ \mathbf{X}'_r \mathbf{W} \mathbf{X}_s & \mathbf{X}'_r \mathbf{W} \mathbf{X}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{X}'_s \mathbf{W} \mathbf{X}_s & \mathbf{X}'_s \mathbf{W} \mathbf{X}_r^{(1)} & \dots & \mathbf{X}'_s \mathbf{W} \mathbf{X}_r^{(p-1)} \\ \mathbf{X}'_r^{(1)} \mathbf{W} \mathbf{X}_s & \mathbf{X}'_r^{(1)} \mathbf{W} \mathbf{X}_r^{(1)} & \dots & \mathbf{X}'_r^{(1)} \mathbf{W} \mathbf{X}_r^{(p-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}'_r^{(p-1)} \mathbf{W} \mathbf{X}_s & \mathbf{X}'_r^{(p-1)} \mathbf{W} \mathbf{X}_r^{(1)} & \dots & \mathbf{X}'_r^{(p-1)} \mathbf{W} \mathbf{X}_r^{(p-1)} \end{bmatrix} \end{aligned} \quad (3.11)$$



The information matrices of  $\mathbf{X}'_i \mathbf{W} \mathbf{X}_j$  in (3.11) are as the following:

$$\begin{aligned}
\mathbf{X}'_i \mathbf{W} \mathbf{X}_j &= \mathbf{X}'_i \mathbf{V}' (\mathbf{I} - \mathcal{P}_{\mathbf{V}_s}) \mathbf{V} \mathbf{X}_j \\
&= \mathbf{X}'_i \mathbf{V}' \mathbf{V} \mathbf{X}_j - (\mathbf{j}'_T \mathbf{V}' \mathbf{V} \mathbf{j}_T)^{-1} \mathbf{X}'_i \mathbf{V}' \mathbf{V} \mathbf{j}_T \mathbf{j}'_T \mathbf{V}' \mathbf{V} \mathbf{X}_j \\
&= \mathbf{X}'_i \mathbf{X}_j - (\mathbf{j}'_T \mathbf{j}_T)^{-1} \mathbf{X}'_i \mathbf{j}_T \mathbf{j}'_T \mathbf{X}_j \\
&= \begin{cases} \mathbf{X}'_s \mathbf{X}_s - (T)^{-1} (\mathbf{L}^{i-1} \mathbf{X}_s)' \mathbf{j}_T \mathbf{j}'_T (\mathbf{L}^{j-1} \mathbf{X}_s), & \text{if } i = j \\ -(T)^{-1} (\mathbf{L}^{i-1} \mathbf{X}_s)' \mathbf{j}_T \mathbf{j}'_T (\mathbf{L}^{j-1} \mathbf{X}_s), & \text{if } i \neq j \end{cases} \\
&= \begin{cases} \mathbf{X}'_s \mathbf{X}_s - (T)^{-1} m^2 \mathbf{j}_{QK_s} \mathbf{j}'_{QK_s}, & \text{if } i = j \\ -(T)^{-1} m^2 \mathbf{j}_{QK_s} \mathbf{j}'_{QK_s}, & \text{if } i \neq j \end{cases} \\
&= \begin{cases} \mathbf{X}'_s \mathbf{W} \mathbf{X}_s, & \text{if } i = j; \\ -(T)^{-1} m^2 \mathbf{j}_{QK_s} \mathbf{j}'_{QK_s}, & \text{if } i \neq j. \end{cases}
\end{aligned}$$

The fifth equality based on  $(\mathbf{L}^{j-1} \mathbf{X}_s)' \mathbf{j}_T = m \mathbf{j}_{QK_s}$ , and  $\mathbf{X}'_s \mathbf{j}_T = m \mathbf{j}_{QK_s}$ , where  $\mathbf{X}_j = \mathbf{L}^{j-1} \mathbf{X}_s$  and  $m$  corresponds to the frequency of each stimulus. Set  $\mathbf{D} = (T)^{-1} m^2 \mathbf{j}_{QK_s} \mathbf{j}'_{QK_s}$ , the information matrix in (3.11) can be written as:

$$\begin{aligned}
\mathbf{X}' \mathbf{W} \mathbf{X} &= \begin{bmatrix} \mathbf{X}'_s \mathbf{W} \mathbf{X}_s & -\mathbf{D} & \cdots & -\mathbf{D} \\ -\mathbf{D} & \mathbf{X}'_s \mathbf{W} \mathbf{X}_s & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\mathbf{D} \\ -\mathbf{D} & \cdots & -\mathbf{D} & \mathbf{X}'_s \mathbf{W} \mathbf{X}_s \end{bmatrix} \\
&= \mathbf{I}_p \otimes \mathbf{X}'_s \mathbf{X}_s - \mathbf{J}_p \otimes \mathbf{D}. \tag{3.12}
\end{aligned}$$

Using Lemma 3.1.4 stated above and substitute in (3.12):

$$\begin{aligned}
(\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} &= (\mathbf{I}_p \otimes \mathbf{X}'_s \mathbf{X}_s - \mathbf{J}_p \otimes m^2 (T)^{-1} \mathbf{J}_{QK_s})^{-1} \\
&= (\mathbf{I}_p \otimes \mathbf{X}'_s \mathbf{X}_s)^{-1} - \frac{(-m^2 T^{-1}) (\mathbf{I}_p \otimes \mathbf{X}'_s \mathbf{X}_s)^{-1} (\mathbf{J}_p \otimes \mathbf{J}_{QK_s}) (\mathbf{I}_p \otimes \mathbf{X}'_s \mathbf{X}_s)^{-1}}{1 + (-m^2 T^{-1}) (\mathbf{j}'_p \otimes \mathbf{j}'_{QK_s}) (\mathbf{I}_p \otimes \mathbf{X}'_s \mathbf{X}_s)^{-1} (\mathbf{j}_p \otimes \mathbf{j}_{QK_s})} \\
&= \mathbf{I}_p \otimes (\mathbf{X}'_s \mathbf{X}_s)^{-1} + \frac{m^2 T^{-1} [\mathbf{J}_p \otimes (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{J}_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1}]}{1 - m^2 T^{-1} [\mathbf{p} \otimes \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}]}.
\end{aligned}$$

In order to check the structure of  $(\mathbf{X}'_s \mathbf{W} \mathbf{X}_s)^{-1}$ , we have:

$$\begin{aligned} (\mathbf{X}'_s \mathbf{W} \mathbf{X}_s)^{-1} &= (\mathbf{X}'_s \mathbf{X}_s - (T)^{-1} \mathbf{X}'_s \mathbf{J}_T \mathbf{X}_s)^{-1} \\ &= (\mathbf{X}'_s \mathbf{X}_s)^{-1} + \frac{m^2 T^{-1} [(\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{J}_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1}]}{1 - m^2 T^{-1} [\mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}]} \end{aligned}$$

For designs that make  $\mathbf{h}$  estimable, we have:

$$\begin{aligned} \text{tr}((\mathbf{X}' \mathbf{W} \mathbf{X})^{-1}) &= \text{tr}(\mathbf{I}_p) \text{tr}(\mathbf{X}'_s \mathbf{X}_s)^{-1} + \frac{m^2 T^{-1} \text{tr}(\mathbf{J}_p) \text{tr}((\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{J}_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1})}{1 - m^2 T^{-1} [p \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}]} \\ &= p \text{tr}(\mathbf{X}'_s \mathbf{X}_s)^{-1} + \frac{m^2 T^{-1} [p \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-2} \mathbf{j}_{QK_s}]}{1 - m^2 T^{-1} [p \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}]} \end{aligned} \quad (3.13)$$

And similarly for designs that make  $\mathbf{h}_1$ , we have:

$$\text{tr}((\mathbf{X}'_s \mathbf{W} \mathbf{X}_s)^{-1}) = \text{tr}(\mathbf{X}'_s \mathbf{X}_s)^{-1} + \frac{m^2 T^{-1} [\mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-2} \mathbf{j}_{QK_s}]}{1 - m^2 T^{-1} [\mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}]} \quad (3.14)$$

For  $D$ -optimality, we substitute Lemma 3.1.5 in (3.12):

$$\begin{aligned} \det(\mathbf{X}' \mathbf{W} \mathbf{X}) &= \det(\mathbf{I}_p \otimes \mathbf{X}'_s \mathbf{X}_s - \mathbf{J}_p \otimes m^2 (T)^{-1} \mathbf{J}_{QK_s}) \\ &= [1 + (-m^2 T^{-1}) (\mathbf{j}'_s \otimes \mathbf{j}'_{QK_s}) (\mathbf{I}_p \otimes (\mathbf{X}'_s \mathbf{X}_s)^{-1}) (\mathbf{j}_s \otimes \mathbf{j}_{QK_s})] \det(\mathbf{I}_p \otimes (\mathbf{X}'_s \mathbf{X}_s)) \\ &= [1 - m^2 T^{-1} p \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}] [\det(\mathbf{X}'_s \mathbf{X}_s)]^p \\ &= [1 - p m^2 T^{-1} \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}] [\det(\mathbf{X}'_s \mathbf{X}_s)]^p \end{aligned}$$

$$\log \det(\mathbf{X}' \mathbf{W} \mathbf{X}) = p \log([1 - p m^2 T^{-1} \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}] \det(\mathbf{X}'_s \mathbf{X}_s)). \quad (3.15)$$

And similarly for  $\det(\mathbf{X}'_s \mathbf{W} \mathbf{X}_s)$ :

$$\begin{aligned} \det(\mathbf{X}'_s \mathbf{W} \mathbf{X}_s) &= \det(\mathbf{X}'_s \mathbf{X}_s - (T)^{-1} \mathbf{X}'_s \mathbf{J}_T \mathbf{X}_s)^{-1} \\ \log \det(\mathbf{X}'_s \mathbf{W} \mathbf{X}_s) &= \log \det(\mathbf{X}'_s \mathbf{X}_s - (T)^{-1} \mathbf{X}'_s \mathbf{J}_T \mathbf{X}_s)^{-1} \\ &= [1 - m^2 T^{-1} \mathbf{j}'_{QK_s} (\mathbf{X}'_s \mathbf{X}_s)^{-1} \mathbf{j}_{QK_s}] \log \det(\mathbf{X}'_s \mathbf{X}_s). \end{aligned} \quad (3.16)$$

We simultaneously substitute (3.13) and (3.14) in (2.5) for  $A$ -optimality:

$$\begin{aligned}
\phi(\mathbf{d}|\mathbf{X}) &= \mathcal{R} / \text{tr}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \\
&= \mathcal{R} / p(\text{tr}(\mathbf{X}'_s\mathbf{X}_s)^{-1} + \frac{m^2T^{-1}[\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-2}\mathbf{j}_{QK_s}]}{1 - m^2T^{-1}[p\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{j}_{QK_s}]}) \\
&= \mathcal{R}_s / (\text{tr}(\mathbf{X}'_s\mathbf{X}_s)^{-1} + \frac{m^2T^{-1}[\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-2}\mathbf{j}_{QK_s}]}{1 - m^2T^{-1}[p\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{j}_{QK_s}]}) \\
\phi(\mathbf{d}|\mathbf{X}_s) &= \mathcal{R}_s / (\text{tr}(\mathbf{X}'_s\mathbf{X}_s)^{-1} + \frac{m^2T^{-1}[\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-2}\mathbf{j}_{QK_s}]}{1 - m^2T^{-1}[\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{j}_{QK_s}]}) .
\end{aligned}$$

Also, we simultaneously substitute (3.15) and (3.16) in (2.5) for  $D$ -optimality:

$$\begin{aligned}
\phi(\mathbf{d}|\mathbf{X}) &= \frac{1}{\mathcal{R}} \log \det(\mathbf{X}'\mathbf{W}\mathbf{X}) \\
&= \frac{p}{\mathcal{R}} [1 - pm^2T^{-1}\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{j}_{QK_s}] \log \det(\mathbf{X}'_s\mathbf{X}_s) \\
&= \frac{1}{\mathcal{R}_s} [1 - pm^2T^{-1}\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{j}_{QK_s}] \log \det(\mathbf{X}'_s\mathbf{X}_s) . \\
\phi(\mathbf{d}|\mathbf{X}_s) &= \frac{1}{\mathcal{R}_s} [1 - m^2T^{-1}\mathbf{j}'_{QK_s}(\mathbf{X}'_s\mathbf{X}_s)^{-1}\mathbf{j}_{QK_s}] \log \det(\mathbf{X}'_s\mathbf{X}_s) .
\end{aligned}$$

### 3.2 Stimulation Results with $\Delta T = \tau_{TR}$

In this section, we apply our proposed approach described previously in Step X-1 to X-3 to optimize the surrogate criterion  $\phi(\mathbf{d}|\mathbf{X}_s)$  of  $\phi(\mathbf{d}|\mathbf{X})$  by considering some realistic situations. Note that directly optimizing the latter criterion is normally time consuming. As demonstrated in the previous section,  $\phi(\mathbf{d}|\mathbf{X}_s)$  is a good surrogate for  $\phi(\mathbf{d}|\mathbf{X})$  under simplified models. Moving away from these models, we demonstrate the usefulness of the surrogate criterion by some numerical studies.

In this section, we focus on the estimation of the HRFs with  $\tau_{TR} = \Delta T$ . In the first case study, we consider two  $(Q, N)$  combinations:  $(1, 255)$ , and  $(2, 242)$  with  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds. We study the effect of different subsampling rates  $p$ . In particular, the value of  $p$  will be set to the positive divisors of  $m_{ISI}$ ; in addition, we will also consider the uniform subsampling of the columns of  $\mathbf{X}$ . In the second case study, we consider the  $(Q, N)$  combinations studied by Kao *et al.* (2009):  $(Q, N) = (3, 255), (4, 624),$  and  $(6, 342)$  with  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds and we set  $p$  to  $m_{ISI}$ . For both cases, we set  $\tau_{TR} = \Delta T = 0.1$  seconds; this means that the response sampling rate is 10 Hz. At an activated brain region, each stimulus is assumed to evoke an HRF of  $\tau_{dur} = 32$  seconds. With  $\Delta T = 0.1$  seconds, the HRF parameter vector  $\mathbf{h}_q$  for the  $q^{\text{th}}$ -type stimulus has a length of  $321 (= \lfloor \tau_{dur} / \Delta T \rfloor + 1)$ , and  $\mathbf{h} = (\mathbf{h}'_1, \dots, \mathbf{h}'_Q)'$  of Model (2.2) has  $321Q$  elements. The drift of the time series, i.e.  $\mathbf{S}\boldsymbol{\gamma}$  is assumed to be a second-order Legendre polynomial (Liu and Frank, 2004) and, for simplicity, the error term is assumed to follow a stationary AR(1) process with an autocorrelation coefficient of 0.3. We now present the simulation results under  $A$ -optimality in the next subsection. All the computations in the following subsection are conducted on a desktop computer of a 4.00 GHz Intel Core i7-4790k quad-core processor with 16GB RAM.

### 3.2.1 Estimation of the HRF Under the A-optimality Criterion

Table 3.1 presents the results for the first case with one stimulus type with different  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds under the  $A$ -optimality criterion. It also provides the CPU times for obtaining the designs with and without the subsampling approach. For Method I when  $p = 1$ , we adopt the genetic algorithm of Kao *et al.* (2009) to obtain an optimal design  $\mathbf{d}^*$  that maximizes the  $A$ -optimality criterion  $\phi(\mathbf{d}^*|\mathbf{X}) = \mathcal{R}/\text{tr}(\mathbf{M}^{-1})$ , where  $\mathbf{M}$  is the information matrix of  $\mathbf{h}$  defined in (2.4) using the full matrix  $\mathbf{X}$  and  $\mathcal{R}$  is set to  $321Q$  for Model (2.2). For Method II when  $p$  is the divisors of  $m_{ISI}$ , we again adopt the genetic algorithm of Kao *et al.* (2009) to obtain optimal designs  $\mathbf{d}_s^*$  that maximize the  $A$ -optimality criterion  $\phi(\mathbf{d}_s^*|\mathbf{X}_s) = \mathcal{R}_s/\text{tr}(\mathbf{M}_s^{-1})$ , where  $\mathbf{M}_s$  is the information matrix defined in (2.4) using the subsampled matrix  $\mathbf{X}_s$  and  $\mathcal{R}_s$  is set to  $K_sQ = (\lfloor \frac{K-1}{p} \rfloor + 1)Q$  for Model (3.2). Then, we evaluate the obtained  $\mathbf{d}_s^*$  using  $\phi(\mathbf{d}_s^*|\mathbf{X})$ . The good performance of  $\mathbf{d}_s^*$  under  $\phi(\mathbf{d}_s^*|\mathbf{X})$  is consistently demonstrated when  $p$  is  $m_{ISI}$  or its divisors. As presented in the last two columns of Table 3.1, all the  $\mathbf{d}_s^*$ 's achieve at least 99.5% efficiency of that  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$ . The CPU time needed for obtaining  $\mathbf{d}_s^*$  under  $\phi(\mathbf{d}_s^*|\mathbf{X}_s)$  is significantly less than obtaining  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$ . In addition, Figure 3.1 shows that all the optimal designs that we obtained have high relative efficiencies since the ratios  $\phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  are all close to 1 for any design  $\mathbf{d}_s^*$ . Furthermore, the reductions in the CPU times are almost 50% when  $p = 2$  and about 80% when  $p = 20, 30,$  or  $40$ .

Table 3.2 presents the results for the first case with two stimulus types with different  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds under the  $A$ -optimality criterion. The designs  $\mathbf{d}^*$ 's and  $\mathbf{d}_s^*$ 's are obtained by optimizing  $\phi(\mathbf{d}^*|\mathbf{X})$  and  $\phi(\mathbf{d}_s^*|\mathbf{X}_s)$  respectively and we re-evaluate the performance of  $\mathbf{d}_s^*$  using  $\phi(\mathbf{d}_s^*|\mathbf{X})$ . For  $Q = 2$ , the size of the full information matrix becomes quite large with  $\mathcal{R} = 642$ . Using Method I for which

Table 3.1:  $A$ -optimality for estimation when  $Q = 1$  and  $p =$  positive divisors of  $m_{ISI}$ .

$\tau_{ISI}$	Method	$p$	$K_s$	$\phi(\mathbf{d}_s^* \mathbf{X}_s)$	CPU (mins)	# Iteration	$\phi(\mathbf{d}^* \mathbf{X})$	$\phi(\mathbf{d}_s^* \mathbf{X})$
2	I	1	321	-	31.37	601	62.9524	-
		2	161	77.0052	28.12	1001	-	62.7911
	II	4	81	76.9677	10.16	601	-	62.8413
		5	65	76.8982	14.64	1001	-	62.8991
		10	33	76.6638	6.70	601	-	62.8126
		20	17	76.5416	9.80	1001	-	62.8662
3	I	1	321	-	43.41	401	66.3760	-
		2	161	81.7802	82.61	1401	-	66.0930
	II	3	107	81.5915	35.81	801	-	66.3710
		5	65	81.6226	31.52	1001	-	66.4020
		6	54	81.7675	17.13	601	-	66.3345
		10	33	81.3937	13.43	601	-	66.3727
		15	22	81.4009	24.80	1201	-	66.2364
		30	11	81.4616	15.38	801	-	66.3737
4	I	1	321	-	111.76	601	70.0395	-
		2	161	86.5834	40.62	401	-	69.8069
	II	4	81	86.4405	49.25	801	-	69.7370
		5	65	86.1111	21.70	601	-	69.9362
		8	41	86.0752	60.38	1401	-	69.9214
		10	33	86.0041	23.79	601	-	69.9608
		20	17	85.5231	26.62	801	-	69.9277
		40	9	84.6738	24.66	801	-	69.9970
5	I	1	321	-	169.55	601	73.7280	-
		2	161	90.8793	124.24	801	-	73.6164
	II	5	65	90.5200	48.72	601	-	73.6687
		10	33	90.4331	35.29	601	-	73.6009
		25	13	90.7760	37.60	801	-	73.5775
		50	7	89.7312	28.39	601	-	73.7613

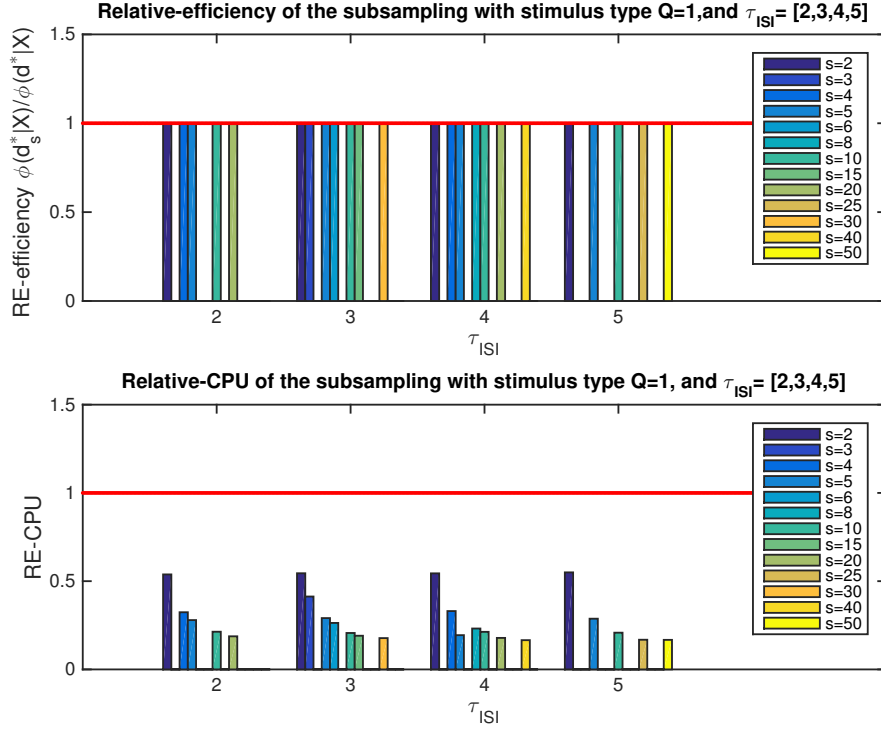


Figure 3.1: Relative design efficiencies and relative CPU times with the  $A$ -optimality criterion for estimation when  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds: this plot provides the relative efficiency of  $\mathbf{d}_s^*$  to  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$  and relative CPU times of different subsampling rates  $p$  with  $Q = 1$ .

the full information matrix (or equivalently, the full design matrix  $\mathbf{X}$ ) is used, it becomes computationally very difficult as indicated in the table when  $p = 1$ . With Method II when the subsampled design matrix  $\mathbf{X}_s(p = m_{ISI}$  or its divisors), obtaining optimal design  $\mathbf{d}_s^*$  requires about half of the computing time needed for achieving the optimal design  $\mathbf{d}^*$  by Method I. Nevertheless, all the obtained  $\mathbf{d}_s^*$ 's still attain very high efficiencies under  $\phi(\mathbf{d}_s^*|\mathbf{X})$  and the good performance of  $\mathbf{d}_s^*$  under  $\phi(\mathbf{d}_s^*|\mathbf{X})$  is consistently demonstrated when  $p$  is  $m_{ISI}$  or its divisors in all scenarios ( $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds). As presented in the last two columns, all the  $\mathbf{d}_s^*$ 's achieve at least

99.6% efficiency of that of  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$ . These results suggest that obtaining designs optimizing  $\phi(\mathbf{d}|\mathbf{X}_s)$  requires less computing time than optimizing  $\phi(\mathbf{d}|\mathbf{X})$ . Designs obtained by maximizing the former criterion have a similar performance to those obtained with the latter criterion. Additionally, Figure 3.2 shows that all the optimal designs that were obtained in Table 3.2 have high relative efficiencies since the ratios  $\phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  are all close to 1 for any design  $\mathbf{d}_s^*$ . Moreover, the CPU times are greatly reduced when the number of stimulus types,  $Q$ , increases. In particular, the reductions in the CPU times are greater than that when  $Q = 1$ , i.e. it's more than 50% when  $p = 2$  and more than 80% when  $p = 20, 30$ , or 40.

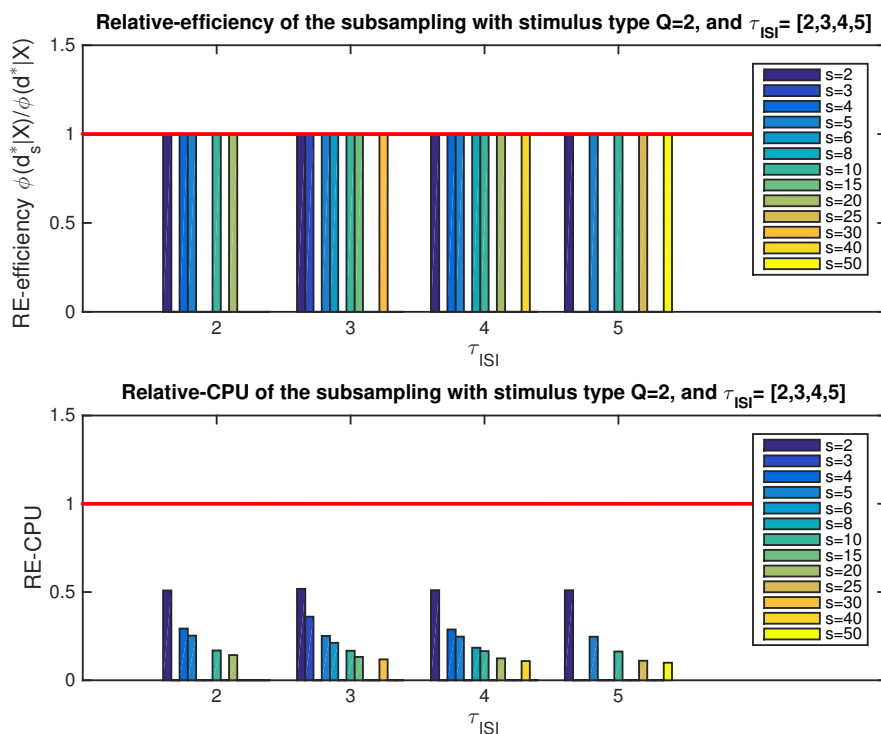


Figure 3.2: Relative design efficiencies and relative CPU times with the  $A$ -optimality criterion for estimation when  $\tau_{ISI} = 2, 3, 4$ , and 5 seconds: this plot provides the relative efficiency under  $\phi = \phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  and relative CPU times of different subsampling rates  $p$  with  $Q = 2$ .



Table 3.2:  $A$ -optimality for estimation when  $Q = 2$  and  $p =$  positive divisors of  $m_{ISI}$ .

$\tau_{ISI}$	Method	$p$	$K_s$	$\phi(\mathbf{d}_s^* \mathbf{X}_s)$	CPU (mins)	# Iteration	$\phi(\mathbf{d}^* \mathbf{X})$	$\phi(\mathbf{d}_s^* \mathbf{X})$
2	I	1	321	-	100.75	1001	39.7195	-
		2	161	47.9008	41.01	801	-	39.5845
		4	81	48.0507	23.61	801	-	39.7022
	II	5	65	47.8828	20.37	801	-	39.5852
		10	33	47.8193	17.02	1001	-	39.6115
		20	17	47.7517	14.37	1001	-	39.6808
3	I	1	321	-	161.36	801	41.3618	-
		2	161	50.2944	62.71	601	-	41.4221
		3	107	50.1728	43.63	601	-	41.3561
		5	65	50.1845	30.43	601	-	41.4332
	II	6	54	50.2227	42.82	1001	-	41.4309
		10	33	50.1782	27.05	801	-	41.3439
		15	22	50.2161	16.07	601	-	41.4362
		30	11	50.1510	19.09	801	-	41.4351
4	I	1	321	-	274.60	801	43.0156	-
		2	161	52.5907	175.15	1001	-	42.9608
		4	81	52.5386	78.98	801	-	43.0462
		5	65	52.3913	118.81	1401	-	43.0931
	II	8	41	52.3451	50.61	801	-	43.0463
		10	33	52.2746	68.02	1201	-	43.0578
		20	17	51.9440	42.61	1001	-	43.0486
		40	9	51.6380	22.40	601	-	43.0479
5	I	1	321	-	413.35	801	44.6604	-
		2	161	54.4765	315.63	1201	-	44.7003
		5	65	54.5324	178.35	1401	-	44.7519
	II	10	33	54.3475	101.01	1201	-	44.6942
		25	13	54.2885	57.16	1001	-	44.6511
		50	7	53.9324	51.37	1001	-	44.7206

Table 3.3 provides the results of the relative efficiency under  $\phi = \phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  and relative CPU time for the second case with  $Q = 1, 2, 3, 4,$  and  $6$  and  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds under the  $A$ -optimality criterion. For  $Q = 3, 4,$  and  $6,$  the size of the full information matrix  $\mathbf{M}$  become quite large with  $\mathcal{R} = 963, 1284,$  and  $1926,$  respectively. With Method I for which the full design matrix  $\mathbf{X}$  is used, it becomes computationally very difficult. With Method II when the subsampled design matrix  $\mathbf{X}_s(p = m_{ISI}),$  obtaining optimal designs  $\mathbf{d}_s^*$  require no more than 20% of the computing time needed for achieving the  $\mathbf{d}^*$  in Method I. Furthermore, all the obtained  $\mathbf{d}_s^*$ 's consistently demonstrate very high efficiencies under  $\phi(\mathbf{d}_s^*|\mathbf{X})$  when  $p$  is  $m_{ISI}$  in all cases ( $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds). Our experience suggests that some other subsampling rates when replacing  $p$  with its divisor can also lead to a good surrogate criterion for  $\phi(\mathbf{d}^*|\mathbf{X}).$  However, the aforementioned procedure when  $p = m_{ISI}$  gives a greater reduction in the size of the information matrix and is thus recommended. As presented in the table, all the  $\mathbf{d}_s^*$ 's achieve at least 99.7% efficiency of that of  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X}).$  Also, these results suggest that the computing time needed for obtaining designs with  $\phi(\mathbf{d}_s^*|\mathbf{X})$  are significantly less than that of  $\phi(\mathbf{d}^*|\mathbf{X}).$

In Table 3.4, we compare the performance of the two methods on obtaining optimal designs,  $\mathbf{d}_s^*$ 's for  $(Q, N) = (1, 255)$  and  $(2, 242)$  when  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds. When comparing our proposed approach (Method II) versus the uniform subsampling (Method III), our proposed approach is more efficient. For evaluating the performance of the obtained  $\mathbf{d}_s^*$ 's from Method II and III, the number of the parameters of interest  $QK_s$  is set to the same and the uniform subsampling is randomly generated five times. We note that the total time needed for obtaining the optimal designs for Method III is nearly similar to Method II presented in Table 3.1, so it's not reported again here. As presented in Table 3.4, the optimal designs  $\mathbf{d}_s^*$ 's in Method III compared to  $\mathbf{d}_s^*$ 's obtained from Method II attain a relative high efficiency when  $K_s = 161, 107$  or  $81$

Table 3.3: Relative efficiency and CPU for estimation when  $Q = 1, 2, 3, 4,$  and  $6$  and  $p = m_{ISI}$ : the designs are obtained by optimizing  $\phi(\mathbf{d}^*|\mathbf{X})$  to obtain  $\mathbf{d}^*$  and  $\phi(\mathbf{d}_s^*|\mathbf{X})$  to obtain  $\mathbf{d}_s^*$ , and re-evaluate  $\mathbf{d}_s^*$  using  $\mathbf{X}$  where  $\phi = \phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$ .

$Q$	$\tau_{ISI} = 2$		$\tau_{ISI} = 3$		$\tau_{ISI} = 4$		$\tau_{ISI} = 5$	
	$\phi$	CPU	$\phi$	CPU	$\phi$	CPU	$\phi$	CPU
1 <sup>1</sup>	0.9986	0.1876	1.0000	0.1774	0.9994	0.1656	1.0005	0.1675
2 <sup>2</sup>	0.9990	0.1426	1.0018	0.1183	1.0007	0.1087	1.0013	0.0994
3	1.0001	0.1160	1.0047	0.0913	1.0022	0.0868	0.9997	0.0778
4	1.0004	0.0931	1.0022	0.0716	0.9984	0.0645	0.9998	0.0585
6	1.0016	0.0820	0.9995	0.0690	1.001	0.0618	1.0041	0.0547

<sup>1</sup> The result is obtained from Table 3.1 when  $p = m_{ISI}$ .

<sup>2</sup> The result is obtained from Table 3.2 when  $p = m_{ISI}$ .

which corresponds to  $1/2, 1/3$  or  $1/4$  of  $K$ , respectively and do not outperform the optimal designs obtained from Method II when  $K_s$  is less than 81.

Table 3.5 presents the results for cases with three stimulus types under the  $A$ -optimality criterion. In particular, we compare the optimal design of our proposed method with some traditional designs that are widely used in functional brain imaging studies when  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds. These traditional designs include those obtained by Kao *et al.* (2009) with  $(\tau_{ISI}, \tau_{TR}) = (2, 2), (3, 3), (4, 4),$  and  $(5, 5)$ , an m-sequence-based design, and ten randomly generated designs. To generate m-sequence-based designs, see Kao and Stufken (2015) for external review and more details. The efficiency ratio  $\phi(\mathbf{d}^*|\mathbf{X})/\phi(\mathbf{d}_s^*|\mathbf{X})$  of these traditional fMRI designs to our  $\mathbf{d}_s^*$  are shown in Table 3.5. Kao's optimal designs and m-sequence-based designs yield a relative high efficiency in very few cases. However, they do not perform as well as

Table 3.4:  $A$ -optimality for estimation for the two methods when  $Q = 1$  and  $2$ ,  $\tau_{ISI} = 2, 3, 4$ , and  $5$  seconds, and  $p =$  positive divisors of  $m_{ISI}$ .

		$Q = 1$			$Q = 2$		
$\tau_{ISI}$	$K_s$	Method II	Method III	RE- $\phi$	Method II	Method III	RE- $\phi$
2	161	62.7911	62.6548-62.8683	0.9978-1.0012	39.5845	39.6143-39.6667	1.0008-1.0021
	81	62.8413	60.9113-62.1017	0.9693-0.9882	39.7022	38.7930-39.2498	0.9771-0.9886
	65	62.8991	58.2924-61.0203	0.9268-0.9701	39.5852	37.0867-39.1572	0.9369-0.9892
	33	62.8126	47.1060-49.2941	0.7499-0.7848	39.6115	22.7042-36.4829	0.5732-0.9210
	17	62.8662	17.8622-29.8622	0.2841-0.4750	39.6808	18.7155-25.0344	0.4717-0.6309
3	161	66.0930	65.7819-66.3306	0.9953-1.0036	41.4221	41.1582-41.4021	0.9936-0.9995
	107	66.3710	64.3253-65.5023	0.9692-0.9869	41.3561	40.8846-41.2906	0.9886-0.9984
	65	66.4020	59.8574-64.1348	0.9014-0.9659	41.4332	37.6830-40.3402	0.9095-0.9736
	54	66.3345	43.4007-60.3234	0.6543-0.9094	41.4309	26.7554-38.6573	0.6458-0.9331
	33	66.3727	25.6785-46.2155	0.3869-0.6963	41.3439	24.5349-35.7459	0.5934-0.8646
	22	66.2364	21.3279-39.6732	0.3220-0.5990	41.4362	18.0055-30.3853	0.4345-0.7333
	11	66.3737	23.8907-35.2724	0.3599-0.5314	41.4351	24.7551-29.0121	0.5974-0.7002
4	161	69.8069	69.1700-69.5758	0.9909-0.9967	42.9608	42.8195-42.9743	0.9967-1.0003
	81	69.7370	66.0326-67.8153	0.9469-0.9724	43.0462	39.8435-42.3704	0.9256-0.9843
	65	69.9362	58.1888-65.9958	0.8320-0.9437	43.0931	38.3214-41.9116	0.8893-0.9726
	41	69.9214	33.7177-59.2346	0.4822-0.8472	43.0463	32.9234-38.8192	0.7648-0.9018
	33	69.9608	49.3798-62.2369	0.7058-0.8896	43.0578	21.1133-37.3173	0.4903-0.8667
	17	69.9277	15.1377-40.8665	0.2165-0.5844	43.0486	24.3041-31.5537	0.5646-0.7330
	9	69.9970	0-46.3227	0-0.6618	43.0479	0-30.0134	0-0.6972
5	161	73.6164	72.2788-72.9827	0.9818-0.9914	44.7003	44.1978-44.5887	0.9888-0.9975
	65	73.6687	63.2600-69.9918	0.8587-0.9501	44.7519	40.9784-42.3776	0.9157-0.9469
	33	73.6009	53.3151-64.5504	0.7244-0.8770	44.6942	31.6534-41.1509	0.7082-0.9207
	13	73.5775	0-41.8363	0-0.5686	44.6511	0-35.42594	0-0.7934
	7	73.7613	0-43.4324	0-0.5888	44.7206	0-0.5745	0-0.0128

the  $\mathbf{d}_s^*$ 's obtained from our proposed method. We randomly generate ten design sequences for each of the three scenarios that we consider. The range of the relative efficiency of these random designs are also reported in Table 3.5. As a result, the good performance of  $\mathbf{d}_s^*$ 's are consistently demonstrated in all the cases that we studied. It is noteworthy that no design  $\mathbf{d}^*$  has a higher  $\phi(\mathbf{d}^*|\mathbf{X})$  than  $\phi(\mathbf{d}_s^*|\mathbf{X})$  among the designs  $\mathbf{d}_s^*$  we considered since the relative efficiencies are all less than 1.

Table 3.5: The optimal designs  $\mathbf{d}_s^*$ 's by Method II versus traditional designs when  $Q = 1, 2,$  and  $3$ : these traditional designs are evaluated by  $\phi(\mathbf{d}^*|\mathbf{X})$ .

$Q$	$\tau_{ISI}$	$\mathbf{d}_s^*$	Kao's 2009	RE- $\phi$	m-sequence	RE- $\phi$	ten-random	RE- $\phi$
1	2	62.8662	57.7170	0.9181	61.0835	0.9716	54.7318-59.3504	0.8706-0.9441
	3	66.3737	59.8481	0.9017	63.2743	0.9533	56.0296-62.9419	0.8442-0.9483
	4	69.9970	61.5093	0.8787	65.1409	0.9306	61.7131-64.1944	0.8817-0.9171
	5	73.7613	62.7391	0.8506	66.8517	0.9063	63.2341-66.4731	0.8573-0.9012
2	2	39.6808	37.1531	0.9363	36.5342	0.9207	31.0420-34.1798	0.7823-0.8614
	3	41.4351	38.4164	0.9271	37.9727	0.9164	32.4428-37.7605	0.7830-0.9113
	4	43.0479	39.1643	0.9098	39.1621	0.9097	34.8907-39.5613	0.8105-0.9190
	5	44.7206	39.7066	0.8879	40.2614	0.9003	37.2482-40.7129	0.8329-0.9104
3	2	31.8883	29.8278	0.9354	29.8491	0.9361	21.9519-25.8041	0.6884-0.8092
	3	33.3286	30.7399	0.8906	30.7619	0.9230	25.8754-28.6065	0.7764-0.8583
	4	34.5154	31.7774	0.9207	31.4990	0.9126	26.1701-29.9911	0.7582-0.8689
	5	35.5721	32.2323	0.9061	32.2191	0.9057	28.5772-32.1269	0.8034-0.9031

Table 3.6 compares the optimal design of our proposed method with Kao *et al.* (2009) and an m-sequence-based design designs under the setting when  $(\tau_{ISI}, \tau_{TR}) = (2, 2), (3, 3), (4, 4),$  and  $(5, 5)$ . The reason for making this comparison is that we know that Kao's designs and an m-sequence-based designs perform very well when  $\tau_{ISI} = \tau_{TR}$ . The efficiency ratio  $\phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  of our  $\mathbf{d}_s^*$  compared to these

traditional fMRI designs are shown in Table 3.6. The first relative efficiency is for our design compared to Kao *et al.* (2009). The second relative efficiency is for our design compared to m-sequence-based designs. Even though there are few cases where the relative efficiency are high, the  $\mathbf{d}_s^*$ 's obtained from our proposed method mostly do not perform these designs under the traditional setting when  $\tau_{ISI} = \tau_{TR}$ .

Table 3.6: The optimal designs  $\mathbf{d}_s^*$ 's by Method II evaluated at  $(\tau_{ISI}, \tau_{TR}) = (2, 2), (3, 3), (4, 4),$  and  $(5, 5)$  versus traditional designs when  $Q = 1, 2,$  and  $3$ .

$Q$	$\tau_{ISI}$	$\mathbf{d}_s^*$	Kao's 2009	RE- $\phi$	m-sequence	RE- $\phi$
1	2	48.0511	60.6687	0.7920	58.6349	0.8195
	3	42.9352	61.1009	0.7027	58.9738	0.7280
	4	39.3274	61.2904	0.6417	59.2056	0.6643
	5	36.1891	61.4694	0.5887	59.6180	0.6070
2	2	34.9701	39.2256	0.8915	35.4570	0.9863
	3	33.1534	39.6783	0.8356	36.4108	0.9105
	4	30.9641	39.8243	0.7775	36.7738	0.8420
	5	27.3491	39.9828	0.6840	37.2080	0.7350
3	2	28.9458	31.7754	0.9109	29.1179	0.9941
	3	27.4155	32.4720	0.8443	29.5276	0.9285
	4	26.4010	32.6280	0.8092	29.6551	0.8903
	5	25.7343	32.8082	0.7844	29.8194	0.8630

Figures 3.3 and 3.4 present these designs for  $Q = 1$  and  $Q = 2$ , respectively with  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds. Different colors indicate different stimulus types and white represents the control. These designs also show the stimulus frequency for estimation ( $\approx 1/(Q + \sqrt{Q})$ ) similar to Liu and Frank (2004). These designs look

random in appearance and do not seem to have perceivable patterns. However, none of these traditional designs that look random perform as well as  $\mathbf{d}_s^*$ .

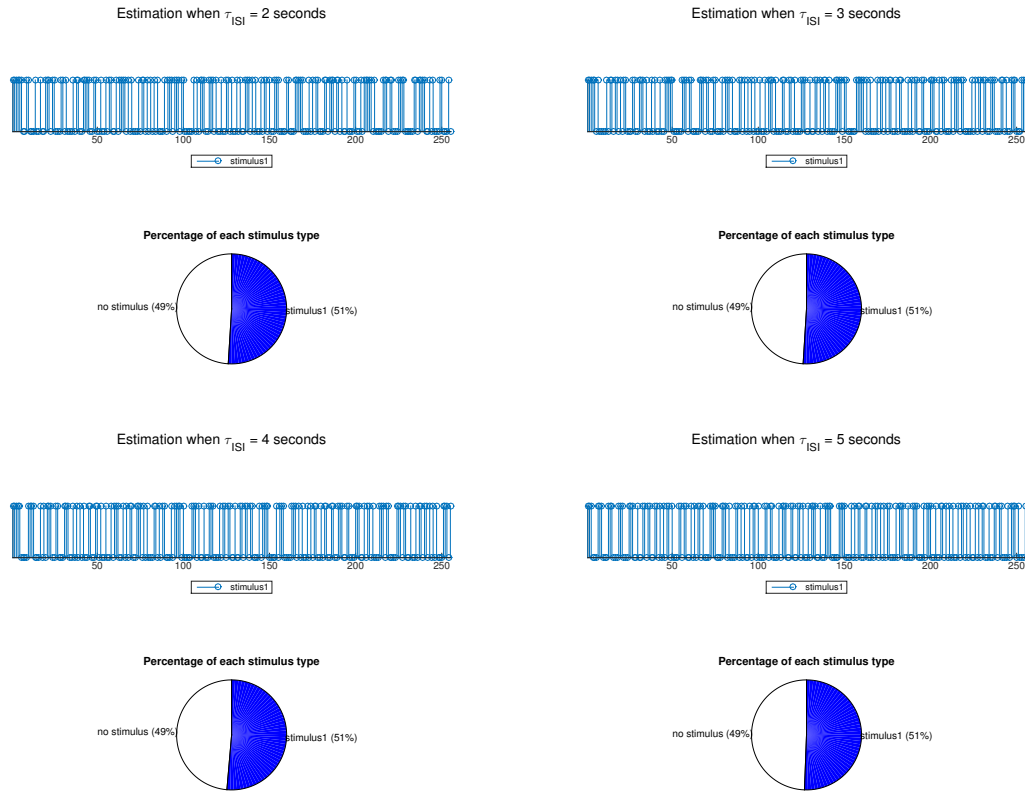


Figure 3.3: The optimal  $\mathbf{d}_s^*$  obtained from Method II when  $Q = 1$  and  $\tau_{ISI} = 2, 3, 4,$  and 5 seconds.

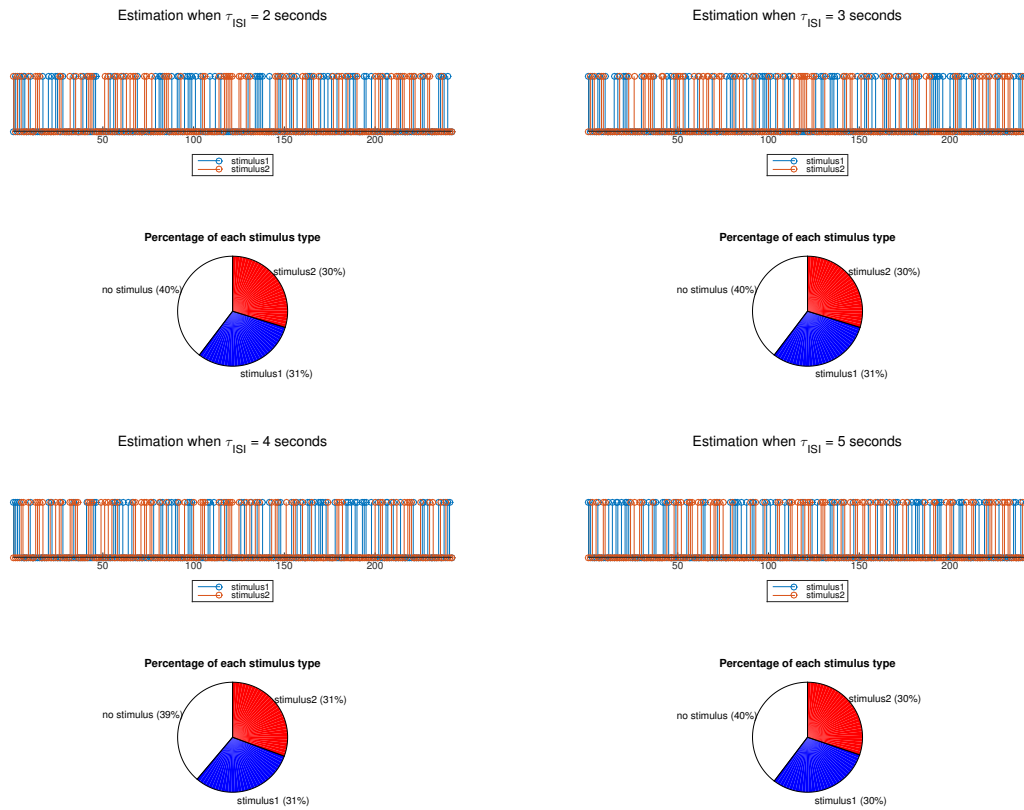


Figure 3.4: The optimal  $\mathbf{d}_s^*$  obtained from Method II when  $Q = 2$  and  $\tau_{ISI} = 2, 3, 4,$  and  $5$  seconds.



### 3.3 Stimulation Results with $\Delta T \neq \tau_{TR}$

In the previous section we studied cases where  $\Delta T = \tau_{TR}$ . Here, we apply the proposed approach again with  $\Delta T \neq \tau_{TR}$ .

In the first case study, we consider the same two  $(Q, N)$  combinations considered in the previous simulation results:  $(1, 255)$ , and  $(2, 242)$  but  $\tau_{ISI} = 2, 4,$  and  $5$  seconds. In the last results, we studied the effect of different subsampling rates  $p$  when  $p$  was set to the positive divisors of  $m_{ISI}$ . Similarly, the value of  $p$  will be set to the positive divisors of  $m_{ISI}$ , but  $p > m_{TR}$ ; in addition, we will consider the second case of subsampling of the columns of  $\mathbf{X}$  in (3.3) because  $\Delta T \neq \tau_{TR}$ . We will also consider the uniform subsampling of the columns of  $\mathbf{X}$ . In the second case study, we consider the  $(Q, N)$  combinations studied previously:  $(Q, N) = (3, 255), (4, 624),$  and  $(6, 342)$  but  $\tau_{ISI} = 2, 4,$  and  $5$  seconds and we set  $p$  to  $m_{ISI}$ . For both cases, we set  $\tau_{TR} = 0.3$  seconds and  $\Delta T = 0.1$  seconds; this means that the response sampling rate is  $3.3$  Hz. At an activated brain region, each stimulus is assumed to evoke an HRF of  $\tau_{dur} = 32$  seconds. With  $\Delta T = 0.1$  seconds, the HRF parameter vector  $\mathbf{h}_q$  for the  $q^{\text{th}}$ -type stimulus has a length of  $321$ , and  $\mathbf{h} = (\mathbf{h}'_1, \dots, \mathbf{h}'_Q)'$  of Model (2.2) has  $321Q$  elements. The drift of the time series, i.e.  $\mathbf{S}\boldsymbol{\gamma}$  is assumed to be a second-order Legendre polynomial and the error term is assumed to follow a stationary AR(1) process with an autocorrelation coefficient of  $0.3$ . We now present the simulation results under  $A$ -optimality in the next subsection. All the computations in the following subsection are conducted on a desktop computer of a  $4.00$  GHz Intel Core i7-4790k quad-core processor with  $16$ GB RAM.

### 3.3.1 Estimation of the HRF Under the A-optimality Criterion

Table 3.7 presents the results for the first case with one stimulus type with different  $\tau_{ISI} = 2, 4,$  and  $5$  seconds under the  $A$ -optimality criterion. It also provides the CPU times for obtaining the designs with and without the subsampling approach. For Method I when  $p = 1$ , we will do the same as last simulation when we start to adopt the genetic algorithm to obtain an optimal design  $\mathbf{d}^*$  that maximizes the  $A$ -optimality criterion  $\phi(\mathbf{d}^*|\mathbf{X}) = \mathcal{R}/\text{tr}(\mathbf{M}^{-1})$ , where  $\mathbf{M}$  is the information matrix of  $\mathbf{h}$  defined in (2.4) using the full matrix  $\mathbf{X}$  and  $\mathcal{R}$  is set to  $321Q$  for Model (2.2). For Method II when  $p$  is the divisors of  $m_{ISI}$  and  $m_{TR} = 3$ ; this means that we keep columns  $j, (j + p), (j + 2p), \dots, (j + \lfloor \frac{K-j}{p} \rfloor p)$  of  $\mathbf{X}_q$  for  $j = 1, 2, 3$  and leave out the other columns to form  $\mathbf{X}_{sq}$  for  $q = 1, \dots, Q$ . Also  $T$  in both  $\mathbf{X}$  and  $\mathbf{X}_s$  are  $1700, 3400, 4250$  for  $\tau_{ISI} = 2, 4,$  and  $5$ , respectively. Again, we adopt the genetic algorithm of Kao *et al.* (2009) to obtain optimal designs  $\mathbf{d}_s^*$  that maximize the  $A$ -optimality criterion  $\phi(\mathbf{d}_s^*|\mathbf{X}_s) = \mathcal{R}_s/\text{tr}(\mathbf{M}_s^{-1})$ , where  $\mathbf{M}_s$  is the information matrix defined in (2.4) using the subsampled matrix  $\mathbf{X}_s$  and  $\mathcal{R}_s$  is set to  $K_sQ = ((\lfloor \frac{K-1}{p} \rfloor + 1) + (\lfloor \frac{K-2}{p} \rfloor + 1) + (\lfloor \frac{K-3}{p} \rfloor + 1))Q$  for Model (3.2). The good performance of  $\mathbf{d}_s^*$  under  $\phi(\mathbf{d}_s^*|\mathbf{X})$  is consistently demonstrated when  $p$  is  $m_{ISI}$  or its divisors when evaluating the obtained  $\mathbf{d}_s^*$  using  $\phi(\mathbf{d}_s^*|\mathbf{X})$ . As presented in the last two columns of Table 3.7, all the  $\mathbf{d}_s^*$ 's achieve at least 99.4% efficiency of that  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$ . The CPU time needed for obtaining  $\mathbf{d}_s^*$  under  $\phi(\mathbf{d}_s^*|\mathbf{X}_s)$  is significantly less than obtaining  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$ .

Figure 3.5 verifies that all the obtained optimal designs have high relative efficiencies as the ratios  $\phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  are all close to 1 for any design  $\mathbf{d}_s^*$ . Additionally, the reductions in the CPU times are almost 50% when  $p = 2$  and about 80% when  $p = 20, 30,$  or  $40$ .

Table 3.7:  $A$ -optimality for estimation when  $Q = 1$ ,  $\tau_{ISI} = 2, 4$  and  $5$  seconds and  $p =$  positive divisors of  $m_{ISI}$  with cluster every  $m_{TR} = 3$ .

$\tau_{ISI}$	Method	$p$	$K_s$	CPU (mins)	# Iteration	$\phi(\mathbf{d}^* \mathbf{X})$	$\phi(\mathbf{d}_s^* \mathbf{X})$	RE-eff	RE-time
2	I	1	321	3.79	601	20.6952	-		
		4	241	3.16	601	-	20.7209	1.0012	0.8324
	II	5	193	1.87	401	-	20.7153	1.0010	0.7381
		10	97	2.17	601	-	20.7218	1.0013	0.5729
		20	49	2.56	801	-	20.6209	0.9964	0.5060
4	I	1	321	10.35	601	23.3181	-		
		4	241	10.97	801	-	23.1681	0.9936	0.7952
	II	5	193	7.15	601	-	23.2704	0.9980	0.6906
		8	121	7.65	801	-	23.3023	0.9993	0.5545
		10	97	8.80	1001	-	23.2749	0.9981	0.5104
	20	49	4.48	601	-	23.2435	0.9968	0.4323	
	40	25	4.13	601	-	23.2534	0.9972	0.3989	
5	I	1	321	19.69	801	24.5620	-		
		5	193	6.65	401	-	24.4266	0.9945	0.6741
	II	10	97	9.55	801	-	24.4707	0.9963	0.4850
		25	39	5.85	601	-	24.4883	0.9970	0.3963
		50	21	5.45	601	-	24.4953	0.9973	0.3685

Table 3.8 presents the results for the first case with two stimulus types with different  $\tau_{ISI} = 2, 4$ , and  $5$  seconds under the  $A$ -optimality criterion. The designs  $\mathbf{d}^*$ 's and  $\mathbf{d}_s^*$ 's in Method I and Method II are obtained by optimizing  $\phi(\mathbf{d}^*|\mathbf{X})$  and  $\phi(\mathbf{d}_s^*|\mathbf{X}_s)$ ; respectively. We re-evaluate the performance of  $\mathbf{d}_s^*$  using  $\phi(\mathbf{d}_s^*|\mathbf{X})$ . For  $Q = 2$ , the size of the full information matrix becomes quite large with  $\mathcal{R} = 642$ . Using Method I for which the full design matrix  $\mathbf{X}$  is used, it becomes computationally slight difficult as indicated in the table when  $p = 1$ . Compared with the result of

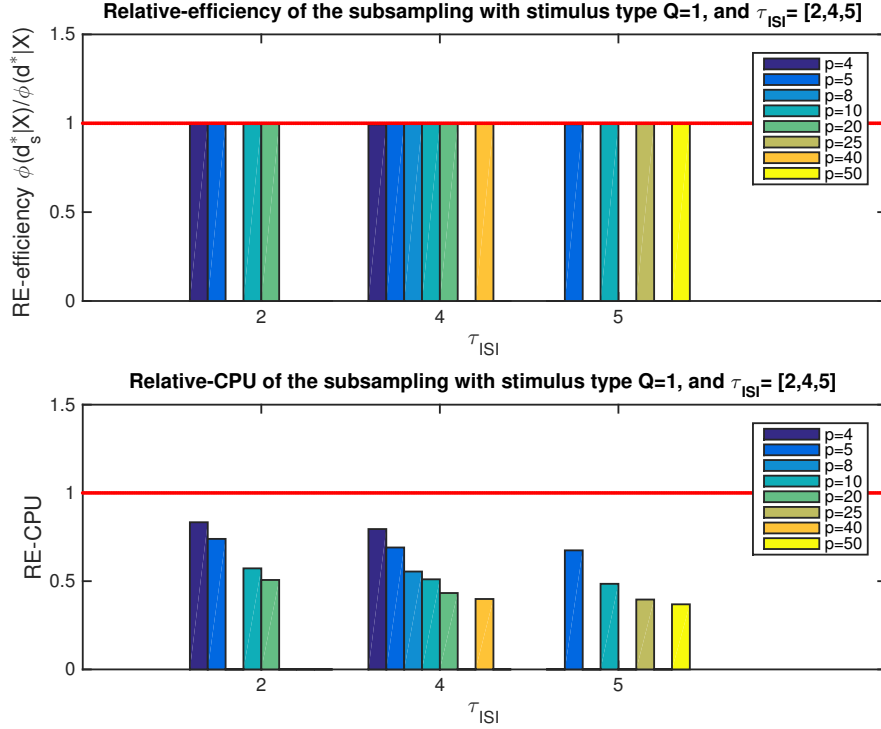


Figure 3.5: Relative design efficiencies and Relative CPU times with the  $A$ -optimality criterion for estimation when  $Q = 1$  with  $\tau_{ISI} = 2, 4,$  and  $5$  seconds: this plot provides the relative efficiency under  $\phi = \phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  and relative CPU times of different subsampling rates  $p$  with  $m_{TR} = 3$ .

Table 3.2, the CPU time is less because the length of the response vector  $T$  with  $\Delta T \neq \tau_{TR}$  here is  $1/3$  of  $T$  with  $\Delta T = \tau_{TR}$  in the previous section. With Method II when the subsampled design matrix  $\mathbf{X}_s$  ( $p = m_{ISI}$  or one of its divisors), obtaining optimal design  $\mathbf{d}_s^*$  requires 29%- 81% of the computing time needed for achieving the optimal design  $\mathbf{d}^*$  by Method I. Nevertheless, all the obtained  $\mathbf{d}_s^*$ 's still attain very high efficiencies under  $\phi(\mathbf{d}_s^*|\mathbf{X})$  as demonstrated when  $p$  is  $m_{ISI}$  or its divisors in all scenarios ( $\tau_{ISI} = 2, 4,$  and  $5$  seconds). As shown in the last two columns, all the  $\mathbf{d}_s^*$ 's achieve at least 99.81% efficiency of that of  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$ .

Table 3.8:  $A$ -optimality for estimation when  $Q = 2, 4$  and  $5$  seconds and  $p =$  positive divisors of  $m_{ISI}$  with cluster every  $m_{TR} = 3$ .

$\tau_{ISI}$	Method	$p$	$K_s$	CPU (mins)	# Iteration	$\phi(\mathbf{d}^* \mathbf{X})$	$\phi(\mathbf{d}_s^* \mathbf{X})$	RE-eff	RE-time
2	I	1	321	22.33	1601	12.4675	-		
		4	241	10.73	1001	-	12.4681	1.0000	0.7686
	II	5	193	9.14	1001	-	12.5032	1.0029	0.6545
		10	97	6.38	1001	-	12.5393	1.0058	0.4567
		20	49	5.40	1001	-	12.5688	1.0081	0.3864
4	I	1	321	38.74	1001	14.1872	-		
		4	241	50.48	1601	-	14.2077	1.0014	0.8146
	II	5	193	33.32	1201	-	14.1600	0.9981	0.7168
		8	121	36.51	1801	-	14.1904	1.0002	0.5238
		10	97	37.45	2001	-	14.1994	1.0009	0.4836
		20	49	15.82	1001	-	14.1986	1.0008	0.4082
		40	25	11.66	801	-	14.2023	1.0011	0.3760
5	I	1	321	69.45	1401	14.6978	-		
		5	193	47.47	1401	-	14.7547	1.0039	0.6836
	II	10	97	28.81	1401	-	14.7680	1.0048	0.4148
		25	39	12.47	801	-	14.7465	1.0033	0.3141
		50	21	8.52	601	-	14.7040	1.0004	0.2859

Additionally, Figure 3.6 demonstrates that all the optimal designs that were obtained in Table 3.8 have high relative efficiencies since the ratios  $\phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  are all close to 1 for any design  $\mathbf{d}_s^*$ . Furthermore, the CPU times are greatly reduced when  $p = m_{ISI}$  and the number of stimulus types,  $Q$ , increases. In particular, the reductions in the CPU times are greater than that when  $Q = 1$ , i.e. it's more than 60% when  $p = 20, 40$ , or  $50$ .

Table 3.9 provides the results of the relative efficiency  $\phi = \phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  and

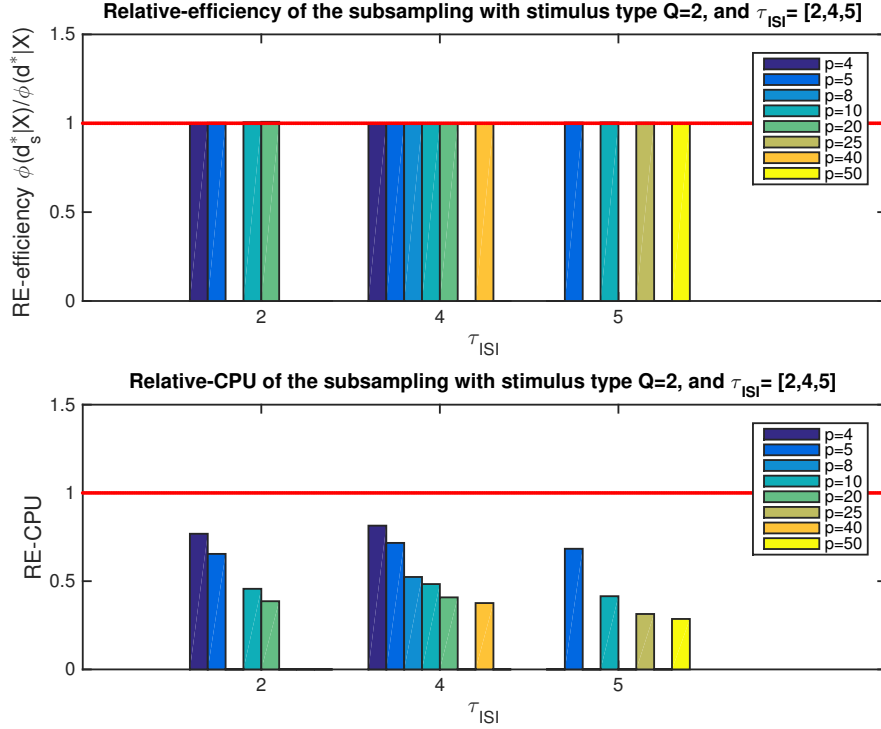


Figure 3.6: Relative design efficiencies and relative CPU times with the  $A$ -optimality criterion for estimation when  $Q = 2$  with  $\tau_{ISI} = 2, 4,$  and  $5$  seconds: this plot provides the relative efficiency under  $\phi = \phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$  and relative CPU times of different subsampling rates  $p$  with  $m_{TR} = 3$ .

relative CPU time for the second case with  $Q = 1, 2, 3, 4,$  and  $6$  and  $\tau_{ISI} = 2, 4,$  and  $5$  seconds under the  $A$ -optimality criterion. The size of the full information matrix  $\mathbf{M}$  become quite large for  $Q = 3, 4,$  and  $6$  with  $\mathcal{R} = 963, 1284,$  and  $1926$ , respectively which make the computation very difficult. With Method II when the subsampled design matrix  $\mathbf{X}_s(p = m_{ISI})$ , obtaining optimal designs  $\mathbf{d}_s^*$  require no more than 50% of the computing time needed for achieving the  $\mathbf{d}^*$  in Method I. In contrast to the result of  $\Delta T = \tau_{TR}$ , the reduction of CPU time previously is larger and this is expected because  $m_{TR} = 3$  when  $\Delta T \neq \tau_{TR}$  versus  $m_{TR} = 1$  when  $\Delta T = \tau_{TR}$ . Furthermore,

all the obtained  $\mathbf{d}_s^*$ 's consistently attain very high efficiencies at least 99% efficiency of that of  $\mathbf{d}^*$  under  $\phi(\mathbf{d}^*|\mathbf{X})$  when  $p$  is  $m_{ISI}$  in all cases ( $\tau_{ISI} = 2, 4,$  and  $5$  seconds). Our experience suggests that some other subsampling rates when replacing  $p$  with its divisor can also lead to a good surrogate criterion for  $\phi(\mathbf{d}^*|\mathbf{X})$ . However, we focus in  $p = m_{ISI}$  as it gives a greater reduction in the size of the information matrix and is thus recommended. Also, these results suggest that obtaining designs optimizing  $\phi(\mathbf{d}|\mathbf{X}_s)$  requires less computing time than optimizing  $\phi(\mathbf{d}|\mathbf{X})$ . Designs obtained by maximizing the former criterion have a similar performance to those obtained with the latter criterion.

Table 3.9: Relative efficiency and CPU for estimation when  $Q = 1, 2, 3, 4,$  and  $6$  and  $p = m_{ISI}$  with cluster of  $m_{TR} = 3$ : the designs are obtained by optimizing  $\phi(\mathbf{d}^*|\mathbf{X})$  to obtain  $\mathbf{d}^*$  and  $\phi(\mathbf{d}_s^*|\mathbf{X})$  to obtain  $\mathbf{d}_s^*$ , and re-evaluate  $\mathbf{d}_s^*$  using  $\mathbf{X}$  where  $\phi = \phi(\mathbf{d}_s^*|\mathbf{X})/\phi(\mathbf{d}^*|\mathbf{X})$ .

$Q$	$\tau_{ISI} = 2$		$\tau_{ISI} = 4$		$\tau_{ISI} = 5$	
	$\phi$	CPU	$\phi$	CPU	$\phi$	CPU
1 <sup>1</sup>	0.9964	0.5060	0.9972	0.3989	0.9973	0.3685
2 <sup>2</sup>	1.0081	0.3864	1.0011	0.3760	1.0004	0.2859
3	1.0167	0.3262	0.9908	0.2936	0.9998	0.2769
4	1.0039	0.3056	1.0003	0.2836	1.0001	0.2689
6	1.0056	0.2910	0.9906	0.2797	0.9947	0.2500

<sup>1</sup> The result is obtained from Table 3.7 when  $p = m_{ISI}$ .

<sup>2</sup> The result is obtained from Table 3.8 when  $p = m_{ISI}$ .

Figures 3.7 and 3.8 show different shades of colors that represent different stimulus types and a white bar corresponds to a zero. These designs also present the stimulus

frequency for estimation ( $\approx 1/(Q + \sqrt{Q})$ ). These designs look random in appearance and do not seem to have perceivable patterns.

Table 3.10 compares the performance of the proposed approach (Method II) and the uniform subsampling (Method III) on obtaining optimal designs,  $\mathbf{d}_s^*$ 's for  $(Q, N) = (1, 255)$  and  $(2, 242)$  when  $\tau_{ISI} = 2, 4,$  and  $5$  seconds. Similar to the result of Table 3.4 when  $\Delta T = \tau_{TR}$ , our proposed approach is still more efficient. We note that the number of the parameters of interest  $QK_s$  is set to the same as Method II and the uniform subsampling is randomly generated five times. Also, the total time needed for obtaining the optimal designs for Method III is nearly similar to Method II presented in Table 3.7, so it's not reported again here. As presented in Table 3.10, the optimal designs  $\mathbf{d}_s^*$ 's in Method III compared to  $\mathbf{d}_s^*$ 's obtained from Method II achieve a relative high efficiency when  $K_s$  is large and close to the original  $K$  but they do not outperform the optimal designs obtained from Method II when  $K_s$  is small (e.g., less than 97).

Table 3.11 compares the optimal design of our proposed method with m-sequence-based design and ten randomly generated designs when  $\tau_{ISI} = 2, 4,$  and  $5$  seconds. The efficiency ratio  $\phi(\mathbf{d}^*|\mathbf{X})/\phi(\mathbf{d}_s^*|\mathbf{X})$  of these traditional fMRI designs to our  $\mathbf{d}_s^*$  are shown in Table 3.11. The range of the relative efficiency of these random designs are also reported. These designs do not perform as well as the  $\mathbf{d}_s^*$ 's obtained from our proposed method. It is noteworthy that no design  $\mathbf{d}^*$  has a higher  $\phi(\mathbf{d}^*|\mathbf{X})$  than  $\phi(\mathbf{d}_s^*|\mathbf{X})$  among the designs  $\mathbf{d}_s^*$  we considered.



Table 3.10:  $A$ -optimality for estimation for the two methods when  $Q = 1$  and 2,  $\tau_{ISI} = 2, 4,$  and 5 seconds, and  $p =$  positive divisors of  $m_{ISI}$ .

		$Q = 1$			$Q = 2$		
$\tau_{ISI}$	$K_s$	Method II	Method III	RE- $\phi$	Method II	Method III	RE- $\phi$
2	241	20.7209	20.7097-20.8328	0.9995-1.0054	12.4681	12.4076-12.5724	0.9952-1.0084
	193	20.7153	20.6325-20.7715	0.9960-1.0027	12.5032	12.3100-12.5705	0.9846-1.0054
	97	20.7218	19.6981-20.2955	0.9506-0.9794	12.5393	11.9260-12.3112	0.9511-0.9818
	49	20.6209	12.8379-14.0283	0.6226-0.6803	12.5688	7.3169-8.5959	0.5822-0.6839
4	241	23.1681	23.1651-23.3027	0.9999-1.0058	14.2077	14.0988-14.1948	0.9923-0.9991
	193	23.2704	23.1185-23.1871	0.9935-0.9964	14.1600	13.9592-14.1319	0.9858-0.9980
	121	23.3023	22.5482-22.9624	0.9676-0.9854	14.1904	13.5327-13.9350	0.9537-0.9820
	97	23.2749	21.8266-22.6340	0.9378-0.9725	14.1994	12.5929-13.8268	0.8869-0.9738
	49	23.2435	13.2986-15.2454	0.5721-0.6559	14.1986	8.0438-11.0293	0.5665-0.7768
	25	23.2534	8.5456-14.0673	0.3675-0.6050	14.2023	7.2364-9.2106	0.5096-0.6485
5	193	24.4266	24.1885-24.4066	0.9903-0.9992	14.7547	14.6042-14.7427	0.9898-0.9992
	97	24.4707	22.4473-23.5042	0.9173-0.9605	14.7680	13.7540-14.2069	0.9313-0.9620
	39	24.4883	7.8927-16.0912	0.3223-0.6571	14.7465	8.7185-10.5465	0.5912-0.7152
	21	24.4953	0-15.5380	0-0.6343	14.7040	1.4686-11.0190	0.0999-0.7494

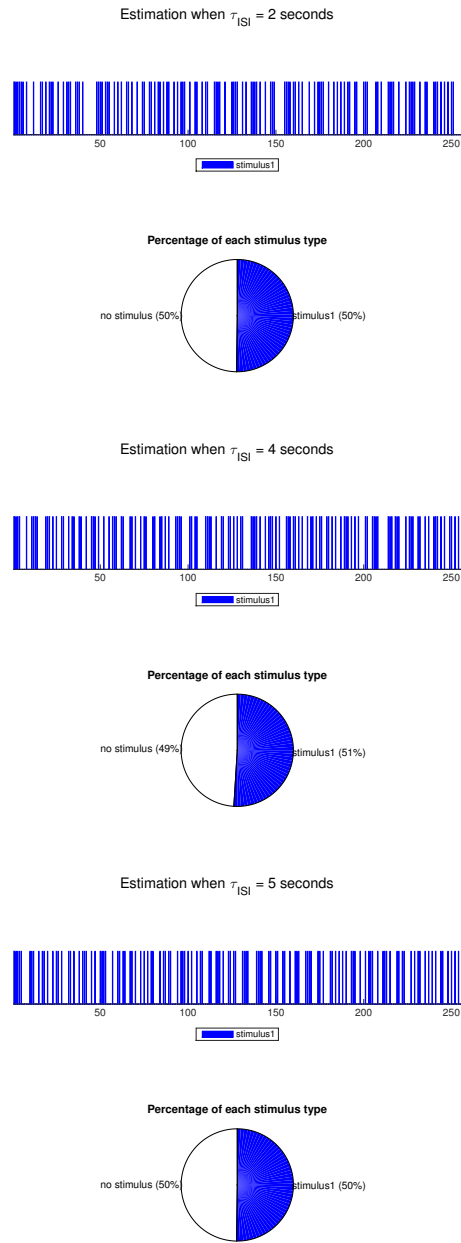


Figure 3.7: The optimal  $\mathbf{d}_s^*$  obtained from Method II when  $Q = 1$  and  $\tau_{ISI} = 2, 4,$  and 5 seconds.

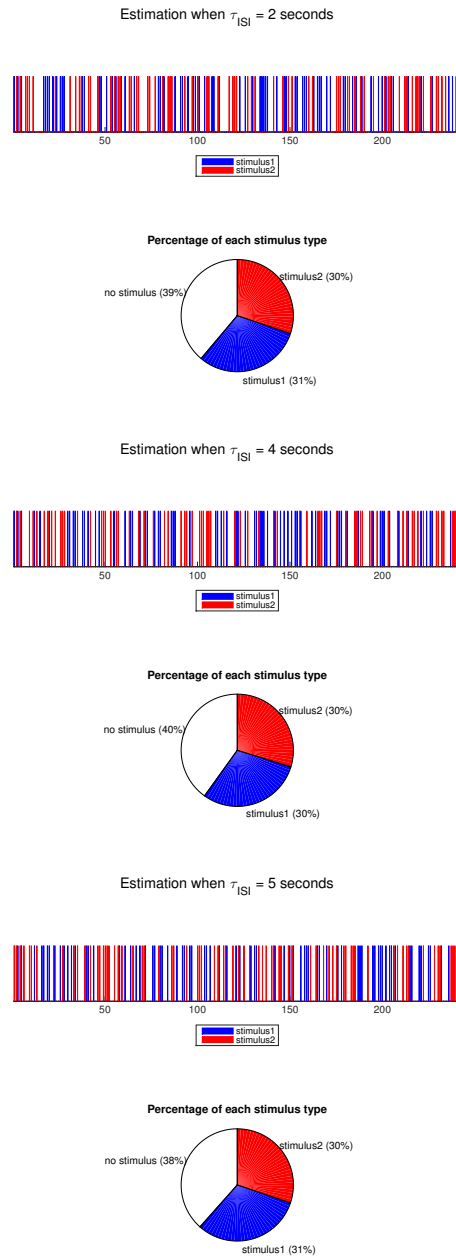


Figure 3.8: The optimal  $\mathbf{d}_s^*$  obtained from Method II when  $Q = 2$  and  $\tau_{ISI} = 2, 4,$  and 5 seconds.

Table 3.11: The optimal designs  $\mathbf{d}_s^*$ 's by Method II versus traditional designs when  $Q = 1, 2,$  and  $3$ .

$Q$	$\tau_{ISI}$	$\mathbf{d}_s^*$	m-sequence	RE- $\phi$	ten-random	RE- $\phi$
1	2	20.6209	17.7382	0.8602	15.2161-16.7100	0.7379-0.8103
	4	23.2534	20.6756	0.8891	18.2696-20.7990	0.7857-0.8945
	5	24.4953	21.4345	0.8751	20.1571-21.6528	0.8229-0.8840
2	2	12.5688	8.0774	0.6427	6.5931-8.2077	0.5247-0.6530
	4	14.2023	11.1442	0.7847	10.3382-11.5046	0.7279-0.8101
	5	14.7040	11.8807	0.8080	10.4219-12.1136	0.7088-0.8238
3	2	9.2056	5.3098	0.5768	3.7459-4.7732	0.4069-0.51851
	4	11.0670	0.7656	0.8443	5.9939-7.8567	0.5416-0.7099
	5	11.5630	9.1914	0.7949	7.3875-9.4560	0.6389 -0.8178

## Chapter 4

### A GENERAL LINEAR MODEL FOR DETECTION

#### 4.1 Temporal Derivative

Another study objective in many functional imaging experiments is identification of the brain region, e.g. voxel/channel that are activated by the stimuli, called detection problem. For this study objective, the HRF is commonly approximated by the product of an assumed shape of the HRF, e.g.  $h^*(t)$  in (2.1) and an unknown magnitude/amplitude,  $\theta$ . In order to allow some variations in the HRF shape, the partial derivatives of the canonical HRF (2.1) with respect to its delay and dispersion can be added as further basis functions. The choice of basis functions depends on the acquired data and more robust results are obtained when the basis set is small and temporally compact (Friston *et al.*, 1998). Furthermore, including many parameters would increase the estimation error, decrease the statistical power and the degrees of freedom especially when multicollinearity occurs. The most commonly considered basis set for fMRI and fNIRS studies is canonical HRF plus its derivatives approach developed by Friston *et al.* (1998). For example, Uga *et al.* (2014), Pinti *et al.* (2017), Tak *et al.* (2016) used the convolution model with a basis set consisting of the canonical HRF and its temporal derivative. The reason for including the temporal derivative is that it can capture the small offsets or latency in the time to the onset/peak of the HRF (Henson and Friston, 2007; Henson *et al.*, 2002; Liao *et al.*, 2000). Another reason is that the precise HRF shape appears to vary considerably across people and across the brain regions in most existing fMRI/fNIRS studies when performing group analysis.

Given  $h_q(t) = \theta_q h^*(t)$ , where  $h^*(t)$  is an assumed common HRF shape such as the double-gamma function in (2.1), and  $\theta_q$  is an unknown coefficient for the HRF of the  $q^{\text{th}}$ -type stimulus,  $q = 1, \dots, Q$ . A time-shifted version  $\tau$  of the HRF can be described by  $h_q(t + \tau)$ , where  $\tau$  allows a time shift for the HRF. The first-order Taylor expansion of  $h_q(t + \tau)$  can give a linear approximation of the shifted HRF, i.e.  $h_q(t + \tau) \approx \theta_q h^*(t) + \tau \theta_q \frac{d}{dt} h^*(t)$ . Thus, if  $h^*(t)$  and  $\frac{d}{dt} h^*(t)$  are used as two basis functions in the general linear model to estimate the parameters  $\theta_{cq}$  and  $\theta_{tq}$  respectively of  $h_q(t + \tau) \approx \theta_{cq} h^*(t) + \theta_{tq} \frac{d}{dt} h^*(t)$ , then small changes in the latency of the HRF can be captured by the parameter estimate of the temporal derivative (more precisely,  $\tau_q \approx \frac{\theta_{tq}}{\theta_{cq}}$ ). Henson *et al.* (2002) identified responses that are earlier than the canonical HRF when the ratio,  $\frac{\theta_{tq}}{\theta_{cq}}$ , is positive, and responses that are later than the canonical HRF when the ratio parameter,  $\frac{\theta_{tq}}{\theta_{cq}}$ , is negative. A similar logic applies to the use of dispersion derivative to capture the very small differences in the duration of the peak response. However, we focus here on the temporal derivative.

## 4.2 Methodology

The linear model (2.3) is considered for modeling the data collected from experiments having high temporal resolution brain imaging. Incorporating the temporal derivative, the model for detection can thus be written as:

$$\mathbf{y} = \mathbf{X} \mathbf{h}^* \boldsymbol{\theta}_c + \mathbf{X} d\mathbf{h}^* \boldsymbol{\theta}_t + \mathbf{S} \boldsymbol{\gamma} + \boldsymbol{\epsilon}. \quad (4.1)$$

where  $\boldsymbol{\theta}_c = [\theta_{c1}, \theta_{c2}, \theta_{c3}, \dots, \theta_{cQ}]'$ , and  $\boldsymbol{\theta}_t = [\theta_{t1}, \theta_{t2}, \theta_{t3}, \dots, \theta_{tQ}]'$ , the vector  $\mathbf{h}^*$  corresponds to the canonical HRF of (2.1), and the vector  $d\mathbf{h}^*$  corresponds to its temporal derivative. All the remaining terms are defined as in (2.3). The Model 4.1 can be rewritten in the following form:

$$\mathbf{y} = \mathbf{X} \mathbf{H} \boldsymbol{\theta} + \mathbf{S} \boldsymbol{\gamma} + \boldsymbol{\epsilon}. \quad (4.2)$$

Here  $\mathbf{H}_{Q[K,2]} = \mathbf{I}_Q \otimes [\mathbf{h}^* \quad d\mathbf{h}^*]$  and  $\boldsymbol{\theta} = [\boldsymbol{\theta}'_c, \boldsymbol{\theta}'_t]'$ . Note, the basis functions are orthogonalized via Gram-Schmidt orthogonalization as in the SPM software package (Henson and Friston, 2007; Wager *et al.*, 2005).

#### 4.2.1 Estimation of the Amplitude

When model (4.2) is considered for detection, it is common in some studies to only use the non-derivative term as an estimate of the HRF amplitude. However, this solution could be biased and the more accurate way for amplitude estimation is the use of the parametric function  $\alpha_q$ , proposed by Calhoun *et al.* (2004) with  $\alpha_q = \text{sgn}(\theta_{cq})\sqrt{\theta_{cq}^2 + \theta_{tq}^2}$  (Worsley and Taylor, 2006; Lindquist and Wager, 2007; Lindquist *et al.*, 2009). We note that a drawback of this parametric function,  $\alpha_q$ , is that it does not fully accommodate the case where there is an activation with  $\theta_{tq} \neq 0$  but  $\theta_{cq} = 0$ . This is especially the case when the *signum* function of a real number  $x$  is defined as:

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0; \\ 0 & \text{if } x = 0; \\ +1 & \text{if } x > 0, \end{cases} \quad (4.3)$$

which it is not differentiable at 0 in the ordinary sense and note that

$$x = \text{sgn}(x) |x|. \quad (4.4)$$

Therefore, we proposed a new parametric function with slight modification as:

$$\alpha_q = \begin{cases} \text{sgn}(\theta_{cq}) \sqrt{\theta_{cq}^2 + \theta_{tq}^2} & , \text{ if } \theta_{cq} \neq 0; \\ \theta_{tq} & , \text{ if } \theta_{cq} = 0. \end{cases} \quad (4.5)$$

$$= \begin{cases} \text{sgn}(\theta_{cq}) |\theta_{cq}| \sqrt{1 + (\theta_{tq}/\theta_{cq})^2} & , \text{ if } \theta_{cq} \neq 0; \\ \theta_{tq} & , \text{ if } \theta_{cq} = 0. \end{cases} \quad (4.6)$$

$$= \begin{cases} \theta_{cq} \sqrt{1 + (\theta_{tq}/\theta_{cq})^2} & , \text{ if } \theta_{cq} \neq 0; \\ \theta_{tq} & , \text{ if } \theta_{cq} = 0. \end{cases} \quad (4.7)$$

The third equality (4.7) is based on (4.4). This formulation would allow us to take into account the activation of  $\theta_{tq}$  even when  $\theta_{cq} = 0$  for obtaining the optimal designs.

By Delta method, the approximated covariance matrix of a parametric function  $\boldsymbol{\alpha} = g(\boldsymbol{\theta})$  under a given design  $\mathbf{d}$  that ensures estimability is:

$$\text{Cov}(\boldsymbol{\alpha}) = (\partial\boldsymbol{\alpha}/\partial\boldsymbol{\theta}')(\mathbf{M})^{-1}(\partial\boldsymbol{\alpha}/\partial\boldsymbol{\theta}'), \quad (4.8)$$

where the information matrix of  $\boldsymbol{\theta}$  is:

$$\mathbf{M} = \mathbf{H}'\mathbf{X}'\mathbf{V}'(\mathbf{I}_T - \mathcal{P}_{VS})\mathbf{V}\mathbf{X}\mathbf{H},$$

and  $\frac{\partial\alpha_q}{\partial\boldsymbol{\theta}'_q}$  of (4.7) for the first case with  $\theta_{cq} \neq 0$  can be written as:

$$\left. \frac{\partial\alpha_q}{\partial\boldsymbol{\theta}'_q} \right|_{\theta_{cq} \neq 0} = \left[ \frac{1}{\sqrt{1 + \left(\frac{\theta_{tq}}{\theta_{cq}}\right)^2}}, \frac{\theta_{tq}}{\theta_{cq} \sqrt{1 + \left(\frac{\theta_{tq}}{\theta_{cq}}\right)^2}} \right] \quad (4.9)$$

but the second case with  $\theta_{cq} = 0$  is as following:

$$\left. \frac{\partial\alpha_q}{\partial\boldsymbol{\theta}'_q} \right|_{\theta_{cq}=0} = [0, \quad 1] \quad (4.10)$$



For  $q = 1, \dots, Q$ , (4.9) can be written in matrix form as:

$$\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}_c \neq \mathbf{0}} = \begin{bmatrix} \frac{1}{\sqrt{1+\left(\frac{\theta_{t1}}{\theta_{c1}}\right)^2}} & \frac{\theta_{t1}}{\theta_{c1}\sqrt{1+\left(\frac{\theta_{t1}}{\theta_{c1}}\right)^2}} & 0 & 0 & \dots & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{1+\left(\frac{\theta_{t2}}{\theta_{c2}}\right)^2}} & \frac{\theta_{t2}}{\theta_{c2}\sqrt{1+\left(\frac{\theta_{t2}}{\theta_{c2}}\right)^2}} & \ddots & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \ddots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \frac{1}{\sqrt{1+\left(\frac{\theta_{tQ}}{\theta_{cQ}}\right)^2}} & \frac{\theta_{tQ}}{\theta_{cQ}\sqrt{1+\left(\frac{\theta_{tQ}}{\theta_{cQ}}\right)^2}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{1+\left(\frac{\theta_{t1}}{\theta_{c1}}\right)^2}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{1+\left(\frac{\theta_{t2}}{\theta_{c2}}\right)^2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sqrt{1+\left(\frac{\theta_{tQ}}{\theta_{cQ}}\right)^2}} \end{bmatrix} \times \begin{bmatrix} 1 & \frac{\theta_{t1}}{\theta_{c1}} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \frac{\theta_{t2}}{\theta_{c2}} & \ddots & \vdots & \\ \vdots & \ddots & \ddots & \ddots & 0 & 0 & \\ 0 & 0 & \dots & 0 & 0 & 1 & \frac{\theta_{tQ}}{\theta_{cQ}} \end{bmatrix}$$

$$\text{Let } \mathbf{N} = \begin{bmatrix} \frac{1}{\sqrt{1+\left(\frac{\theta_{t1}}{\theta_{c1}}\right)^2}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{1+\left(\frac{\theta_{t2}}{\theta_{c2}}\right)^2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sqrt{1+\left(\frac{\theta_{tQ}}{\theta_{cQ}}\right)^2}} \end{bmatrix}, \mathbf{G} = \begin{bmatrix} 1 & \frac{\theta_{t1}}{\theta_{c1}} & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \frac{\theta_{t2}}{\theta_{c2}} & \ddots & \vdots & \\ \vdots & \ddots & \ddots & \ddots & 0 & 0 & \\ 0 & 0 & \dots & 0 & 0 & 1 & \frac{\theta_{tQ}}{\theta_{cQ}} \end{bmatrix},$$

and then  $\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}'} = \mathbf{N} \times \mathbf{G}$ .

Denote  $\boldsymbol{\theta}_c$  as  $\boldsymbol{\theta}_0$  when at least one  $\theta_{cq} = 0$  for  $q = 1, \dots, Q$ . For example, when  $\theta_{ck} = 0$ , then  $\boldsymbol{\theta}_0 = [\theta_{c1}, \dots, \theta_{c,k-1}, 0, \theta_{c,k+1}, \dots, \theta_{cQ}]$ , then  $\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}'} \Big|_{\theta_{ck}=0}$  can be written as following:

$$\frac{\partial \boldsymbol{\alpha}}{\partial \boldsymbol{\theta}'} \Big|_{\theta_{ck}=0} = \begin{bmatrix} \frac{1}{\sqrt{1+\left(\frac{\theta_{t1}}{\theta_{c1}}\right)^2}} & \frac{\theta_{t1}}{\theta_{c1}\sqrt{1+\left(\frac{\theta_{t1}}{\theta_{c1}}\right)^2}} & 0 & 0 & \dots & \dots & \dots & \dots & 0 & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \frac{1}{\sqrt{1+\left(\frac{\theta_{t,k-1}}{\theta_{c,k-1}}\right)^2}} & \frac{\theta_{t,k-1}}{\theta_{c,k-1}\sqrt{1+\left(\frac{\theta_{t,k-1}}{\theta_{c,k-1}}\right)^2}} & \dots & 0 & 0 \\ 0 & 0 & \dots & \dots & \dots & \dots & \dots & 0 & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & \dots & \dots & \dots & \frac{1}{\sqrt{1+\left(\frac{\theta_{t,k+1}}{\theta_{c,k+1}}\right)^2}} & \frac{\theta_{t,k+1}}{\theta_{c,k+1}\sqrt{1+\left(\frac{\theta_{t,k+1}}{\theta_{c,k+1}}\right)^2}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Also as another argument,  $\alpha_q$  in (4.5) can be written as:

$$\alpha_q = \begin{cases} [2H(\theta_{cq}) - 1]\sqrt{\theta_{cq}^2 + \theta_{tq}^2} & , \text{ if } \theta_{cq} \neq 0; \\ \theta_{tq} & , \text{ if } \theta_{cq} = 0. \end{cases} \quad (4.11)$$

where  $\text{sgn}(x) = 2H(x) - 1$ , and  $H(x)$  is the Heaviside step function as

$$H(x) = \begin{cases} 0 & \text{if } x < 0; \\ 1/2 & \text{if } x = 0; \\ 1 & \text{if } x > 0. \end{cases}$$

Then,  $\left. \frac{\partial \alpha_q}{\partial \boldsymbol{\theta}'_q} \right|_{\theta_{cq} \neq 0}$  is as following:

$$\left. \frac{\partial \alpha_q}{\partial \boldsymbol{\theta}'_q} \right|_{\theta_{cq} \neq 0} = \left[ 2\delta(\theta_{cq})\sqrt{\theta_{cq}^2 + \theta_{tq}^2} + [2H(\theta_{cq}) - 1] \frac{\theta_{cq}}{\sqrt{\theta_{cq}^2 + \theta_{tq}^2}}, \quad [2H(\theta_{cq}) - 1] \frac{\theta_{tq}}{\sqrt{\theta_{cq}^2 + \theta_{tq}^2}} \right] \quad (4.12)$$

where  $\delta(x)$  is the Dirac delta, and  $\frac{\partial H(x)}{\partial x} = \delta(x)$

$$\delta(x) = \begin{cases} +\infty & \text{if } x = 0; \\ 0 & \text{if } x \neq 0. \end{cases}$$

Then,

$$\left. \frac{\partial \alpha_q}{\partial \boldsymbol{\theta}'_q} \right|_{\theta_{cq} \neq 0} = \left[ \pm \frac{\theta_{cq}}{\sqrt{\theta_{cq}^2 + \theta_{tq}^2}}, \pm \frac{\theta_{tq}}{\sqrt{\theta_{cq}^2 + \theta_{tq}^2}} \right] \quad (4.13)$$

Again,  $\left. \frac{\partial \alpha_q}{\partial \boldsymbol{\theta}'_q} \right|_{\theta_{cq} \neq 0}$  in (4.13) is deduced to  $N$  and  $G$  that are reported previously,

and  $\left. \frac{\partial \alpha_q}{\partial \boldsymbol{\theta}'_q} \right|_{\theta_{cq} = 0}$  is the same as (4.10).

### 4.3 $D$ -optimality Criterion

We would like a design minimizing the determinant of the covariance matrix stated in (4.8). The answer will depend on the unknown  $\boldsymbol{\theta}_c$  and  $\boldsymbol{\theta}_t$ . One way for tackling this

problem is by obtaining a locally optimal design that is optimal for a given  $(\boldsymbol{\theta}_c, \boldsymbol{\theta}_t)$ -vector (Chernoff, 1953). Another recommended way is the maximin approach because a good guess value for the parameter vector maybe unavailable. Moreover, the selected design should be relatively robust to a mis-specification of the  $(\boldsymbol{\theta}_c, \boldsymbol{\theta}_t)$ -vector. The Bayesian approach is another way as well to obtain optimal designs. The previously mentioned  $D$ -optimality criteria in (2.5) can be equivalently written as following:

$$\phi(\text{Cov}(g(\boldsymbol{\theta}))) = |(\partial\boldsymbol{\alpha}/\partial\boldsymbol{\theta}')\mathbf{M}^{-1}(\partial\boldsymbol{\alpha}/\partial\boldsymbol{\theta}')'|^{-\frac{1}{\mathcal{R}}}. \quad (4.14)$$

For simplicity, we set  $\phi(\mathbf{M}) = 0$  when  $\mathbf{M}$  is singular.  $\mathcal{R}$  is the number of the parameters of interest.

When all  $\theta_{cq} \neq 0$ , (4.14) can be expressed as a function of the ratio, say  $\eta_q$ , where  $\eta_q = \frac{\theta_{tq}}{\theta_{cq}}$ . Specifically,

$$\begin{aligned} \phi(\mathbf{d}; \boldsymbol{\alpha}) \Big|_{\boldsymbol{\theta}_c \neq \mathbf{0}} &= \phi(\mathbf{d}; \boldsymbol{\eta}) = \left| \left( \frac{\partial\boldsymbol{\alpha}}{\partial\boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}_c \neq \mathbf{0}} \right) \mathbf{M}^{-1} \left( \frac{\partial\boldsymbol{\alpha}}{\partial\boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}_c \neq \mathbf{0}} \right)' \right|^{-\frac{1}{\mathcal{R}}} \\ &= (|\mathbf{G}\mathbf{M}^{-1}\mathbf{G}'| |\mathbf{N}\mathbf{N}'|)^{-\frac{1}{\mathcal{R}}}. \end{aligned} \quad (4.15)$$

Because  $|\mathbf{N}\mathbf{N}'|$  does not depend on the design  $\mathbf{X}\mathbf{H}$ , we have the following Lemma 4.3.1 that is useful for the maximin approach.

**Lemma 4.3.1** *Any design  $\mathbf{d}$  maximizes  $|\mathbf{G}\mathbf{M}^{-1}\mathbf{G}'|^{-\frac{1}{\mathcal{R}}}$  if and only if it maximizes  $(|\mathbf{G}\mathbf{M}^{-1}\mathbf{G}'| |\mathbf{N}\mathbf{N}'|)^{-\frac{1}{\mathcal{R}}}$ .*

However, for the case where at least one  $\theta_{cq} = 0$ , the  $D$ -optimality criteria may or may not be free of the unknown parameter,  $\eta_q$ . In particular,

$$\phi(\mathbf{d}; \boldsymbol{\alpha}) \Big|_{\boldsymbol{\theta}_c = \boldsymbol{\theta}_0} = \phi(\mathbf{d}; \boldsymbol{\theta}_0) = \left| \left( \frac{\partial\boldsymbol{\alpha}}{\partial\boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}_c = \boldsymbol{\theta}_0} \right) \mathbf{M}^{-1} \left( \frac{\partial\boldsymbol{\alpha}}{\partial\boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}_c = \boldsymbol{\theta}_0} \right)' \right|^{-\frac{1}{\mathcal{R}}}, \quad (4.16)$$

where,  $\frac{\partial\boldsymbol{\alpha}}{\partial\boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}_c = \boldsymbol{\theta}_0}$  is reported under (4.10) for  $q = 1, \dots, Q$ .

#### 4.4 The Applied Maximin Approach

Because the value of the ratio parameter vector,  $\boldsymbol{\eta}$ , is uncertain, we would like to find a design that is robust to a mis-specification of the  $\boldsymbol{\eta}$ -value. The applied approach for obtaining such a design is the maximin approach considered in Maus *et al.* (2010), and Kao and Mittelman (2014). Solving a maximin design problem can be very computationally expensive, and it is to obtain a maximin design  $\mathbf{d}_{Mm}^*$  that maximizes:

$$\min_{\boldsymbol{\eta} \in \Omega} (RE(\mathbf{d}; \mathbf{d}_{\boldsymbol{\eta}}^*), RE(\mathbf{d}; \mathbf{d}_{\boldsymbol{\theta}_0}^*)) = \min_{\boldsymbol{\eta} \in \Omega} \left( \frac{\phi(\mathbf{d}; \boldsymbol{\eta})}{\phi(\mathbf{d}_{\boldsymbol{\eta}}^*; \boldsymbol{\eta})}, \frac{\phi(\mathbf{d}; \boldsymbol{\theta}_0, \boldsymbol{\eta})}{\phi(\mathbf{d}_{\boldsymbol{\theta}_0}^*; \boldsymbol{\theta}_0, \boldsymbol{\eta})} \right), \quad (4.17)$$

where  $\Omega$  is the parameter space of  $\boldsymbol{\eta}$ . The optimization will involve three steps:

1. Obtain the following locally optimal designs
  - (a)  $\mathbf{d}_{\boldsymbol{\eta}}^*$  that maximizes  $\phi(\mathbf{d}; \boldsymbol{\eta})$  for each of the (many) possible values of  $\boldsymbol{\eta}$ .
  - (b)  $\mathbf{d}_{\boldsymbol{\theta}_0}^*$  that maximizes  $\phi(\mathbf{d}; \boldsymbol{\theta}_0, \boldsymbol{\eta})$ .
2. With these  $\mathbf{d}_{\boldsymbol{\eta}}^*$ s and  $\mathbf{d}_{\boldsymbol{\theta}_0}^*$ s, we then find  $\min_{\boldsymbol{\eta} \in \Omega} (RE(\mathbf{d}; \mathbf{d}_{\boldsymbol{\eta}}^*), RE(\mathbf{d}; \mathbf{d}_{\boldsymbol{\theta}_0}^*))$  for each candidate design  $\mathbf{d}$ .
3. Achieve a design yielding the maximal value of  $\min_{\boldsymbol{\eta} \in \Omega} (RE(\mathbf{d}; \mathbf{d}_{\boldsymbol{\eta}}^*), RE(\mathbf{d}; \mathbf{d}_{\boldsymbol{\theta}_0}^*))$ .

To search for a maximin design, we will replace  $\Omega$  by a grid, and use the genetic algorithm to find the target optimal designs in step 1 and 3.

#### 4.5 Simulation Results

In this section, we focus on the detection of the HRFs by estimating the amplitude,  $\boldsymbol{\alpha}$ , in (4.7). We consider three different scenarios that, respectively, have  $(Q, N)$ :  $(1, 127)$ ,  $(2, 121)$  and  $(3, 127)$  with  $\tau_{ISI} = 2$  seconds. For this case, we set  $\tau_{TR} = \Delta T =$

0.1 seconds; this means that the response sampling rate is 10 Hz. At an activated brain region, each stimulus is assumed to evoke an HRF of  $\tau_{dur} = 32$  seconds. The drift of the time series, i.e.  $\mathbf{S}\boldsymbol{\gamma}$  is assumed to be a second-order Legendre polynomial, (Liu and Frank, 2004) and, for simplicity, the error term is assumed to follow a stationary AR(1) process with an autocorrelation coefficient of 0.3. We now present the simulation results under  $D$ -optimality. All the computations in the following subsection are conducted on a desktop computer of a 4.00 GHz Intel Core i7-4790k quad-core processor with 16GB RAM.

Two methods are considered for obtaining maximin designs. Method IA is our proposed method described in the previous section when  $\alpha_q$  as in (4.7). We first adopt the genetic algorithm of Kao *et al.* (2009) to obtain the required locally optimal designs  $\mathbf{d}_\eta^*$  that maximizes the  $D$ -optimality criterion  $\phi(\mathbf{d}; \boldsymbol{\eta})$  stated in (4.15). In particular, for  $Q = 1, 2$ , and 3, we obtain (9, 81, 729) locally optimal design  $\mathbf{d}_\eta^*$  that maximizes  $\phi(\mathbf{d}; \boldsymbol{\eta})$  for the (9, 81, 729) grid points of  $\boldsymbol{\eta}$  in  $\tilde{G}_1$ ,  $\tilde{G}_2$ , and  $\tilde{G}_3$ ; respectively. To reflect the change in the latency of the response, it typically varies between  $[-2, 2]$  seconds and the corresponding ratio parameter also varies between  $[-2, 2]$  seconds; see also Figure 1.d of Henson *et al.* (2002), Worsley and Taylor (2006), Henson and Friston (2007), and Liao *et al.* (2000). We set  $\tilde{G}_1 = \{\eta_1 \mid \eta_i = -2, -1.5, -1, \dots, 2, i = 1\}$ . For  $Q = 2$ ,  $\tilde{G}_2 = \{(\eta_1, \eta_2) \mid \eta_i = -2, -1.5, -1, \dots, 2, i = 1, 2\}$  and similarly for  $Q = 3$ ,  $\tilde{G}_3 = \{(\eta_1, \eta_2, \eta_3) \mid \eta_i = -2, -1.5, -1, \dots, 2, i = 1, 2, 3\}$ . Then, we again adopt the genetic algorithm to obtain the required locally optimal designs  $\mathbf{d}_{\theta_0}^*$  that maximizes the  $D$ -optimality criterion  $\phi(\mathbf{d}; \boldsymbol{\theta}_c = \boldsymbol{\theta}_0)$  stated in (4.16). In particular, for  $Q = 1, 2$ , and 3, we obtain (1, 19, 271) locally optimal design  $\mathbf{d}_{\theta_0}^*$  that maximizes  $\phi(\mathbf{d}; \boldsymbol{\theta}_0, \boldsymbol{\eta})$  in  $G_1$ ,  $G_2$ , and  $G_3$ ; respectively. We define  $\boldsymbol{\theta}_0^{(k)}$  be a vector contains  $k$  zero elements. Thus,  $G_1 = \{\boldsymbol{\theta}_0^{(1)}\}$ . For  $Q = 2$ ,  $G_2 = \{\boldsymbol{\theta}_0^{(2)}\} \cup \{\{\boldsymbol{\theta}_0^{(1)}\} \times \tilde{G}_1\}$  and similarly for  $Q = 3$ ,  $G_3 = \{\boldsymbol{\theta}_0^{(3)}\} \cup \{\{\boldsymbol{\theta}_0^{(2)}\} \times \tilde{G}_1\} \cup \{\{\boldsymbol{\theta}_0^{(1)}\} \times \tilde{G}_2\}$ .

Table 4.1 summarizes the number of locally optimal designs needed for Method IA and the total computing time spent on obtaining all these locally optimal designs for  $q = 1, 2$ , and  $3$  in  $\tilde{G}_1 \cup G_1, \tilde{G}_2 \cup G_2$ , and  $\tilde{G}_3 \cup G_3$ ; respectively.

Table 4.1: Locally Optimal Designs; Method IA

	$Q = 1$	$Q = 2$	$Q = 3$
# of LODs	10	100	1000
CPU(mins)	10.73	279.65	5300.66

In Table 4.2, we compare the performance of Method IA on obtaining maximin designs,  $\mathbf{d}_{Mm}^*$ , for  $(Q, N) = (1, 127), (2, 121)$ , and  $(3, 127)$ . The minimum relative efficiency of the obtained  $\mathbf{d}_{Mm}^*$  and the computing time are reported there (the first row). This is only the time spent on the genetic algorithm to search for  $\mathbf{d}_{Mm}^*$ . The total CPU time needed should also include the time required for obtaining the locally optimal designs presented in Table 4.1. Consequently, the total times needed for obtaining  $\mathbf{d}_{Mm}^*$  are approximately 12, 283 and 5307 minutes, respectively. The times needed for different stimulus are increasing, especially for  $Q = 3$ . Our proposed approach (first row) is the most efficient. For evaluating the performance of the obtained  $\mathbf{d}_{Mm}^*$ , we consider the optimal design  $\mathbf{d}_{\theta_c}^*$ , and  $\mathbf{d}_{\theta_c|\theta_t}^*$  that maximize  $\phi(\mathbf{d}^*; \boldsymbol{\theta}_c) = |\mathbf{M}_1|^{\frac{1}{\mathcal{R}}}$ , and  $\phi(\mathbf{d}^*; \boldsymbol{\theta}_c|\boldsymbol{\theta}_t) = |\mathbf{M}_2|^{\frac{1}{\mathcal{R}}}$  where,

$$\mathbf{M}_1 = (\mathbf{I}_Q \otimes \mathbf{h}^*)' \mathbf{X}' \mathbf{V}' (\mathbf{I}_T - \mathcal{P}_{\mathbf{V}S}) \mathbf{V} \mathbf{X} (\mathbf{I}_Q \otimes \mathbf{h}^*);$$

$$\mathbf{M}_2 = (\mathbf{I}_Q \otimes \mathbf{h}^*)' \mathbf{X}' \mathbf{V}' (\mathbf{I}_T - \mathcal{P}_{\mathbf{V}[\mathbf{x}(\mathbf{I}_Q \otimes \mathbf{a}\mathbf{h}^*), S]}) \mathbf{V} \mathbf{X} (\mathbf{I}_Q \otimes \mathbf{h}^*).$$

We evaluate these designs to calculate  $\min_{\eta \in \Omega} (RE(\mathbf{d}; \mathbf{d}_\eta^*), RE(\mathbf{d}; \mathbf{d}_{\theta_0}^*))$ . By comparing the first row with the last two in 4.2, all  $\mathbf{d}_{Mm}^*$ s obtained by different stimulus perform similarly, and they all achieve at least 97% relative efficiency over

$\tilde{G}_1 \cup G_1, \tilde{G}_2 \cup G_2,$  and  $\tilde{G}_3 \cup G_3$ . These results indicate that our proposed Method IA can efficiently obtain a good maximin design.

Table 4.2:  $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_\eta^*, \mathbf{d}_{\theta_0}^*)$  of  $\mathbf{d}_{Mm}^*$  Method IA Versus  $\mathbf{d}_{\theta_c}^*$  and  $\mathbf{d}_{\theta_c|\theta_t}^*$

	$Q = 1$		$Q = 2$		$Q = 3$	
	$\phi$	CPU(#Iter)	$\phi$	CPU(#Iter)	$\phi$	CPU (#Iter)
$\mathbf{d}_{Mm}^*$	0.9775	1.53(601)	0.9704	3.21(801)	0.9653	6.24(801)
$\mathbf{d}_{\theta_c}^*$	0.9049	0.90(401)	0.8340	2.25(601)	0.7974	2.17(401)
$\mathbf{d}_{\theta_c \theta_t}^*$	0.9051	29.30(601)	0.8949	44.47(1001)	0.8875	41.65(801)

In Table 4.3, we compare the performance of the  $\mathbf{d}_{Mm}^*$  obtained by Method IA with some traditional designs that are widely used in functional brain imaging studies. These traditional designs include blocked designs, an  $m$ -sequence-based design, and ten randomly generated designs. We consider blocked designs having a 16 seconds-on-16 seconds-off pattern or 12 seconds-on-12 seconds-off pattern. These designs perform well for detecting brain activation under linear models, (Maus *et al.*, 2010). Even though blocked design of size six performs similarly as the maximin design in  $Q = 1$ , for larger  $Q$ 's, they do not perform well for detection under the linear model when the parametric function is non linear. An  $m$ -sequence (or maximum length sift-register sequence) of length  $N = (Q + 1)^r - 1$ . These sequences are commonly used for estimating the HRF in fMRI studies (i.e. for estimating  $\mathbf{h}$  of Model (2.2)). However, they do not perform as well as the  $\mathbf{d}_{Mm}^*$  obtained from Method IA. We also randomly generate ten design sequences and the range of the  $\min_{\eta \in \Omega} (RE(\mathbf{d}; \mathbf{d}_\eta^*), RE(\mathbf{d}; \mathbf{d}_{\theta_0}^*))$  of these random designs are also reported in Table 4.2. Similarly, they do not outperform the maximin design obtained from our proposed method.

In Figure 4.2, we provide these maximin designs by presenting the value of the

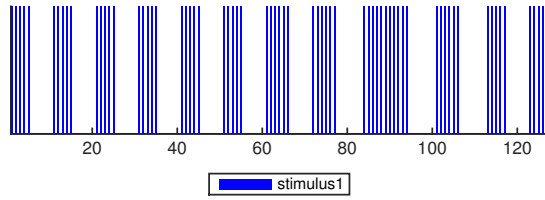
Table 4.3:  $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_{\eta}^*, \mathbf{d}_{\theta_0}^*)$  of  $\mathbf{d}_{Mm}^*$  From Method IA Versus Some Traditional Designs

	$Q = 1$	$Q = 2$	$Q = 3$
$\mathbf{d}_{Mm}^*$	0.9775	0.9730	0.9653
$m$ -sequence-based design	0.8476	0.8396	0.8423
blocked design (size=8)	0.9047	0.8348	0.7982
blocked design (size=6)	0.9648	0.8669	0.8275
ten random designs	0.8196-0.8693	0.7899-0.8720	0.8104-0.8421

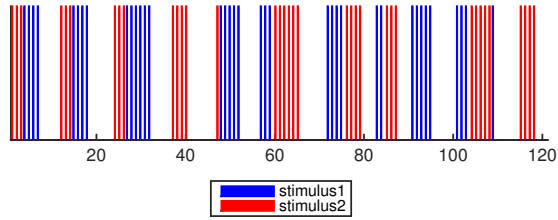
$n$ th element  $d_n$  of  $\mathbf{d}_{Mm}^*$ . These  $\mathbf{d}_{Mm}^*$ 's seem to be a mixture of a blocked and a rapid event-related design, i.e.,  $m$ -sequences or random designs, (Maus *et al.*, 2010).



(Q, N)= (1, 127)



(Q, N)= (2, 121)



(Q, N)= (3, 127)

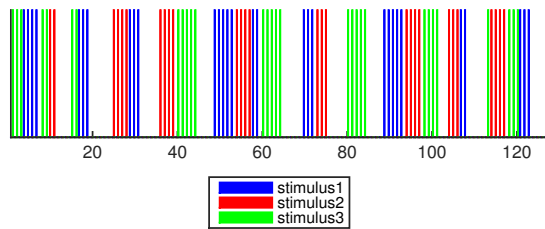


Figure 4.1: The maximin designs  $\mathbf{d}_{Mm}^* = (d_1, \dots, d_N)$  obtained from Method IA.

Method IB is the alternative method that we proposed. It is partially described in the previous section when considering the first case of  $\alpha_q$  in 4.7. Similar to method IA, we use the obtained locally optimal designs  $\mathbf{d}_\eta^*$  that maximizes the  $D$ -optimality criterion  $\phi(\mathbf{d}; \boldsymbol{\eta})$ . In particular, for  $Q = 1, 2$ , and  $3$  in previous results, we obtained  $(9, 81, 729)$  locally optimal designs that maximizes  $\phi(\mathbf{d}; \boldsymbol{\eta})$  in  $\tilde{G}_1, \tilde{G}_2$ , and  $\tilde{G}_3$ ; respectively.  $\tilde{G}_1, \tilde{G}_2, \tilde{G}_3$  are as described in Method IA.

Table 4.4 summarizes the number of locally optimal designs needed for Method IB and the total computing time spent on obtaining all these locally optimal designs.

Table 4.4: Locally Optimal Designs; Method IB

	$Q = 1$	$Q = 2$	$Q = 3$
# of LODs	9	81	729
CPU(mins)	9.68	228	3888.50

In Table 4.5, we compare the performance of Method IB on obtaining maximin designs,  $\mathbf{d}_{Mm}^*$ , for  $(Q, N) = (1, 127), (2, 121)$ , and  $(3, 127)$ . The minimum relative efficiency of the obtained  $\mathbf{d}_{Mm}^*$  and the computing time are reported in the first row. The total CPU time needed for obtaining  $\mathbf{d}_{Mm}^*$  are approximately 11, 234 and 3894 minutes, respectively. For evaluating the performance of the obtained  $\mathbf{d}_{Mm}^*$ , we consider the locally optimal design  $\mathbf{d}_{\theta_c}^*$ , and  $\mathbf{d}_{\theta_c|\theta_t}^*$  that maximize  $\phi(\mathbf{d}^*; \boldsymbol{\theta}_c)$  and  $\phi(\mathbf{d}^*; \boldsymbol{\theta}_c | \boldsymbol{\theta}_t)$  similar to the previous section. This results is very similar to the obtained results on Method IA.

In Table 4.6, we compare the performance of the  $\mathbf{d}_{Mm}^*$  obtained by Method IB with some traditional designs. These traditional designs include blocked designs, an  $m$ -sequence-based design, and ten randomly generated designs. We consider blocked designs having a 16 seconds-on-16 seconds-off pattern or 12 seconds-on-12 seconds-off

Table 4.5:  $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_\eta^*)$  of  $\mathbf{d}_{Mm}^*$  Versus  $\mathbf{d}_{\theta_c}^*$  and  $\mathbf{d}_{\theta_c|\theta_t}^*$ ; Method IB

	$Q = 1$		$Q = 2$		$Q = 3$	
	$\phi$	CPU(#Iter)	$\phi$	CPU(#Iter)	$\phi$	CPU (#Iter)
$\mathbf{d}_{Mm}^*$	0.9777	1.54(601)	0.9732	5.56(1401)	0.9669	5.75(801)
$\mathbf{d}_{\theta_c}^*$	0.9049	0.90(401)	0.8340	2.25(601)	0.7974	2.17(401)
$\mathbf{d}_{\theta_c \theta_t}^*$	0.9051	29.30(601)	0.8949	44.47(1001)	0.8875	41.65(801)

pattern. These designs perform the same as the traditional designs in Method IA.

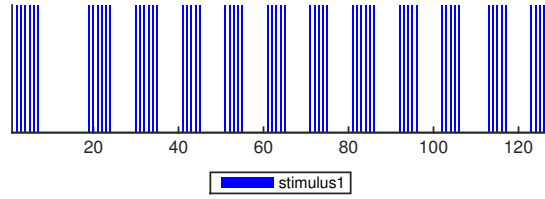
Table 4.6:  $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_\eta^*)$  of  $\mathbf{d}_{Mm}^*$  Versus Some Traditional Designs; Method IB

	$Q = 1$	$Q = 2$	$Q = 3$
$\mathbf{d}_{Mm}^*$	0.9777	0.9730	0.9669
$m$ -sequence-based design	0.8476	0.8396	0.8423
blocked design (size=8)	0.9047	0.8348	0.7982
blocked design (size=6)	0.9648	0.8669	0.8275
ten random designs	0.8030-0.8679	0.8086-0.8635	0.7998-0.8525

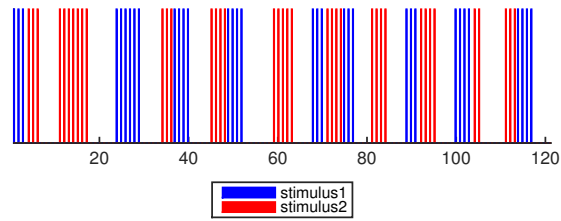
Figure 4.2, presents the value of the  $n$ th element  $d_n$  of  $\mathbf{d}_{Mm}^*$  maximin designs. Similar to the obtained results in Method IA, these  $\mathbf{d}_{Mm}^*$ 's seem to be a mixture of a blocked and a rapid event-related design.

In Table 4.7, we evaluate  $\mathbf{d}_{Mm}^*$  by Method IB over  $\tilde{G}_1 \cup G_1, \tilde{G}_2 \cup G_2$ , and  $\tilde{G}_3 \cup G_3$ ; respectively. Also, we evaluate  $\mathbf{d}_{Mm}^*$  by Method IA over  $\tilde{G}_1, \tilde{G}_2$ , and  $\tilde{G}_3$ ; respectively. The minimum relative efficiencies are reported for  $q = 1, 2, 3$ . Both methods can efficiently generate high quality designs. However, the proposed Method IB is more efficient than Method IA in terms of the achieved design efficiency and required CPU

(Q, N)= (1, 127)



(Q, N)= (2, 121)



(Q, N)= (3, 127)

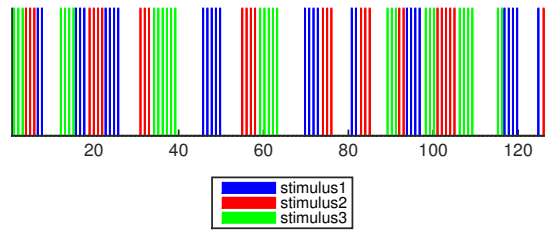


Figure 4.2: The maximin designs  $\mathbf{d}_{Mm}^* = (d_1, \dots, d_N)$  obtained from Method IB.

time. These results indicate that Method IB can efficiently obtain a good maximin design.

Table 4.7: Evaluation of  $\mathbf{d}_{Mm}^*$  Method IB Versus  $\mathbf{d}_{Mm}^*$  Method IA

	$Q = 1$	$Q = 2$	$Q = 3$
$\min_{\eta \in \Omega} RE(\mathbf{d}_{MmIB}^*; \mathbf{d}_\eta^*, \mathbf{d}_{\theta_0}^*)$	0.9777	0.9732	0.9669
$\min_{\eta \in \Omega} RE(\mathbf{d}_{MmIA}^*; \mathbf{d}_\eta^*)$	0.9775	0.9704	0.9653

As mentioned previously that Method IB is more efficient, Table 4.8 reports minimum relative efficiencies of  $\mathbf{d}_{Mm}^*$  Method IB,  $\mathbf{d}_{\theta_c}^*$ , and  $\mathbf{d}_{\theta_c|\theta_t}^*$  over  $G_1, G_2$ , and  $G_3$ . All these results show that  $\min_{\eta \in \Omega} RE(\mathbf{d}; \mathbf{d}_{\theta_0}^*) \geq \min_{\eta \in \Omega} RE(\mathbf{d}; \mathbf{d}_\eta^*)$  where  $\mathbf{d} = \mathbf{d}_{Mm}^*, \mathbf{d}_{\theta_c}^*$ , and  $\mathbf{d}_{\theta_c|\theta_t}^*$ . It suggests that GA tends to perform well by focusing on the reduced parameter space  $\tilde{G}_1, \tilde{G}_2$ , and  $\tilde{G}_3$ ; respectively. In addition, these designs are still performing well with or without including  $\theta_c = \theta_0$  in the parameter space.

Table 4.8:  $\min_{\eta \in \Omega} RE(\cdot; \mathbf{d}_{\theta_0}^*)$  of  $\mathbf{d}_{Mm}^*$  Method IB Versus  $\mathbf{d}_{\theta_c}^*$  and  $\mathbf{d}_{\theta_c|\theta_t}^*$

	$Q = 1$	$Q = 2$	$Q = 3$
$\mathbf{d}_{Mm}^*$	0.9781	0.9758	0.9686
$\mathbf{d}_{\theta_c}^*$	0.9197	0.8454	0.8103
$\mathbf{d}_{\theta_c \theta_t}^*$	0.9197	0.9031	0.8946

## CONCLUSION

In this dissertation, we investigate two issues in high temporal resolution neuro-imaging studies when obtaining optimal designs.

In the first part, we propose an efficient approach to obtain optimal designs for high temporal resolution neuro-imaging studies to estimate the HRF. The major challenge is that the design matrix is large. In order to tackle this issue, the main idea is by considering the information matrix,  $\mathbf{M}_s$  of  $\mathbf{h}_s$  in Model (3.2) instead of that of the ‘full’ information matrix  $\mathbf{M}$  of  $\mathbf{h}$  of Model (3.1). Thus, we apply the proposed approach described in Step X-1 to X-3 to optimize the surrogate criterion  $\phi(\mathbf{d}|\mathbf{X}_s)$  of  $\phi(\mathbf{d}|\mathbf{X})$ . Under certain conditions, we show mathematically that a judiciously selected  $\mathbf{X}_s$  not only allows an unbiased estimate of  $\mathbf{h}_s$  as suggested by Lemma 3.1.1, but also makes  $\phi(\mathbf{d}|\mathbf{X}_s)$  a good surrogate of  $\phi(\mathbf{d}|\mathbf{X})$ . Moving away from these conditions, we consider the genetic algorithm of Kao *et al.* (2009) to obtain an optimal design  $\mathbf{d}^*$  that maximizes  $\phi(\mathbf{d}_s^*|\mathbf{X}_s)$  using the subsampled matrix  $\mathbf{X}_s$ . We demonstrate the usefulness of our method for estimating the HRF with both  $\tau_{TR} = \Delta_T$  and  $\tau_{TR} \neq \Delta_T$  under the  $A$ -optimality criteria through case studies. We find that our obtained designs outperform some traditional fMRI designs in all cases. We also observe that  $\phi(\mathbf{d}_s^*|\mathbf{X}_s)$  provides a very good surrogate for  $\phi(\mathbf{d}^*|\mathbf{X})$ . We show that the proposed method can obtain designs that perform very well in terms of the  $\phi(\mathbf{d}^*|\mathbf{X})$ -value with much less CPU time. Our observation of cases with different stimulus types and different  $\tau_{ISI}$  consistently show the usefulness of the proposed method. Thus, we infer further that this method will provide optimal designs for experiments with much faster scanning time and also can be extended to other optimality criteria, such

as  $D$ -optimality criterion.

In the second part, we extend the existing approaches to obtain optimal design for high temporal resolution neuro-imaging studies when HRF is modeled by a linear combination of the double-gamma function and its temporal derivative. The proposed parametric function was inspired by Calhoun *et al.* (2004) for estimating the amplitude. We consider a slight modification of the parametric function to accommodate cases where the temporal HRF parameter,  $\theta_t$  is active but  $\theta_c = 0$ . Because the value of the latency parameter or the ratio parameter vector,  $\boldsymbol{\eta}$ , is uncertain, we apply the maximin approach considered in Maus *et al.* (2010), and Kao and Mittelman (2014). We incorporate GA and demonstrate the usefulness of the modified approach through several case studies. The performance of the maximin designs is compared with different designs along with some traditional designs under the  $D$ -optimality criterion. We observe that these designs are relatively robust to the misspecification of the latency parameter vector. The optimal designs obtained by Method IA and Method IB always perform better than the traditional designs. Furthermore, the GA tends to perform well when the parameter space is reduced. For comparison purposes, we choose the  $D$ -optimality criterion because of the simplified result by Lemma 4.16, however, the proposed approach can be further extended to other optimality criteria, such as the  $A$ -optimality criterion. The results show that Method IB helps to obtain maximin designs with better efficiencies and requires significantly less time to obtain the locally optimal designs compared to Method IA. In addition, to check the robustness of the obtained designs from Method IB, we evaluate these designs under the setting of Method IA. We observe that obtaining maximin designs without including  $\theta_c = \theta_0$  can further reduce the computational burden without having a negative effect on the design efficiencies.

## REFERENCES

- Aguirre, G. K., M. G. Mattar and L. Magis-Weinberg, “de bruijn cycles for neural decoding”, *NeuroImage* **56**, 3, 1293–1300 (2011).
- Bandettini, P. A. and R. W. Cox, “Event-related fmri contrast when using constant interstimulus interval: theory and experiment”, *Magnetic Resonance in Medicine* **43**, 4, 540–548 (2000).
- Barker, J. W., A. Aarabi and T. J. Huppert, “Autoregressive model based algorithm for correcting motion and serially correlated errors in fnirs”, *Biomedical optics express* **4**, 8, 1366–1379 (2013).
- Bullmore, E., M. Brammer, S. C. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard and P. Sham, “Statistical methods of estimation and inference for functional MR image analysis”, *Magnetic Resonance in Medicine* **35**, 2, 261–277 (1996).
- Buracas, G. T. and G. M. Boynton, “Efficient design of event-related fMRI experiments using M-sequences”, *Neuroimage* **16**, 3 Pt 1, 801–13 (2002).
- Calhoun, V. D., M. Stevens, G. D. Pearlson and K. Kiehl, “fmri analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms”, *Neuroimage* **22**, 1, 252–257 (2004).
- Cheng, C.-S., M.-H. Kao and F. K. H. Phoa, “Optimal and efficient designs for functional brain imaging experiments”, *Journal of Statistical Planning and Inference* **181**, 71–80 (2017).
- Cheng, C.-S., M.-H. Kao *et al.*, “Optimal experimental designs for fmri via circulant biased weighing designs”, *The Annals of Statistics* **43**, 6, 2565–2587 (2015).
- Chernoff, H., “Locally optimal designs for estimating parameters”, *The Annals of Mathematical Statistics* pp. 586–602 (1953).
- Dale, A. M., “Optimal experimental design for event-related fmri”, *Human brain mapping* **8**, 2-3, 109–114 (1999).
- Efromovich, S. and Z. A. Valdez-Jasso, “Aggregated wavelet estimation and its application to ultra-fast fmri”, *Journal of Nonparametric Statistics* **22**, 7, 841–857 (2010).
- Friston, K. J., P. Fletcher, O. Josephs, A. Holmes, M. Rugg and R. Turner, “Event-related fmri: characterizing differential responses”, *Neuroimage* **7**, 1, 30–40 (1998).
- Friston, K. J., A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. Williams, R. S. Frackowiak and R. Turner, “Analysis of fMRI time-series revisited”, *Neuroimage* **2**, 1, 45–53 (1995).



- Friston, K. J., E. Zarahn, O. Josephs, R. Henson and A. M. Dale, “Stochastic designs in event-related fmri”, *Neuroimage* **10**, 5, 607–619 (1999).
- Glover, G. H., “Deconvolution of impulse response in event-related BOLD fMRI”, *Neuroimage* **9**, 4, 416–29 (1999).
- Harville, D. A., *Matrix Algebra from a Statistician’s Perspective*, Wiley series in probability and mathematical statistics (Springer, Secaucus, 1977).
- Henson, R. and K. Friston, “Convolution models for fmri”, *Statistical parametric mapping: The analysis of functional brain images* pp. 178–192 (2007).
- Henson, R. N., C. J. Price, M. D. Rugg, R. Turner and K. J. Friston, “Detecting latency differences in event-related bold responses: application to words versus nonwords and initial versus repeated face presentations”, *Neuroimage* **15**, 1, 83–97 (2002).
- Hofmann, M. J., M. J. Herrmann, I. Dan, H. Obrig, M. Conrad, L. Kuchinke, A. M. Jacobs and A. J. Fallgatter, “Differential activation of frontal and parietal regions during visual word recognition: an optical topography study”, *Neuroimage* **40**, 3, 1340–1349 (2008).
- Huppert, T. J., “Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy”, *Neurophotonics* **3**, 1, 010401–010401 (2016).
- Huppert, T. J., S. G. Diamond, M. A. Franceschini and D. A. Boas, “Homer: a review of time-series analysis methods for near-infrared spectroscopy of the brain”, *Applied optics* **48**, 10, D280–D298 (2009).
- Jang, K.-E., S. Tak, J. Jung, J. Jang, Y. Jeong and Y. C. Ye, “Wavelet minimum description length detrending for near-infrared spectroscopy”, *Journal of biomedical optics* **14**, 3, 034004 (2009).
- Kao, M.-H., “Multi-objective optimal experimental designs for ER-fMRI using MATLAB”, *Journal of Statistical Software* **30**, 11, 1–13 (2009).
- Kao, M.-H., “On the optimality of extended maximal length linear feedback shift register sequences”, *Statistics & Probability Letters* **83**, 6, 1479–1483 (2013).
- Kao, M.-H., “A new type of experimental designs for event-related fmri via hadamard matrices”, *Statistics & Probability Letters* **84**, 108–112 (2014).
- Kao, M.-H., “Universally optimal fmri designs for comparing hemodynamic response functions”, *Statistica Sinica* pp. 499–506 (2015).
- Kao, M.-H., D. Majumdar, A. Mandal, J. Stufken *et al.*, “Maximin and maximin-efficient event-related fMRI designs under a nonlinear model”, *The Annals of Applied Statistics* **7**, 4, 1940–1959 (2013).

- Kao, M.-H., A. Mandal, N. Lazar and J. Stufken, “Multi-objective optimal experimental designs for event-related fMRI studies”, *Neuroimage* **44**, 3, 849–56 (2009).
- Kao, M.-H., A. Mandal and J. Stufken, “Constrained multiobjective designs for functional magnetic resonance imaging experiments via a modified non-dominated sorting genetic algorithm”, *Journal of the royal statistical society: series c (applied statistics)* **61**, 4, 515–534 (2012).
- Kao, M.-H. and H. D. Mittelman, “A fast algorithm for constructing efficient event-related functional magnetic resonance imaging designs”, *Journal of Statistical Computation and Simulation* **84**, 11, 2391–2407 (2014).
- Kao, M.-H. and J. Stufken, *Optimal Design for Event-Related fMRI Studies*, Chapman & Hall/CRC in the Handbook of Design and Analysis of Experiments (Taylor & Francis, 2015).
- Koh, P. H., D. E. Glaser, G. Flandin, S. Kiebel, B. Butterworth, A. Maki, D. T. Delpy and C. E. Elwell, “Functional optical signal analysis: a software tool for near-infrared spectroscopy data processing incorporating statistical parametric mapping”, *Journal of biomedical optics* **12**, 6, 064010–064010 (2007).
- Lazar, N., *The statistical analysis of functional MRI data* (Springer Science & Business Media, 2008).
- Liao, C., K. Worsley, J. Poline, G. Duncan and A. Evans, *Estimating the delay of the hemodynamic response in fMRI data*, Ph.D. thesis, McGill University Libraries (2000).
- Lin, F.-H., J. Ahveninen, T. Raij, T. Witzel, Y.-H. Chu, I. P. Jääskeläinen, K. W.-K. Tsai, W.-J. Kuo and J. W. Belliveau, “Increasing fmri sampling rate improves granger causality estimates”, *PloS one* **9**, 6, e100319 (2014).
- Lin, Y.-L., F. K. H. Phoa, M.-H. Kao *et al.*, “Optimal design of fmri experiments using circulant (almost-) orthogonal arrays”, *The Annals of Statistics* **45**, 6, 2483–2510 (2017).
- Lindquist, M. A., J. M. Loh, L. Y. Atlas and T. D. Wager, “Modeling the hemodynamic response function in fmri: efficiency, bias and mis-modeling”, *Neuroimage* **45**, 1, S187–S198 (2009).
- Lindquist, M. A. and T. D. Wager, “Validity and power in hemodynamic response modeling: a comparison study and a new approach”, *Human brain mapping* **28**, 8, 764–784 (2007).
- Lindquist, M. A. *et al.*, “The statistical analysis of fmri data”, *Statistical science* **23**, 4, 439–464 (2008).
- Liu, T. T., “Efficiency, power, and entropy in event-related fmri with multiple trial types: Part ii: design of experiments”, *Neuroimage* **21**, 1, 401–413 (2004).

- Liu, T. T. and L. R. Frank, “Efficiency, power, and entropy in event-related fmri with multiple trial types: Part i: Theory”, *Neuroimage* **21**, 1, 387–400 (2004).
- Liu, T. T., L. R. Frank, E. C. Wong and R. B. Buxton, “Detection power, estimation efficiency, and predictability in event-related fmri”, *Neuroimage* **13**, 4, 759–773 (2001).
- Logothetis, N. K. and B. A. Wandell, “Interpreting the bold signal”, *Annu. Rev. Physiol.* **66**, 735–769 (2004).
- Maus, B., G. J. P. van Breukelen, R. Goebel and M. P. F. Berger, “Robustness of optimal design of fMRI experiments with application of a genetic algorithm”, *Neuroimage* **49**, 3, 2433–43 (2010).
- Pinti, P., A. Merla, C. Aichelburg, F. Lind, S. Power, E. Swingler, A. Hamilton, S. Gilbert, P. W. Burgess and I. Tachtsidis, “A novel glm-based method for the automatic identification of functional events (aide) in fnirs data recorded in naturalistic environments”, *NeuroImage* **155**, 291–304 (2017).
- Plichta, M., S. Heinzl, A.-C. Ehlis, P. Pauli and A. Fallgatter, “Model-based analysis of rapid event-related functional near-infrared spectroscopy (fnirs) data: a parametric validation study”, *Neuroimage* **35**, 2, 625–634 (2007).
- Plichta, M., M. Herrmann, C. Baehne, A.-C. Ehlis, M. Richter, P. Pauli and A. Fallgatter, “Event-related functional near-infrared spectroscopy (fnirs): are the measurements reliable?”, *Neuroimage* **31**, 1, 116–124 (2006).
- Proulx, S., M. Safi-Harb, P. LeVan, D. An, S. Watanabe and J. Gotman, “Increased sensitivity of fast bold fmri with a subject-specific hemodynamic response function and application to epilepsy”, *NeuroImage* **93**, 59–73 (2014).
- Purdon, P. L., V. Solo, R. M. Weisskoff and E. N. Brown, “Locally regularized spatiotemporal modeling and model comparison for functional {MRI}”, *NeuroImage* **14**, 4, 912 – 923 (2001).
- Rencher, A. C. and G. B. Schaalje, *Linear models in statistics* (Wiley-Inter science, Hoboken, N.J., 2008).
- Scholkman, F., S. Kleiser, A. J. Metz, R. Zimmermann, J. M. Pavia, U. Wolf and M. Wolf, “A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology”, *Neuroimage* **85**, 6–27 (2014).
- Schroeter, M. L., M. M. Bcheler, K. Miller, K. Uluda, H. Obrig, G. Lohmann, M. Tittgemeyer, A. Villringer and D. von Cramon, “Towards a standard analysis for functional near-infrared imaging”, *NeuroImage*. **21**, 1, 283–290 (2004).
- Seber, G. A. F., *Linear regression analysis*, Wiley series in probability and mathematical statistics (Wiley, New York, 1977).
- Tak, S., M. Uga, G. Flandin, I. Dan and W. Penny, “Sensor space group analysis for fnirs data”, *Journal of neuroscience methods* **264**, 103–112 (2016).

- Tak, S. and J. C. Ye, “Statistical analysis of fnirs data: a comprehensive review”, *Neuroimage* **85**, 72–91 (2014).
- Uga, M., I. Dan, T. Sano, H. Dan and E. Watanabe, “Optimizing the general linear model for functional near-infrared spectroscopy: an adaptive hemodynamic response function approach”, *Neurophotonics* **1**, 1, 015004 (2014).
- Villringer, A. and U. Dirnagl, “Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging.”, *Cerebrovascular and brain metabolism reviews* **7**, 3, 240–276 (1995).
- Villringer, A., J. Planck, C. Hock, L. Schleinkofer and U. Dirnagl, “Near infrared spectroscopy (nirs): a new tool to study hemodynamic changes during activation of brain function in human adults”, *Neuroscience letters* **154**, 1-2, 101–104 (1993).
- Wager, T. D. and T. E. Nichols, “Optimization of experimental design in fmri: a general framework using a genetic algorithm”, *Neuroimage* **18**, 2, 293–309 (2003).
- Wager, T. D., A. Vazquez, L. Hernandez and D. C. Noll, “Accounting for nonlinear bold effects in fmri: parameter estimates and a model for prediction in rapid event-related studies”, *NeuroImage* **25**, 1, 206–218 (2005).
- Worsley, K. and J. Taylor, “Detecting fmri activation allowing for unknown latency of the hemodynamic response”, *Neuroimage* **29**, 2, 649–654 (2006).
- Worsley, K. J. and K. J. Friston, “Analysis of fMRI time-series revisited—again”, *Neuroimage* **2**, 3, 173–81 (1995).
- Worsley, K. J., C. Liao, J. Aston, V. Petre, G. Duncan, F. Morales and A. Evans, “A general statistical analysis for fmri data”, *Neuroimage* **15**, 1, 1–15 (2002).
- Ye, J. C., S. Tak, K. E. Jang, J. Jung and J. Jang, “Nirs-spm: statistical parametric mapping for near-infrared spectroscopy”, *Neuroimage* **44**, 2, 428–447 (2009).