Maximin Designs for Event-Related fMRI with Uncertain Error Correlation

by

Amani Alrumayh

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

Approved November 2019 by the
Graduate Supervisory Committee:

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

ARIZONA STATE UNIVERSITY

December 2019

ABSTRACT

One of the premier technologies for studying human brain functions is the event-related functional magnetic resonance imaging (fMRI). The main design issue for such experiments is to find the optimal sequence for mental stimuli. This optimal design sequence allows for collecting informative data to make precise statistical inferences about the inner workings of the brain. Unfortunately, this is not an easy task, especially when the error correlation of the response is unknown at the design stage. In the literature, the maximin approach was proposed to tackle this problem. However, this is an expensive and time-consuming method, especially when the correlated noise follows high-order autoregressive models. The main focus of this dissertation is to develop an efficient approach to reduce the amount of the computational resources needed to obtain A-optimal designs for event-related fMRI experiments. One proposed idea is to combine the Kriging approximation method, which is widely used in spatial statistics and computer experiments with a knowledge-based genetic algorithm. Through case studies, a demonstration is made to show that the new search method achieves similar design efficiencies as those attained by the traditional method, but the new method gives a significant reduction in computing time. Another useful strategy is also proposed to find such designs by considering only the boundary points of the parameter space of the correlation parameters. The usefulness of this strategy is also demonstrated via case studies. The first part of this dissertation focuses on finding optimal event-related designs for fMRI with simple trials when each stimulus consists of only one component (e.g., a picture). The study is then extended to the case of compound trials when stimuli of multiple components (e.g., a cue followed by a picture) are considered.

Dedicated

to my beloved parents

Mrs. Zubaida Mustaghfir and Mr. Ahmed Alrumayh.

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my academic advisor, Dr. Ming- Hung Kao for his guidance, patience, and support during my Ph.D. journey. Without his knowledge and encouragement, this dissertation would not have been possible.

I would also like to extend my deepest appreciation to my committee members: Dr. Mark Reiser, Dr. John Stufken, Dr. Rong Pan and Dr. Dan Cheng for their valuable suggestions and helpful comments.

Thanks are also due to Northern Border University in Saudi Arabia and the Saudi Arabian Cultural Mission (SACM) for providing me a full scholarship to pursue my Ph.D. studies in the U.S.

I am forever thankful to my wonderful friends: Elham Alkhadar, Valeria Benson-Lira, and Nahee Kim for their understanding and encouragement throughout this long journey. Special thanks to my friend and colleague in the Statistics program, Hazar Khogeer for her frequent help and support. She was the first person to encourage me to go for a Ph.D. and has been like a sister to me since I came to the U.S.

I am heartily thankful to my brother, Khaled Alrumayh. This dissertation would not have been possible without the inspiration and support of him. Thank you for being there when I needed you the most.

Last, but not least, I want to thank my family: my mother, Zubaidah Mustaghfir; my father, Ahmed Alrumayh; my sisters, Sarah, Huda, Afrah, Shahad; and my little brother, Sultan, for their unconditional love, support, and care.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

Studying human brain functions has been an interest to many scientists in different research areas such as psychology, neuroscience, education, and medical studies. One of the most widely used technique in neuroimaging is the functional magnetic resonance imaging (fMRI). This technique identifies what parts of the brain are being used when doing a task or making a decision, and helps to study some characteristics of brain activity. The safety of this non-invasive technology makes it a useful tool to study the inner workings of the human brain to provide insights into some mental disorders such as schizophrenia and Alzheimer's diseases.

The use of event-related fMRI designs has become very common in practice (e.g., Amaro et al., 1999; Buckner, 1998; Buračas & Boynton, 2002; Dale, 1999; D'Esposito, Zarahn, & Aguirre, 1999; Friston et al., 1998; Friston, Zarahn, Josephs, Henson, & Dale, 1999; Josephs, Turner, & Friston, 1997; Kao, Mandal, & Stufken, 2008). This type of advanced fMRI designs is flexible, and allows stimuli of short duration (several milliseconds to a few seconds) to be used instead of traditional designs which require longer presentation time (usually more than 1 minute). Several advantages of using event-related designs for fMRI studies are discussed with some examples such as in Buckner (1998), Henson (2003), Wager and Nichols (2003), Huettel (2012), and Kashou (2014). In this dissertation, the interest is in experiments with event-related fMRI designs.

Before conducting an event-related fMRI experiment, the experimenter prepares a long sequence of mental stimuli that determines the onset times and orders of the stimuli of one or more types. In many cases, these mental stimuli presented to the

experimental subject could be simple, having only one component (e.g., a photo), or compound, consisting of two or more components (e.g., a cue followed by a photo). The primary focus of this research is to find designs that help to obtain informative data about brain functions in response to simple mental stimuli as well as compound mental stimuli.

When analyzing event-related fMRI experiments, two statistical objectives are commonly considered, which are the detection of specific brain regions associated with experimental stimuli, and the estimation of the hemodynamic response function (HRF). The HRF is a function of time describing the noise-free change of the strength of the magnetic field following a stimulus onset. The investigator often studies this function to gain knowledge of the effects of the stimulus to the brain. Several studies in the literature have shown that an fMRI design that is good for one study objective might perform poorly for another objective (Liu & Frank, 2004). When attempting to overcome this issue, a useful tool based on the genetic algorithm technique was developed by Kao, Mandal, Lazar, and Stufken (2009) for obtaining good multi-objective designs. This tool helps researchers to efficiently search over the enormous design space of event-related fMRI and has shown to outperform the previous methods proposed in the literature. It has been applied in obtaining efficient event-related designs for fMRI experiments (e.g., Eck, Kaas, & Goebel, 2013; Kao, 2014; Kubilius, Wagemans, & Op de Beeck, 2011), and adapted in tackling several important design issues (e.g., Kao, Majumdar, Mandal, & Stufken, 2013; Kao & Zhou, 2017; Maus, Van Breukelen, Goebel, & Berger, 2010; Maus, van Breukelen, Goebel, & Berger, 2012).

In this research, we adapt the genetic algorithm of Kao et al. (2009) to accommodate a more complicated situation when the noise is assumed to follow 2nd- or 3rd-order autoregressive models; note that the focus of most previous fMRI design studies

is on AR(1) models. The $AR(2)$ process has been recommended to model the temporal correlation in fMRI data (Lenoski, Baxter, Karam, Maisog, & Debbins, 2008), and it often provides better results compared to $AR(1)$ and $ARMA$ models (Lindquist, 2008). Other complicated autocorrelation structures such as $AR(p), p \geq 2$, were discussed in (Worsley et al., 2002) and were shown to be useful when analyzing the correlated noise in fMRI data.

In contrast to some previous studies, where the value of the correlation coefficient is assumed at the design stage, we target designs that are robust against a misspecification of error correlation. In other words, our goal is to find a design that is relatively efficient for all possible values of the unknown correlation coefficients. In particular, we consider the maximin approach, which is also discussed in Berger and Tan (2004). The idea is to work on the worst-case scenario to find a design that maximizes the minimum relative efficiency (RE). The RE of a design $\boldsymbol{d}$ in this case is defined as its relative value of the optimality criterion such as A-optimality or D-optimality to be maximized, to a locally optimal design for a given parameter vector value. The minimum RE is then taken over all the possible values of correlation coefficients, and the design goal is to identify a design maximizing this minimum RE. Maus et al. (2010) also considered the maximin approach and assumed the noise to follow the first-order autoregressive, $AR(1)$, model. However, their method is computationally expensive and time-consuming. As a consequence, this approach requires much computational effort when considering higher-order autoregressive models, such as $AR(p), p \geq 2$.

Alternatively, we propose to use the Kriging approach with a Gaussian process model that is widely used in spatial statistics and computer experiments (Santner, Williams, & Notz, 2003). In computer experiments, the Kriging approach gives a surrogate model that replaces the more expensive computer simulation models by using a Gaussian process to approximate the output 'surface' with a small set of input

values of the computer model. In this research, we apply Kriging to approximate the surface of the RE-values of each design over the parameter space of the correlation coefficients using a small, finite subset of the parameter space. For obtaining such a small subset, we consider space-filling designs such as the well-known Sobol sequences (Sobol', 1967). We would like to emphasize that the space-filling designs here is a sampling plan for subsampling the parameter space, and they are employed in our proposed approach for obtaining a maximin fMRI design that is a sequence of stimuli. Such a space-filling design spreads points evenly throughout the specified region of interest (i.e. the parameter space in our case). Another interesting finding is the usefulness of using a subset of correlation coefficient values located on the boundaries of the parameter space of $AR(p)$. We noticed this method improves the speed to search for the desired maximin designs with no need to apply the Kriging model. We will call this latter approach as the boundary points approach in this dissertation. The proposed Kriging approach and the boundary points approach are compared to the expensive method by Maus et al. (2010). In terms of the statistical efficiency of the achieved designs, the results are quite similar among the three methods. Our proposed methods require much less CPU time and are very efficient in obtaining good maximin designs for event-related fMRI experiments with uncertain error correlations.

The remainder of the dissertation is organized as the following order.

In Chapter 2, we provide background information about event-related fMRI studies, including the commonly used generalized linear models in the analysis of such experiments and the design optimality criteria for estimating the HRF and detecting the active brain regions. A brief discussion of some standard designs that are known to be highly efficient when studying fMRI experiments also provided in this chapter. Chapter 2 also includes some relevant discussions on computer experiments, space-filling designs, such as Latin hypercube designs, uniform designs, and quasi-Monte Carlo

sampling methods such as Sobol sequences. An overview of the Gaussian process and the ordinary Kriging model are provided at the end of that chapter.

Chapter 3 presents the issue of misspecification of error correlations at the design stage. The chapter gives detailed explanation of our proposed Kriging-based genetic algorithm approach and the use of the boundary points approach. Different case studies for obtaining event-related fMRI designs with simple trials when the correlated noise is assumed to follow $AR(2)$ and $AR(3)$ models are presented there. The previously mentioned approaches for obtaining maximin designs are compared in these case studies.

Chapter 4 introduces the concept of compound trials in event-related fMRI experiments. The main models and design selection criteria for the case of compound studies are presented. The results of different case studies are reported at the end of Chapter 4. Finally, Chapter 5 provides a summary of the main findings and some possible extensions for future research.

Chapter 2

RELEVANT BACKGROUND

2.1 Event-Related fMRI Studies

2.1.1 Introduction

To perform an event-related fMRI experiment, as a first step, a design sequence consisting of mental stimuli of one or more types separated by periods of rest or visual fixation is prepared. An example of an event-related fMRI design with two stimulus types can be written as a sequence of finite integers of length N, e.g., $\boldsymbol{d} = \{1012....0\}$. A positive integer ($q = 1, 2, ..., Q$) indicates the presentation of a qth-type stimulus; here, Q is the total number of stimulus types. A 0 means there is no stimulus presentation at the corresponding time point. The nth element in $\boldsymbol{d}$ corresponds to time $(n-1)\tau_{ISI}$ of the experiment, where τ_{ISI} is a pre-specified inter-stimulus interval ; e.g., $\tau_{ISI} = 4$ seconds. We note that time 0 of the experiment can be set to the time when the first valid fMRI measurement is collected. Each stimulus appears for a short time (e.g., 1 second). The selected sequence is then presented to an experimental subject during the experiment. While the subject is cognitively engaging with the stimuli, the fMRI scanner scans the subject's brain at a fixed sampling rate of τ_{TR} seconds to obtain the subject's blood oxygenated level-dependent (BOLD) signals from each brain voxel (a three-dimensional image unit). Specifically, we observe an fMRI time series from each brain voxel.

The fluctuations of the collected BOLD signals following a brief stimulus reflects the change in the concentration of oxygenated cerebral blood due to the neuronal activity in response to the mental stimulus. This fluctuation can take about 32 seconds

(Henson & Friston, 2007) to return to baseline, and it often is described by a function of time called the hemodynamic response function (HRF). An example of a typical HRF in response of a brief stimulus as shown in Figure 2.1. The HRF is of utmost im-

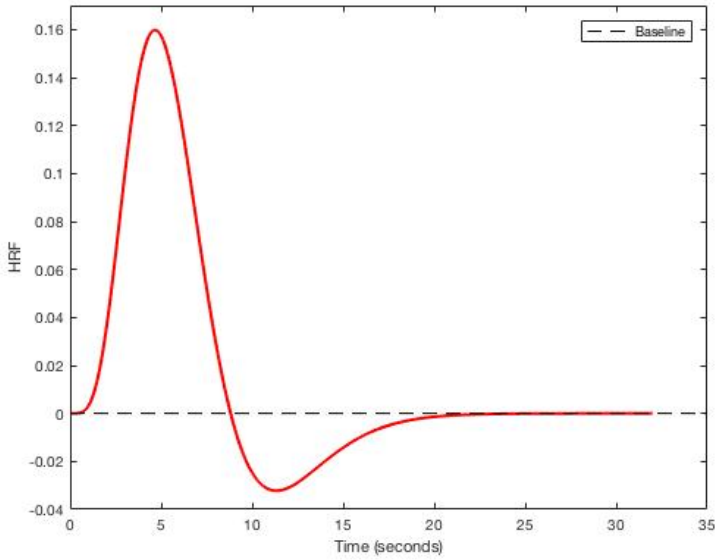


Figure 2.1: A Typical HRF Evoked by a Brief Stimulus.

portance to neuroscientists for understanding the inner workings of the human brain. One of the main assumptions considered when analyzing fMRI data is that for the same voxel, the mental stimuli of the same type presented at different time points will have the same HRF (Lindquist, 2008). In addition, the heights of overlapping HRFs accumulate additively when multiple stimuli are presented in a time interval shorter than the duration of the HRF as shown in Figure 2.2. We note that this additivity assumption may not be valid in the case of very short intervals between stimuli (i.e., $\tau_{ISI} < 2$ seconds) (e.g., Dale, 1999; Dale & Buckner, 1997). Typically, the fMRI time series obtained from a voxel can be expressed as a combination of overlapping HRFs, a possible systematic drift and noise.

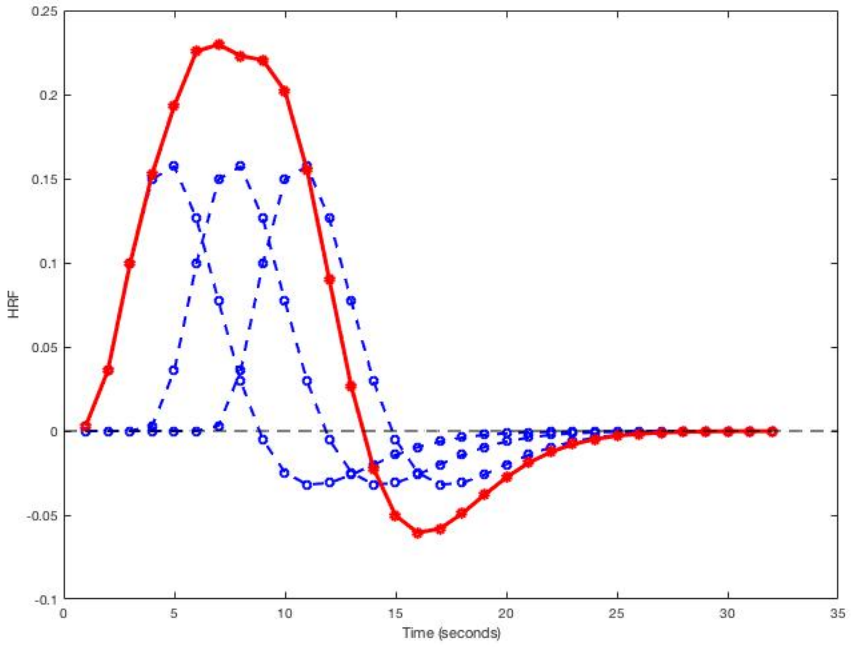


Figure 2.2: The Accumulated HRF (Red Curve) Formed by Three HRFs (Blue Curves) Evoked by the Same Stimulus Type Presented at Different Time Points.

When analyzing event-related fMRI data, the most common statistical objectives, which are considered in this study, include identifying the brain regions that are activated by a mental task and estimating the HRF evoked by each stimulus type. To obtain good designs to help efficiently achieve these objectives, one may also need to consider some unwanted psychological confounds (e.g., anticipation and habituation) that could have a significant impact on the analysis results. Moreover, additional practical constraints that an experimenter might have (e.g., a given stimulus frequency in the design) may also need to be considered at the design stage. To account for all these aspects, a good multi-objective design that strikes the right balance among the competing objectives is needed. An excellent tool based on the genetic algorithm technique was developed by Wager and Nichols (2003) to search for such optimal multiple-objective designs. They formulated the multi-objective criterion as

a weighted sum of the design criteria for common objectives of interest. This genetic algorithm approach is then improved by Kao et al. (2009).

For experiments involving simple trials (i.e. each stimulus involves only one mental task), some popular fMRI designs were introduced in the literature for different study objectives. For instance, m-sequences or systematically computed maximum length sequences (Buračas & Boynton, 2002) are usually recommended for the estimation purposes, due to the low autocorrelation property of these designs. An m-sequence design is known to exists only when Q + 1 (Q represents the number of stimulus types) is a prime or prime power, and the design length is $N = (Q + 1)^r$ for some positive integer r. Moreover, random designs were discussed in Buxton et al. (2000) for being suitable to use when the interest is in estimating the HRF. But, the statistical efficiency achieved by some random designs might not be as high as that of the m-sequence. On the other hand, block designs, where stimuli of the same type are grouped together in the same block, yield high detection power (Friston et al., 1999), and these designs can attain good statistical efficiencies for detecting activated brain regions (when ignoring the possible psychological confounders, including anticipation and habituation). But, block designs perform very poorly for estimating the HRF (Dale, 1999). Clustered m-sequences, permuted block designs, and mixed designs, which are combinations of block designs and random designs (or m-sequences) can be considered in cases when both estimation and detection are of interest (Liu & Frank, 2004). However, there are other efficient designs based on algorithmic approaches that have shown to outperform these traditional designs for different study objectives; see, e.g., Wager and Nichols (2003), Kao et al. (2009), and Saleh, Kao, and Pan (2017).

2.1.2 General Linear Models for Event-Related fMRI Studies

One of the commonly used approaches when analyzing event-related fMRI time series is the voxel-wise general linear model approach (Dale, 1999). In this case, the collected time series from different brain voxels are analyzed separately. For the objective of estimating the HRF, we need first to define the set of HRF parameters associated with the heights of the HRF. This can be done by using the discretization interval $\tau_{\Delta T}$, where $\tau_{\Delta T}$ is the greatest time making both $m\tau_{ISI} = \tau_{ISI}/\tau_{\Delta T}$ and $m\tau_{TR} = \tau_{TR}/\tau_{\Delta T}$ integers. $\tau_{\Delta T}$ may or may not be an integer. We note that the heights of the HRF that contribute to the observed event-related fMRI signals occur at zero seconds and every $\tau_{\Delta T}$ seconds following stimulus onset.

The general linear model (GLM) that is commonly used for the estimation of the HRF is:

$$\boldsymbol{y} = \sum_{q=1}^{Q} \boldsymbol{X}_q \boldsymbol{h}_q + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\boldsymbol{y} = ((y_t))_{t=1,...,T}$ is a $T \times 1$ vector represents the BOLD time series from a brain voxel, $\boldsymbol{h}_q = (h_{q1}, ..., h_{qk})'$ is the HRF parameter vector for the qth-type stimulus, where h_{qj}, $j = 1, 2, ..., k$ corresponds to the jth height of the HRF evaluated $(j-1)\tau_{\Delta T}$ seconds after the onset of the stimulus. We define k as $1 + \lfloor H/\tau_{\Delta T} \rfloor$, where $\lfloor a \rfloor$ is the greatest integer less than or equal to a, and H represents the duration of the HRF (e.g., 32 seconds), counting from the onset of the stimulus and ending at the return of the HRF to baseline. $\boldsymbol{X}_q$ is the $T \times k$, 0-1 design matrix for the qth-type stimulus, where the $(t, j)th$ element of $\boldsymbol{X}_q$ is 1 when h_{qj} contributes to y_t and 0 otherwise. $\boldsymbol{S}\boldsymbol{\gamma}$ describes the trend or drift of the time series with $\boldsymbol{S}$ being a pre-specified matrix and $\boldsymbol{\gamma}$ an unknown parameter vector. $\boldsymbol{\varepsilon}$ is a $T \times 1$ vector that represents the correlated noise and assumed to be normally distributed with mean 0 and covariance matrix $\boldsymbol{\Sigma}$ corresponding to the stationary $AR(p)$ structure with unknown correlation coefficients

parameters $\boldsymbol{\phi}$. A known whitening matrix $\boldsymbol{V}$ is assumed such that $\boldsymbol{V\Sigma V'} = \boldsymbol{I}_T$, $\boldsymbol{I}_T$ is $T \times T$ identity matrix. Note that, the error variance σ^2 is set to equal 1 throughout this dissertation since it does not affect the selection of the optimal design $\boldsymbol{d}$.

For illustrative purposes, we provide two examples to demonstrate the construction of the design matrix $\boldsymbol{X}_q$ when $\tau_{ISI} = \tau_{TR}$ and $\tau_{ISI} \neq \tau_{TR}$.

Example 2.1: Assume $\tau_{ISI} = \tau_{TR} = 2$ seconds, hence $\tau_{\Delta T} = 2$ seconds, and $m_{ISI} = m_{TR} = 1$) . Let the design be $\boldsymbol{d} = \{12012...0\}$ with $Q = 2$, and $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2]$ is the $T \times Qk$, 0-1 design matrix where each column of $\boldsymbol{X}$ corresponds to a parameter vector in $\boldsymbol{h} = (\boldsymbol{h}_1', \boldsymbol{h}_2')'$, and each row in $\boldsymbol{X}$ corresponds to an fMRI scan. If the duration of the HRF is 32 seconds, then $k = 1 + \lfloor 32/2 \rfloor = 17$. We then generate two indicator vectors $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ that have the same length as the design $\boldsymbol{d}$, where $\boldsymbol{\delta}_1 = (10010...0)'$ represents the locations of the first stimulus type in $\boldsymbol{d}$ and $\boldsymbol{\delta}_2 = (01001...0)'$ for the locations of the second stimulus type as occurred in $\boldsymbol{d}$. Then we can write the first column of $\boldsymbol{X}_1$ as $\boldsymbol{\delta}_1$ and the first column of $\boldsymbol{X}_2$ is the same as $\boldsymbol{\delta}_2$. To obtain the following columns, we shift the previous column one position down and then add zeros at the top. The design matrices for the two stimulus types can be written as follows:

$$\boldsymbol{X_1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \end{bmatrix}; \boldsymbol{X_2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots & \end{bmatrix}.$$

A value of 1 at the first row of $\boldsymbol{X}_1$ for instance, represents the first height of the first stimulus type that contributes to the first fMRI scan. The 1's at the fourth row of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ can be interpreted as the first and fourth heights of the HRF of the first stimulus type and the third height of the HRF of the second stimulus type that contribute to the fourth fMRI scan. We note that, the design matrix $\boldsymbol{X}$ can be written as the concatenation of the design matrices for each stimulus type.

Example 2.2: Assume $\tau_{ISI} = 3$ seconds, $\tau_{TR} = 2$ seconds, thus $\tau_{\Delta T} = 1$, $m_{\tau_{ISI}} = 3$, $m_{\tau_{TR}} = 2$ and $k = 1 + \lfloor 32/1 \rfloor = 33$. Assume the design with one stimulus type is $d = \{1011...0\}$. We obtain the design matrix $\boldsymbol{X}$ by following these steps:

1. Construct $\boldsymbol{d}^*$ by adding $(m_{\tau_{ISI}} - 1)$ 0's between any two consecutive elements in design $\boldsymbol{d}$. For this example, $m\tau_{ISI} - 1 = 2$ and $\boldsymbol{d}^* = \{100000100100...0\}$.

2. Use $\boldsymbol{d}^*$ to construct the expanded design matrix $\boldsymbol{X}^*$. The number of rows in $\boldsymbol{X}^*$ is the same as the length of $\boldsymbol{d}^*$ and the number of columns is k. Therefore, the first column in $\boldsymbol{X}^*$ is $\boldsymbol{d}^*$ and the following columns are obtained by shifting the previous column one position down then a zero at the top. Note that $\boldsymbol{X}^*$ is constructed as if ${\tau_{ISI}}^* = \tau_{\Delta T} = 1$ second.

3. The design matrix $\boldsymbol{X}$ is then obtained from the expanded matrix $\boldsymbol{X}^*$ by taking the rows that correspond to the scanning time, which are the row numbers $(j-1)\tau_{TR}/\tau_{\Delta T} + 1, j = 1, ..., T$.

The expanded design matrix $\boldsymbol{X}^*$ and the final form for the design matrix $\boldsymbol{X}$ can be written as follows:

$$\boldsymbol{X}^* = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 1 & 0 & 0 & 1 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}; \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ 1 & 0 & 0 & 1 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}.$$

Following Kao et al. (2008), the formula to construct the design matrix $\boldsymbol{X}_q$ for this case is:

$$\boldsymbol{X}_q = [\boldsymbol{I}_T \otimes (1, \boldsymbol{0}'_{m\tau_{TR}-1})][\boldsymbol{w}_q, \boldsymbol{B}\boldsymbol{w}_q, ..., \boldsymbol{B}^{k-1}\boldsymbol{w}_q], \tag{2.2}$$

where $\otimes$ is the Kronecker product, $\boldsymbol{0}_b$ is the $b \times 1$ zero vector, $\boldsymbol{w}_q$ is defined as $\boldsymbol{w}_q = \boldsymbol{\delta}_q \otimes [1, \boldsymbol{0}'_{m\tau_{ISI}-1}]'$, $\boldsymbol{\delta}_q$ as mentioned previously is the indicator vector represents the location of type q stimulus in a $\boldsymbol{d}$. The matrix $\boldsymbol{B}$ is defined as:

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{0}'_{T-1} & 0 \\ \boldsymbol{I}_{T-1} & \boldsymbol{0}_{T-1} \end{bmatrix}.$$

For the objective of identifying the activated regions in the brain in response to mental stimuli, a commonly considered model has the form of:

$$\boldsymbol{y} = \sum_{q=1}^{Q} \boldsymbol{X}_q \boldsymbol{h}^* \theta_q + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{2.3}$$

where $\boldsymbol{h}^* = (h_1^*, ..., h_k^*)'$ is a $k \times 1$ vector representing the assumed shape of the HRF, and θ_q represents the unknown response amplitude of the HRF for the qth-type stimulus. In this model, the HRF is expressed as the product of a reference waveform (basis function) and an unknown HRF amplitude (Kao & Mittelmann, 2014). We note that a large θ value indicates a voxel that is greatly activated. Here, The term $\boldsymbol{X}_q \boldsymbol{h}^*$ represents the discretized convolution of the stimuli with the assumed basis $\boldsymbol{h}^*$ of the HRF and $\boldsymbol{\eta}$ represents the error that follows $AR(p)$ structure as $\boldsymbol{\varepsilon}$ in model (2.1). The remaining terms in (2.2) are as in (2.1).

The basis $\boldsymbol{h}^*$ is commonly set to the canonical HRF of the widely used software package SPM for fMRI data analysis. The canonical HRF is defined as the following double gamma function:

$$g(t) = \frac{t^5 e^{-t}}{5!} - \frac{1}{6}\frac{t^{15} e^{-t}}{15!}. \tag{2.4}$$

Specifically, $g(t)$ is a combination of two gamma probability density functions. To construct $\boldsymbol{h}^*$, $g(t)$ is first normalized to have a maximum value of 1; i.e., $g(t)/max_s g(s)$. The elements of $\boldsymbol{h}^*$ are then the values of the normalized $g(t)$ at spaced time points, starting from $t = 0$. Here, we assume the same $\boldsymbol{h}^*$ for all the Q stimulus types. It is also possible to allow different shapes of HRF for different stimulus types.

2.1.3 Design Selection Criteria

In almost all experiments, the quality of the collected data depends on the selected design. In this dissertation, a design said to be an optimal design when it allows the most precise least square estimates of the parametric functions of interest. For the estimation objective, let $\boldsymbol{h} = (\boldsymbol{h}_1', ..., \boldsymbol{h}_Q')'$ represent the unknown parameter vector of interest; it contains the Q HRF parameter vectors defined in (2.1). Following Kiefer (1959), the efficiency of a design $\boldsymbol{d}$ in estimating $\boldsymbol{h}$ is evaluated by some functions of the following information matrix of $\boldsymbol{h}$:

$$\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi}) = \boldsymbol{X}'[\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}]\boldsymbol{X}. \tag{2.5}$$

where $\boldsymbol{\Sigma}(\boldsymbol{\phi}) = cov(\boldsymbol{\varepsilon})$ and $\boldsymbol{\phi}$ represents the correlation coefficients in $\boldsymbol{\Sigma}$. Similarly, for the objective of detecting the activated brain voxels, we set $\boldsymbol{\theta} = (\theta_1, ..., \theta_Q)'$ to represent the unknown amplitudes for the Q HRFs, and $\boldsymbol{\Sigma}(\boldsymbol{\phi}) = cov(\boldsymbol{\eta})$. The quality of a design $\boldsymbol{d}$ in estimating $\boldsymbol{\theta}$ is evaluated by some functions of the information matrix of the form:

$$\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi}) =$$
$$(\boldsymbol{I}_Q \otimes \boldsymbol{h}^*)'\boldsymbol{X}'[\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}]\boldsymbol{X}(\boldsymbol{I}_Q \otimes \boldsymbol{h}^*), \tag{2.6}$$

$\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi})$ in models (2.5) and (2.6) depends on the design $\boldsymbol{d}$ through the design matrix $\boldsymbol{X}$, so we aim to find a design $\boldsymbol{d}$ that in some sense, maximizes $\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi})$ to precisely estimate the desired parametric functions. Additionally, the interest might be on estimating some linear combinations of the HRF heights such as $\boldsymbol{C}_h\boldsymbol{h}$, or some linear combinations of the response amplitudes $\boldsymbol{C}_\theta\boldsymbol{\theta}$.

Two popular criteria proposed in the literature of event-related fMRI studies to measure the design efficiency, are A-optimality (e.g., Buračas & Boynton, 2002; Dale, 1999; Friston et al., 1999; Liu & Frank, 2004) and D-optimality (e.g., Kao et al.,

2009; Maus et al., 2010). A-optimal designs aim at minimizing the average variance of estimators of parametric functions, while D-optimal designs minimize the volume of the (asymptotic) ellipsoidal confidence region of linearly independent parametric functions. The A- and D- optimality criteria can respectively be written as:

$$F_i(\boldsymbol{d};\boldsymbol{\phi}) = \begin{cases} m/trace(\boldsymbol{M}^{-1}(\boldsymbol{d};\boldsymbol{\phi})), & for A-optimality \\ det(\boldsymbol{M}(\boldsymbol{d};\boldsymbol{\phi}))^{1/m}, & for D-optimality \end{cases}$$

where m denotes the size of $\boldsymbol{M}(\boldsymbol{d};\boldsymbol{\phi})$, $F_i(\boldsymbol{d};\boldsymbol{\phi}) = F_e(\boldsymbol{d};\boldsymbol{\phi})$ denotes the value of the design criterion for estimating the HRF; this value is referred to as estimation efficiency, and $F_i(\boldsymbol{d};\boldsymbol{\phi}) = F_d(\boldsymbol{d};\boldsymbol{\phi})$ denotes the value of the design criterion for the objective of detection; this value is referred to as detection power. The values of $F_e(\boldsymbol{d};\boldsymbol{\phi})$ and $F_d(\boldsymbol{d};\boldsymbol{\phi})$ are set to zero when the designs do not allow the parameters of interest to be estimable. These optimality criteria are defined as the larger-the-better criteria.

2.2 Some Designs for the Kriging Approach

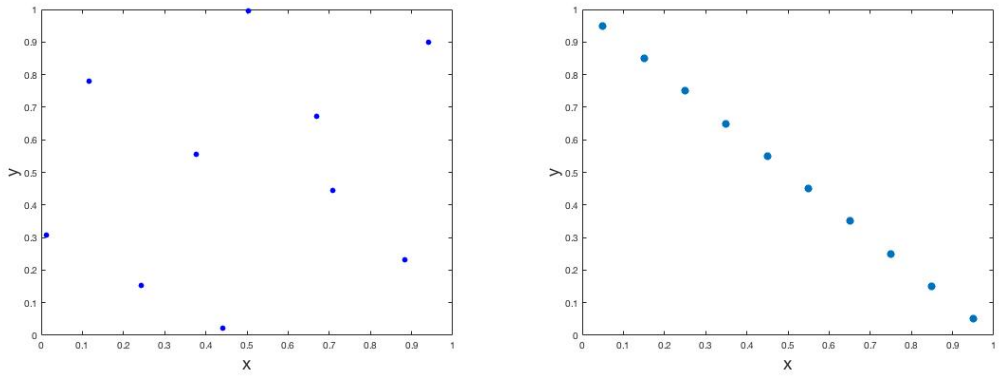
The Kriging approach will be applied to facilitate the search of good event-related fMRI designs. A key step to the success of this approach is by selecting a good design (sampling plan). In the following subsections, we discuss some possible designs. We then briefly introduce the Gaussian process to be considered in our approach.

2.2.1 Space-Filling Designs in Computer Experiments

To develop an efficient surrogate model that helps to predict the outcomes over a region of interest, it is useful to select a good design that determines the locations in the region of interest for making observations. The design discuss in this subsection is a sampling plan used to select a finite sub-set of the region of interest to explore

the output surface. An important application of these designs is in computer experiments (Santner et al., 2003). The key issue in the design and analysis of computer experiments is how to choose a small number of input values for generating a set of informative computer outputs. These computer outputs may then be used to predict the outputs of the untried input values via some statistical models (e.g., Gaussian processes). One possible solution is to consider designs that have a space-filling property where the selected inputs cover the input space as uniformly as possible. We note that a space-filling design without replication is often considered in computer experiments, especially when the same input values will produce identical output. We briefly discuss some common space-filling designs in the literature.

One of the most commonly used space-filling designs in computer experiments is the Latin hypercube design (LHD) which was first introduced by McKay, Beckman, and Conover (1979). Fang, Li, and Sudjianto (2006) represent a LHD with n runs and s variables as a matrix of size $n \times s$ where each column is a random permutation of $1, 2, ..., n$. Possibly after some transformations, each row in this matrix is then used to determine the coordinates of a selected point in the s-dimensional cube. Since there are a large number of column permutations, the number of possible LHDs is large too, and a random selection of LHD does not guarantee to get a design that fills the space uniformly. As an example, an LHD can have all sampled points aligned on the diagonal part of the design space as shown in Figure 2.3(b). Another example of a LHD with 10 runs (design points) and 2 factors x_1, x_2 in the unit square $[0, 1]^2$ is presented in Figure 2.3(a). We may generate LHDs by optimizing some criteria. For instance, the maximin distance criterion, which was proposed by Morris and Mitchell (1995) helps to ensure a good space-filling property of the resulting LHD and avoid the problem of getting a bad design as the one in Figure 2.3 (b).



(a) Space-Filling LHD (b) Non Space-Filling LHD

Figure 2.3: Examples of LHDs With 10 Runs and 2 factors

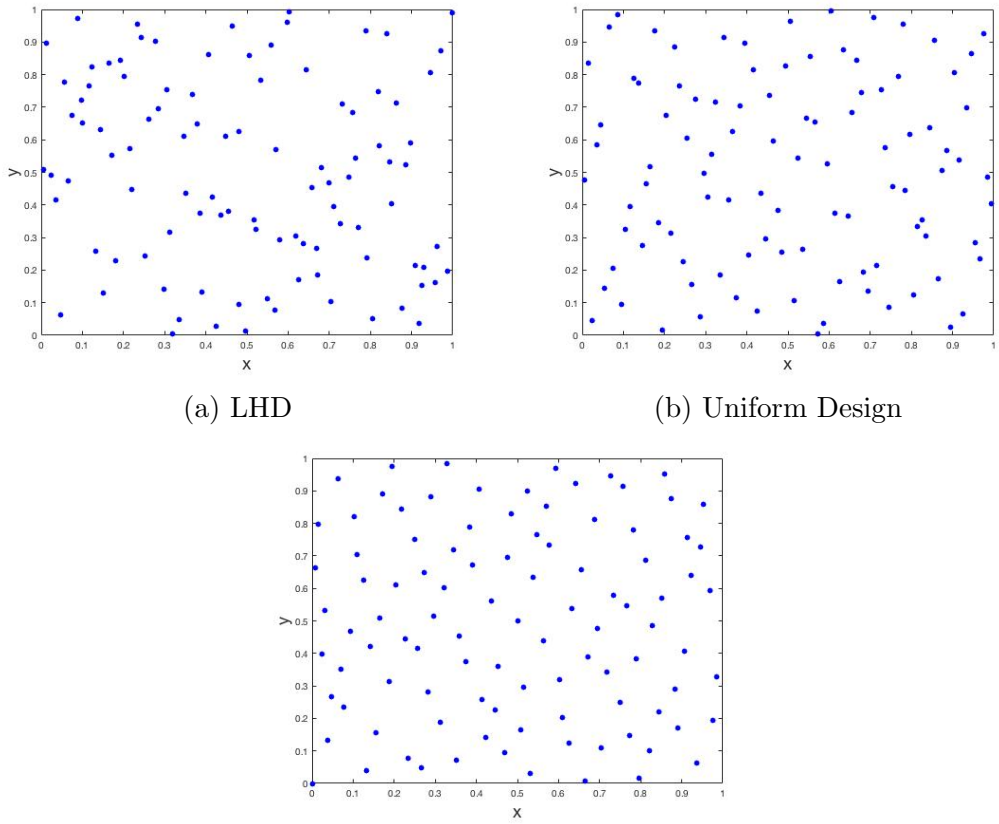
Uniform designs (Fang 1980; Fang, Lin, Winker, and Zhang 2000) are another common type of space-filling designs where the uniformity of a design is measured by comparison against uniform distribution using discrepancy measures (e.g., the star discrepancy L_∞ or the centered L_2 discrepancy). The lower the discrepancy the better uniformity the set of points has. An example of a uniform design $U_n(n^s)$ with $n = 100$ runs and $s = 2$ factors measured by centered L_2 discrepancy is presented in Figure 2.4(b); for more details see Chapter 3 of Fang et al. (2006).

Additionally, distance-based designs such as maximin and minimax distance designs are known for their space-filling property (Johnson, Moore, & Ylvisaker, 1990). Maximin distance designs are more popular in practice than Minimax designs partly because the latter designs are hard to compute (Santner et al., 2003). Constrained maximin designs were proposed in several studies (e.g., Stinstra, den Hertog, Stehouwer, & Vestjens, 2003; Trosset, 1999) for the case of non-rectangular regions.

2.2.2 Quasi-Monte Carlo Sampling Methods

Another way to obtain space-filling designs is to consider the quasi-Monte Carlo sampling methods. The quasi-Monte Carlo (QMC) simulation uses deterministic se-

quences, i.e., low discrepancy sequences, instead of pseudorandom numbers that are used in the traditional Monte Carlo (MC) simulation. There are many advantages of using the QMC sampling methods in computer experiments; one of them is that quasi-random numbers fill the design space better than pseudorandom numbers because of the consideration of the previously sampled points which leads to better coverage of the space by avoiding the form of clusters or gaps. Sobol sequences (Sobol', 1967) are among the most popular and easy to implement quasi-random sequences. Figure 2.4 shows how space-filling designs generated by Sobol sequences covered the unit square space as uniformly as possible compared to Latin hypercube and uniform designs.



(a) LHD

(b) Uniform Design

(c) Sobol Sequences

Figure 2.4: Examples of 2 Factors and 100 Runs Space-Filling Designs for LHD, Uniform, and Sobol Sequences.

Sobol sequences are generated by using primitive polynomials over the Galois field $(GF)_2 = \{0, 1\}$. Following Bratley and Fox (1988), and Chi and Mascagni (2004), we provide a brief explanation of how to generate such sequences. Suppose we would like to generate a sequence of n points $x_1, x_2, ..., x_n$ in the s-dimensional space $[0, 1]^s$. Then for each point $x_i, i = 1, 2, ..., n$, we can write $x_i = (x_{i1}, x_{i2}, ..., x_{is})$, where $x_{ij}, j = 1, ..., s$, represents the ith component in the jth dimension, which can be generated by the following recursive equation:

$$x_{ij} = b_1 v_{1j} \oplus b_2 v_{2j} \oplus b_3 v_{3j} \oplus ...., \tag{2.7}$$

where $\oplus$ is the bitwise exclusive-OR operator (the output is 1 only when inputs differ and 0 otherwise), and b_k is the kth digit from the right in the binary representation of a non-negative integer n (i.e., $n = (...b_3 b_2 b_1)_2$). v_{kj} is the kth direction number in the jth dimension, which can be defined as a binary fraction in the form $v_{kj} = m_{kj}/2^k$, where m_k is an odd integer and $m_k < 2^k$. In order to generate these direction numbers we need first to select a primitive polynomial of degree q in GF_2 which can be expressed as:

$$x^q + a_1 x^{q-1} + a_2 x^{q-2} + ... + a_{q-1} x + 1. \tag{2.8}$$

Here, the coefficients a_i are either 0 or 1 and can be used to obtain the direction numbers v_{kj} using the following recurrence relation:

$$v_{kj} = a_1 v_{(k-1)j} \oplus a_2 v_{(k-2)j} \oplus .... \oplus a_{q-1} v_{(k-q+1)j} \oplus v_{(k-q)j} \oplus v_{(k-q)j}/2^q, i > q, \tag{2.9}$$

To ease the implementation of v_{kj}, we can re-write the previous recurrence relation in terms of m_{kj} as follows:

$$m_{kj} = 2a_1 m_{(k-1)j} \oplus 2^2 a_2 m_{(k-2)j} \oplus .... \oplus 2^{q-1} a_{q-1} m_{(k-q+1)j} \oplus 2^q m_{(k-q)j} \oplus m_{(k-q)j}. \tag{2.10}$$

After choosing the primitive polynomial of degree q, we can choose the initial values of $m_1, m_2, ..., m_q$ to be any odd integers that satisfy $m_i < 2^i$, then we can obtain the values of $m_{q+1}, m_{q+2}, ...$ by using model (2.10).

Another common method in constructing Sobol sequences is the Gray code by Antonov and Saleev (1979), where the binary coefficients $....b_3b_2b_1$ in the recursive equation (2.7) are replaced by the Gray code representation of the nonegative integer n. The updated recursive equation in this case is:

$$x_{ij} = g_1 v_{1j} \oplus g_2 v_{2j} \oplus g_3 v_{3j} \oplus ...., \tag{2.11}$$

where $....g_3g_2g_1 = (....b_3b_2b_1)_2 \oplus (...b_4b_3b_2)_2$. Since the gray code for n and $n+1$ differ in only one position, then we can generate Sobol points recursively by the following recursive equation:

$$x_{(i+1)j} = x_{ij} \oplus v_{cj}, \tag{2.12}$$

where c is the index of the first zero-bit from the right in binary representation of $n = (...b_3b_2b_1)_2$. This is a faster and more efficient way to generate Sobol sequences. More examples are discussed in (Joe & Kuo, 2008).

In Kucherenko, Albrecht, and Saltelli (2015) a comparison was made between QMC approach based on Sobol sequences and other sampling methods, such as LHD and regular Monte Carlo methods. They concluded that Sobol sequences satisfied the uniformity properties (Properties A and A'), that were first introduced by Sobol (1976), in low and high dimensional problems up to 20 dimensions better than other sampling methods that performed only well in low dimensions. In addition, Kucherenko et al. (2015) have also shown that Sobol sequences are in many aspects superior to the other sampling schemes, particularly when prediction is of primary interest. In this study, we consider the stationary $AR(p)$, parameter space which usually presents a non-rectangular shape under the constraints on the parameter val-

ues. We will explore the use of the previously introduced designs to subsample the correlation parameter values from such a parameter space.

2.2.3 The Gaussian Process

In the past decade, there has been an explosive growth in the field of computer experiments to design products for modern industry. Several reasons for using computer experiments as a replacement to the traditional physical experiments were discussed in the literature. A major one is that physical experiments are usually too costly and for some cases they are infeasible to implement (Fang et al., 2006; Jones & Johnson, 2009).

The idea of computer experiments is to use a computer model (a simulator) to generate the data in order to study the input-output relationship. Typically, these are deterministic simulators, which means no matter how many times we run the simulation we will get the same output under the same set of input conditions. However, these computer experiments are based on some complicated mathematical models that sometimes require hours or even days to finish a single run.

A surrogate model that approximates the real one with limited numbers of runs to produce an output is needed (Santner et al., 2003). One of the well-known surrogate models (i.e., metamodels) that are widely used in the spatial statistics literature is the Kriging model. The idea of Kriging was first proposed by geologist D.G. Krige in the 1950s and was then developed by other geologists to predict the location of valuable minerals over a large area of interest given a small set of sites from that area (Matheron, 1963). Sacks, Welch, Mitchell, and Wynn (1989) have introduced Kriging into computer experiments. This method has then gained much popularity in approximating deterministic computer models and optimization purposes for its ability to interpolate the observed data points (Martin & Simpson, 2005). Briefly

speaking, Kriging is a method to build an approximation of a target function from a given finite set of evaluations of the function. The technique is also known as Gaussian process regression.

A Kriging model can be defined as a generalized linear regression model that accounts for the correlation in the residuals between the regression model and the observations (Goldberger, 1962). The mathematical form of the Kriging model can be written as:

$$\boldsymbol{w}(\boldsymbol{\phi}) = \sum_{k=1}^{p} f_k(\boldsymbol{\phi})\beta_k + Z(\boldsymbol{\phi}), \tag{2.13}$$

where $\boldsymbol{w}(\boldsymbol{\phi})$ represents the output from the simulator for the given input $\boldsymbol{\phi} \in \mathbb{R}^n; \boldsymbol{\phi} = (\phi_1, ..., \phi_n)$. $f_1(.), f_2(.), ..., f_p(.)$ are regression functions, $\beta_1, ..., \beta_p$ are regression parameters, and $\boldsymbol{Z}(.)$ corresponds to a stationary Gaussian process with mean zero and covariance $Cov(Z(\boldsymbol{\phi}_i), Z(\boldsymbol{\phi}_j)) = \sigma^2 R(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$, where σ^2 is the variance and $R(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j)$ is the spatial correlation function between the n-dimensional inputs $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$. One of the most popular spatial correlation functions and is given by the following formula:

$$R(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j) = \exp\{-\alpha|\boldsymbol{\phi}_i - \boldsymbol{\phi}_j|^2\}. \tag{2.14}$$

The correlation function parameter α controls the smoothness of the surface, where $\alpha = 0$ means the correlation is 1 and the surface is flat while a large α value corresponds to a low correlation and yields a rather rough surface (Jones & Johnson, 2009). Several methods were discussed in Santner et al. (2003) to estimate the correlation parameter α in R such as the maximum likelihood, restricted maximum likelihood and cross validation methods. According to Martin and Simpson (2005), the maximum likelihood method provides good estimations of the Kriging model parameters in some cases.

Suppose that the training set of data is obtained at given input sites $\boldsymbol{\phi}_1, ...\boldsymbol{\phi}_s$, and the resulting outputs are $\boldsymbol{W} = (\boldsymbol{w}(\boldsymbol{\phi}_1), ..., \boldsymbol{w}(\boldsymbol{\phi}_s))'$. Given these sampled outputs, a

linear predictor of the output at another location $\boldsymbol{\phi}_0$ can be expressed as:

$$\hat{\boldsymbol{w}}(\boldsymbol{\phi}_0) = \sum_{i=1}^{s} \lambda_i \boldsymbol{w}(\boldsymbol{\phi}_i) = \boldsymbol{\lambda}' \boldsymbol{W}, \tag{2.15}$$

where $\boldsymbol{\lambda}$ can be selected by minimizing the mean square error (MSE) for prediction,

$$MSE[\hat{\boldsymbol{w}}(\boldsymbol{\phi}_0)] = E[\hat{\boldsymbol{w}}(\boldsymbol{\phi}_0) - \boldsymbol{w}(\boldsymbol{\phi}_0)]^2, \tag{2.16}$$

under the unbiasedness constraint,

$$E[\hat{\boldsymbol{w}}(\boldsymbol{\phi}_0) - \boldsymbol{w}(\boldsymbol{\phi}_0)] = 0. \tag{2.17}$$

If $\boldsymbol{\lambda}$ solves the minimization problem of Eq. (2.16) subject to the unbiasedness constraint of Eq. (2.17), then $\boldsymbol{\lambda}'\boldsymbol{W}$ is called the best linear unbiased predictor (BLUP) of $\boldsymbol{w}(\boldsymbol{\phi}_0)$. By solving for $\boldsymbol{\lambda}$ and substituting into Eq. (2.15), the BLUP of $\boldsymbol{w}(\boldsymbol{\phi}_0)$ is given by

$$\hat{\boldsymbol{w}}(\boldsymbol{\phi}_0) = \boldsymbol{f}'(\boldsymbol{\phi}_0)\hat{\boldsymbol{\beta}} + \boldsymbol{r}'(\boldsymbol{\phi}_0)\boldsymbol{R}^{-1}(\boldsymbol{W} - \boldsymbol{F}\hat{\boldsymbol{\beta}}), \tag{2.18}$$

where $\boldsymbol{f}'(\boldsymbol{\phi}_0) = [f_1(\boldsymbol{\phi}_0), f_2(\boldsymbol{\phi}_0), ..., f_p(\boldsymbol{\phi}_0)]'$, $\boldsymbol{F}$ is the expanded $s \times p$ matrix of regressors having the *(i,k)th* element as $f_k(\boldsymbol{\phi}_i)$ for $1 \leq i \leq s, 1 \leq k \leq p$, $\boldsymbol{r}'(\boldsymbol{\phi}_0)$ is the $s \times 1$ vector of correlations between the sample points and an untried point $\boldsymbol{\phi}_0$, which is defined as $\boldsymbol{r}'(\boldsymbol{\phi}_0) = \{R(\boldsymbol{\phi}_0, \boldsymbol{\phi}_1), ..., R(\boldsymbol{\phi}_0, \boldsymbol{\phi}_s)\}$. $\boldsymbol{R}$ is the $s \times s$ correlation matrix of $\boldsymbol{W}$, which is composed of spatial correlation functions evaluated at each possible combination of the known points. $\hat{\boldsymbol{\beta}}$ is the generalized least-squares estimator of $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)$ which can be written as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}'\boldsymbol{R}^{-1}\boldsymbol{W}. \tag{2.19}$$

The Gaussian model in (2.18) can interpolates the training data points. To show this, we set $\boldsymbol{\phi}_0 = \boldsymbol{\phi}_i$ for some fixed i, $1 \leq i \leq s$, then $\boldsymbol{f}'(\boldsymbol{\phi}_0) = \boldsymbol{f}'(\boldsymbol{\phi}_i)$ and $\boldsymbol{r}'(\boldsymbol{\phi}_i) =$

$\{R(\boldsymbol{\phi}_i, \boldsymbol{\phi}_1), ..., R(\boldsymbol{\phi}_i, \boldsymbol{\phi}_{\boldsymbol{s}})\}$ which corresponds to the ith row of $\boldsymbol{R}$. Hence the product of $\boldsymbol{r}'(\boldsymbol{\phi}_i)$ and $\boldsymbol{R}^{-1}$ is the unit vector $\boldsymbol{e}_i' = (0, ..., 0, 1, 0, ..., 0)$. This leads to:

$$\hat{\boldsymbol{w}}(\boldsymbol{\phi}_i) = \boldsymbol{f}'(\boldsymbol{\phi}_i)\hat{\boldsymbol{\beta}} + \boldsymbol{e}_i'(\boldsymbol{W} - \boldsymbol{F}\hat{\boldsymbol{\beta}}) = \boldsymbol{f}'(\boldsymbol{\phi}_i)\hat{\boldsymbol{\beta}} + \boldsymbol{W}(\boldsymbol{\phi}_i) - \boldsymbol{f}'(\boldsymbol{\phi}_i)\hat{\boldsymbol{\beta}} = \boldsymbol{w}(\boldsymbol{\phi}_i). \quad (2.20)$$

One of the most commonly used forms of Kriging is the ordinary Kriging model that has shown to provide satisfactory results in many previous studies. An ordinary Kriging model is a special case of model (2.13) by taking $p = 1$ and $f_1(\boldsymbol{\phi}) = 1$. Consequently, the mathematical form of the ordinary Kriging model is:

$$\boldsymbol{w}(\boldsymbol{\phi}) = \beta + Z(\boldsymbol{\phi}), \quad (2.21)$$

and for this case the BLUP of $\boldsymbol{w}(\boldsymbol{\phi}_0)$ is given by

$$\begin{aligned} \hat{\boldsymbol{w}}(\boldsymbol{\phi}_0) &= \hat{\beta} + \boldsymbol{r}'(\boldsymbol{\phi}_0)\boldsymbol{R}^{-1}(\boldsymbol{W} - \boldsymbol{j}_s\hat{\beta}), \\ \hat{\beta} &= (\boldsymbol{j}_s'\boldsymbol{R}^{-1}\boldsymbol{j}_s)^{-1}\boldsymbol{j}_s'\boldsymbol{R}^{-1}\boldsymbol{W}. \end{aligned} \quad (2.22)$$

where $\boldsymbol{j}_s$ is a vector of ones of length s, and the remaining terms are as defined earlier in (2.18). We will use the ordinary Kriging model to facilitate the search for maximin event-related fMRI designs.

Chapter 3

SIMPLE TRIALS EVENT-RELATED FMRI EXPERIMENTS

3.1 Misspecification of Error Correlation Problem

Generally, data acquired from the same experimental subject in an event-related fMRI experiments are known to be correlated. In the voxel-wise analysis, models such as (2.1) and (2.2) allow the BOLD responses obtained from the same voxel to be temporally dependent. The correlated noise is assumed to follow some parametric correlation models, such as the autoregressive models of orders $1, 2$, and $p \geq 2$ (e.g., Lenoski et al., 2008; Lindquist, 2008; Worsley et al., 2002) and ARMA models (e.g., Locascio, Jennings, Moore, & Corkin, 1997; Purdon, Solo, Weisskoff, & Brown, 2001). When obtaining optimal designs, it is important to take the correlation of the BOLD responses into account.

Some studies in the literature assumed that the correlation among the measurements in an event-related fMRI time series could be specified at the design stage. For instance, in Kao et al. (2009), the first-order autoregressive process $AR(1)$ is considered to address the temporal correlation and the correlation coefficient fixed to $\phi = 0.3$. This assumption is not always valid in practice. Maus et al. (2010) claimed that the correlation coefficient can vary between experiments and also from voxel to voxel. Consequently, a design that is good for a given correlation of a voxel might be inefficient for another voxel. Therefore, obtaining a good design that works relatively well for all possible error correlation values is crucially important. To this end, Maus et al. (2010) used the maximin approach (Berger & Tan, 2004) to work on the worst-case scenario and found robust designs against misspecification of the

error correlations. However, their method has shown to be very expensive and quickly becomes infeasible when considering a more complicated such as $AR(p), p \geq 2$. As the use of $AR(p)$ for some $p \geq 2$ is not uncommon in practice, an efficient approach is needed to allow experimenters to efficiently obtain a maximin design that can perform relative well across the parameter space of the p autocorrelation parameters.

In the following sections, we explain in details the maximin design approach considered by Maus et al. (2010) and compare it to our proposed method. As to be explained later, we borrow a useful statistical method from the analysis of computer experiments to develop our proposed method to efficiently obtain maximin designs.

3.2 The Maximin Criterion

In this dissertation, we assume that the error correlations in models (2.1) and (2.2) follow $AR(p), p = 2, 3$, which are not uncommon for the fMRI settings that we consider. Following Maus et al. (2010), we would like to obtain robust designs that have high overall efficiencies over a set of possible values of the correlation coefficients; that is to obtain a maximin design $\boldsymbol{d}^*_{Mm}$ that maximizes the following criterion:

$$\min_{\boldsymbol{\phi} \in \boldsymbol{\Omega}} RE(\boldsymbol{d}; \boldsymbol{d}^*_{\boldsymbol{\phi}}) = \min_{\boldsymbol{\phi} \in \boldsymbol{\Omega}} \frac{F_i(\boldsymbol{d}; \boldsymbol{\phi})}{F_i(\boldsymbol{d}^*_{\boldsymbol{\phi}}; \boldsymbol{\phi})} \tag{3.1}$$

where $\boldsymbol{\Omega}$ represents the parameter space that consists of all possible values of the autocorrelation parameters ϕ. $F_i(.)$ is the optimality criterion (e.g., A-optimality) where $i = e$ for estimation or $i = d$ for detection (see also, Subsection 2.1), $\boldsymbol{d}$ denotes the candidate design being evaluated and $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ represents the locally optimal design which is the best design for the given $\boldsymbol{\phi}$ value. To obtain $\boldsymbol{d}^*_{Mm}$, we may consider the following steps:

- Step 1: Find $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ that maximizes $F_i(\boldsymbol{d}; \boldsymbol{\phi})$ for a large set of values of $\boldsymbol{\phi} \in \boldsymbol{\Omega}$.

- Step 2: For each candidate design $\boldsymbol{d}$, use these $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ to calculate its $RE(\boldsymbol{d}; \boldsymbol{d}^*_{\boldsymbol{\phi}}) = F_i(\boldsymbol{d}; \boldsymbol{\phi})/F_i(\boldsymbol{d}^*_{\boldsymbol{\phi}}; \boldsymbol{\phi})$ values, then find the minimal value of the obtained REs over the set of the $\boldsymbol{\phi}$-values in Step 1.

- Step 3: Find a design $\boldsymbol{d}^*_{Mm}$ that yields the maximal value of $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d}; \boldsymbol{d}^*_{\boldsymbol{\phi}})$.

Note that, when the minimum RE-value of a design is very close to 1, this design performs well for the different values of $\boldsymbol{\phi}$. In addition, Step 2 in the previously described method will need to be repeated for every candidate design and for all possible values of the correlation coefficients. Maus et al. (2010) considered a similar procedure to obtain a maximin design by assuming the $AR(1)$ model with $\Omega = [0, 0.5]$. In particular, they consider a regular grid on Ω, and obtain a locally optimal design for each of the 51 grid points. They then evaluated the $minRE_{\boldsymbol{\phi}\in[0,0.5]}$ of the 51 locally optimal designs, and among these locally optimal designs they then selected the design with the greatest $minRE$ value as $\boldsymbol{d}^*_{Mm}$. In our experience, this method requires much computational effort especially for the case with an autoregressive of order $p \geq 2$, where the parameter space $\boldsymbol{\Omega}$ will be large and finding a design that works well for all possible $\boldsymbol{\phi} \in \boldsymbol{\Omega}$ will be difficult. In the next section, we propose an efficient method to reduce the time needed for finding the desired optimal designs.

3.3 The Proposed Approach

Our idea is on utilizing the Kriging model to reduce the time needed for obtaining maximin event-related fMRI designs that are robust against a misspecification of error correlations. Our proposed method is described below:

- Step 1: Select a small set of $\boldsymbol{\phi}$-values from $\boldsymbol{\Omega}$, which are denoted by $\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_s$ where $\boldsymbol{\phi}_j = (\phi_{1,j}, \phi_{2,j}, ..., \phi_{p,j}); j = 1, ..., s$, for the general $AR(p)$ model. More details on the selection of these $\boldsymbol{\phi}$-values will be provided in the next subsection.

- Step 2: Find $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ that maximizes $F_i(\boldsymbol{d};\boldsymbol{\phi})$ for the selected values of $\boldsymbol{\phi}$.

- Step 3: For each candidate event-related fMRI design $\boldsymbol{d}$, use $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ to calculate the $RE(\boldsymbol{d};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values.

- Step 4: For given $\boldsymbol{d}$, use the BLUP in (2.22) to approximate $\boldsymbol{w}(\boldsymbol{\phi}) = RE(\boldsymbol{d};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ for each $\boldsymbol{\phi}$ on a fine grid of $\boldsymbol{\Omega}$. The $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ is then approximated from these $\boldsymbol{w}(\boldsymbol{\phi})$'s.

- Step 5: Find a design $\boldsymbol{d}^*_{Mm}$ that maximizes the approximated min-RE in Step 4.

To obtain the locally optimal designs in Step 2, and $\boldsymbol{d}^*_{Mm}$ in Step 5, we consider to adapt the genetic algorithm by Kao et al. (2009) (for more details, see the Appendix). Our approach is thus a combination of a Kriging method that helps to approximate the objective function, and the genetic algorithm, that helps to efficiently search over the enormous design space for the needed locally optimal designs $\boldsymbol{d}^*_{\phi}$, and a maximin design $\boldsymbol{d}^*_{Mm}$ optimizing the approximated objective function.

3.3.1 Sampling Methods Over the Parameter Space of the Correlation Coefficients

We first explain how we define the parameter space when the correlation coefficients are assumed to follow $AR(2)$ and $AR(3)$ models, respectively. First, for $AR(2)$, the inverse of the variance-covariance matrix ($\Sigma^{-1} = ((\Sigma^{-1}_{ij}))_{i,j=1,\ldots,T}$) has the following form:

$$\Sigma^{-1}_{i,i} = \begin{cases} 1, & i = 1, T; \\ 1+\phi_1^2, & i = 2, T-1; \\ 1+\phi_1^2+\phi_2^2, & 2 < i < T-1; \end{cases}$$

$$\Sigma^{-1}_{i,j} = \begin{cases} -\phi_1, & |i-j| = 1 \text{ and } i+j = 3, 2T-1; \\ -\phi_1(1-\phi_2), & |i-j| = 1 \text{ and } 3 < i+j < 2T-1; \\ -\phi_2, & |i-j| = 2; \\ 0, & |i-j| > 2. \end{cases}$$

For demonstration purposes, we set both $\rho_1 = \phi_1/(1-\phi_2)$ and $\rho_2 = \phi_1^2/(1-\phi_2)+\phi_2$ to be within the range of $[0, 0.5]$; this interval includes the most commonly observed values of correlation for fMRI data (Maus et al., 2010). Note that, ρ_i stands for the i-th autocorrelation coefficient of the AR(2) process, which describes the correlation between errors at time t and $t+i$, where t=1,2,...,T-i. To satisfy the conditions of stationarity, we have

$$\phi_1 \in \begin{cases} [0, 0.5(1-\phi_2)], & \text{if } \phi_2 \in [0, \frac{1}{3}]; \\ [0, \sqrt{0.5(1-2\phi_2)(1-\phi_2)}], & \text{if } \phi_2 \in (\frac{1}{3}, \frac{1}{2}]. \end{cases}$$

This leads to an irregularly shaped parameter space $\mathbf{\Omega}$ of $AR(2)$, and is the blue shaded area in Figure 3.1.

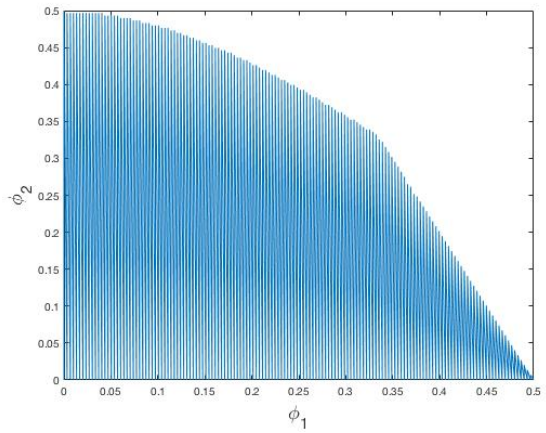


Figure 3.1: The Parameter Space $\mathbf{\Omega}$ of the Stationary $AR(2)$ Model.

For the case when the error correlation is assumed to follow an $AR(3)$ model, the elements of the inverse variance-covariance matrix are defined as:

$$\Sigma_{i,i}^{-1} = \begin{cases} 1, & i = 1, T; \\ 1 + \phi_1^2, & i = 2, T-1; \\ 1 + \phi_1^2 + \phi_2^2, & i = 3, T-2; \\ 1 + \phi_1^2 + \phi_2^2 + \phi_3^2, & 3 < i < T-2; \end{cases}$$

$$\Sigma_{i,j}^{-1} = \begin{cases} -\phi_1, & |i-j| = 1 \text{ and } i+j = 3, 2T-1; \\ -\phi_1(1-\phi_2), & |i-j| = 1 \text{ and } i+j = 5, 2T-3; \\ -\phi_1 + \phi_1\phi_2 + \phi_2\phi_3, & |i-j| = 1 \text{ and } 6 < i+j < 2T-4; \\ -\phi_2, & |i-j| = 2 \text{ and } i+j = 4, 2T-2; \\ -\phi_2 + \phi_1\phi_3, & |i-j| = 2 \text{ and } 5 < i+j < 2T-3; \\ -\phi_3, & |i-j| = 3; \\ 0, & |i-j| > 3. \end{cases}$$

Following Yule-Walker equations, we then have:

$$\rho_1 = \frac{\phi_1}{(1-\phi_2)} + \frac{\phi_3\phi_1(\phi_1+\phi_3) + \phi_3\phi_2(1-\phi_2)}{(1-\phi_2)^2 - \phi_3(\phi_1+\phi_3)(1-\phi_2)},$$

$$\rho_2 = \frac{\phi_1(\phi_1+\phi_3) + \phi_2(1-\phi_2)}{(1-\phi_2) - \phi_3(\phi_1+\phi_3)},$$

$$\rho_3 = \phi_1\rho_2 + \phi_2\rho_1 + \phi_3.$$

Here, we consider the case where ρ_1, ρ_2, ρ_3 are set to be within the range of $[0, 0.5]$. The parameter space for $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$ is then a 3-dimensional irregular space shown in Figure 3.2.

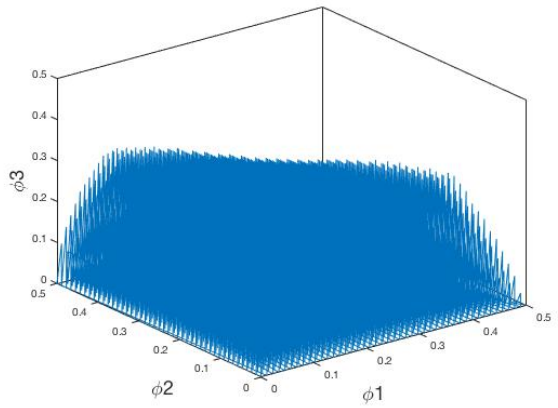


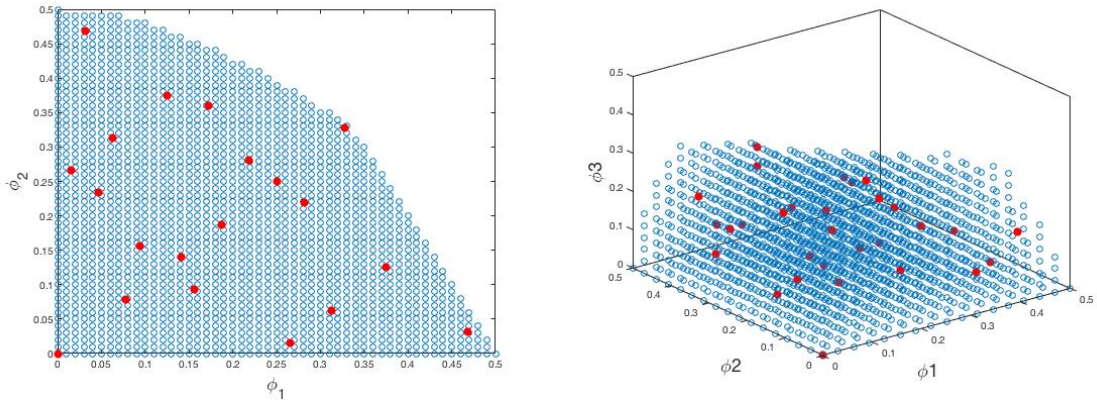
Figure 3.2: The Parameter Space $\boldsymbol{\Omega}$ of the Stationary $AR(3)$ Model.

In contrast to the expensive method by Maus et al. (2010), we propose to approximate the unknown surface of the relative efficiencies $RE(\boldsymbol{d}; \boldsymbol{d}^*_\phi)$ over $\boldsymbol{\Omega}$. This is achieved by using a good sampling plan to subsample $\boldsymbol{\Omega}$ to allow us to explore this surface at different regions of $\boldsymbol{\Omega}$. When choosing a sampling plan we need to take into account its projection (collapsing) properties and the space-filling properties over irregular regions. Sampling plans, such as a regular grid design or the commonly used Latin hypercube designs in computer experiments may lack one of these properties. For instance, points sampled by a regular grid design suffer from the collapsing property even when applied on regular regions, which means when projecting the grid points from d-dimensional space to a lower dimensional space, we will have points that share the same coordinate values, hence, the model will be evaluated repeatedly over the same values. On the other hand, Latin hypercube designs (LHD)s have shown to avoid the collapsing property when used to sample over regular regions. It is guaranteed that there will be only one point at each location whenever we project

the LHD points onto a one- dimensional space. Nevertheless, a major drawback for LHDs is they sometime fail to maintain the space-filling properties and provide bad coverage of the region of interest as shown before in Figure 2.3(b). For irregular regions, LHDs do not guarantee these properties. For example, we might have some big gaps between the selected points after projecting an LHD to a one-dimensional space. Therefore, LHDs are often not recommended for irregular regions. For this reason, we consider the well-known Sobol sequences (Niederreiter, 1988, 1992; Sobol', 1967), to obtain a small set of $\boldsymbol{\phi}$-values as mentioned in Step 1 in Section 3.3. Sobol sequences tend to have good projection and space-filling properties even when the space is irregular. This is because they are constructed so that the selected points that are scattered uniformly over the subregions of different sizes without having a pattern or producing clusters in certain areas.

We follow a rule of thumb to sample 10 observations per dimension. Specifically, we sample 20 points from the specified 2-dimensional space of $AR(2)$ and 30 points from the 3-dimensional space of $AR(3)$ using the Sobol technique. The implementation of such designs is done by using the statistics toolbox function 'Sobolset' in MATLAB to generate a Sobol point set. To do that, an initial point set sequence consisting of 27 points is generated from the regular design space of $AR(2)$; i.e., $[0, 0.5] \times [0, 0.5]$. Then, we eliminate the points outside the region of interest $\boldsymbol{\Omega}$, and keep the remaining 20 Sobol points inside $\boldsymbol{\Omega}$. For the case when the error correlation follows $AR(3)$ model, an initial point set sequence consisting of 65 points is generated from the regular design space $[0, 0.5]^3$ of $AR(3)$, then we eliminate the points outside $\boldsymbol{\Omega}$. These sampled points will be used as inputs in the surrogate model (2.22) to generate a proper approximation of the surface of the $RE(\boldsymbol{d}; \boldsymbol{d}^*_\phi)$-values over $\boldsymbol{\Omega}$ for each candidate event-related fMRI design $\boldsymbol{d}$.

For the expensive method by Maus et al. (2010) where a large set of $\boldsymbol{\phi}$-values is required, we consider the traditional grid sampling plan where points are distributed equidistantly over the region of interest. We set the grid size over the parameter space of $AR(2)$ to 0.01×0.01, and this gives a set of 1797 values of $\boldsymbol{\phi} = (\phi_1, \phi_2)$. For $AR(3)$ model, we set the grid size to 0.03^3, which results in a subset of 1957 points of the corresponding $\boldsymbol{\Omega}$. Those points will then be used to calculate the $RE(\boldsymbol{d}; \boldsymbol{d}^*_\phi)$-values for each design $\boldsymbol{d}$ following the steps in Section 3.2. Figure 3.3 represents the selected Sobol points (red dots) and the grid points (blue dots) over the parameter space of $AR(2)$ and $AR(3)$ respectively.



(a) $\boldsymbol{\Omega}$ for $AR(2)$ Model. (b) $\boldsymbol{\Omega}$ for $AR(3)$ Model.

Figure 3.3: Sampled Points by Sobol Sequences Over the Parameter Space $\boldsymbol{\Omega}$.

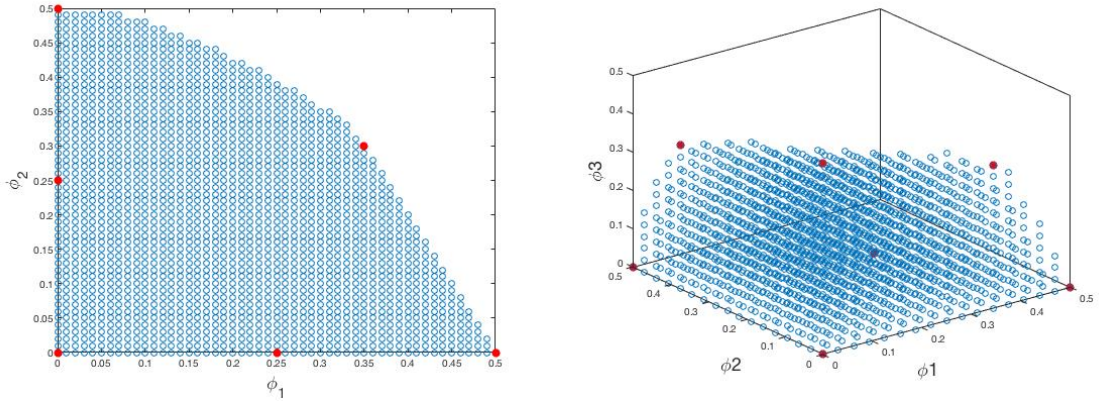
3.3.2 Boundary Points Method

When evaluating the RE-values of the locally optimal designs (LODs) over the parameter space $\boldsymbol{\Omega}$ of $AR(2)$ and $AR(3)$ models, there seems to have a tendency that by moving away form the location where a design is locally optimal, the value of its RE decreases until it reaches its minimum at the boundary of the space. As an example, for the case with an $AR(2)$ model, the locally optimal design for $\boldsymbol{\phi} = (0, 0)$ is evaluated over the fine grid of 1797 points of $\boldsymbol{\phi} = (\phi_1, \phi_2)$, and its min-

RE occurs at points $\boldsymbol{\phi} = (0.31, 0.35)$ and $(0.33, 0.33)$ with 66% and 88% maximin efficiencies for detection and estimation, respectively. Both points are located on the boundaries of $\boldsymbol{\Omega}$ far from the origin. Another example when evaluating LODs that are located at the center of the parameter space $\boldsymbol{\Omega}$ *i.e.*, $\boldsymbol{\phi} = (0.25, 0.25)$, the minimal value among all its RE-values obtained form the fine grid occurs at $(0, 0)$ with 79% maximin efficiency for detection case and at $(0, 0.5)$ with 94% maximin efficiency for estimation, which are also boundary points. In a similar manner, when the error correlation follows an $AR(3)$ model, we evaluate the locally optimal designs for $\boldsymbol{\phi} = (0, 0, 0)$ and $(0.25, 0.25, 0.25)$ over the fine grid of 1957 points of $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$. We noticed that the min-REs occur at points $(0, 0.38, 0.38)$ and $(0, 0, 0)$ for detection case, and at $(0.35, 0.32, 0)$ and $(0, 0.5, 0)$ for estimation case, which are also located at the boundaries of $\boldsymbol{\Omega}$ of $AR(3)$ process. It also is interesting to note that the maximin efficiencies of LODs located at the center of the parameter space of interest are expected to be higher than the maximin efficiencies for LODs located at the boundaries of the space. For instance, the LOD at the origin of $AR(3)$ parameter space has 66% and 88% maximin efficiencies for detection and estimation purposes, and the LOD located at $(0.25, 0.25, 0.25)$ which is at the center of $AR(3)$ parameter space has 78% and 92% maximin efficiencies for detection and estimation purposes. However, in our experience, the LODs obtained at the center points do not guarantee a maximin design, and a search of such a maximin design is needed in many cases.

The previously mentioned phenomenon on the locations where the min-RE occurs is also observed in many other event-related fMRI designs that we studied. We thus treat the boundaries of $\boldsymbol{\Omega}$ as our 'region of interest' for finding maximin designs. Therefore, we consider only to use a small set of $\boldsymbol{\phi}$-values located at the boundaries of $\boldsymbol{\Omega}$ to approximate the $minRE(\boldsymbol{d}; \boldsymbol{d}^*_{\boldsymbol{\phi}})$ of each candidate design $\boldsymbol{d}$. Specifically, this is only done by finding the min-RE over the selected boundary points without the

use of the previously described Kriging approach. A maximin design maximizing this approximated min-RE is then obtained via the genetic algorithm of Kao et al. (2009). We consistently observe that the use of the boundary points requires much less CPU time than all the previously described methods in Sections 2.2 and 2.3, without significantly sacrificing the achieved designs efficiencies. Figure 3.4 shows the set of boundary points (red dots) that we consider for the parameter space $\boldsymbol{\Omega}$ (blue dots) of $AR(2)$ and $AR(3)$ models.



(a) Ω for $AR(2)$ Model. (b) Ω for $AR(3)$ Model.

Figure 3.4: The Selected Boundary Points Over the Parameter Space $\boldsymbol{\Omega}$.

3.4 Case Studies for Simple Trials Event-Related fMRI Experiments

Through simulations, we compare the performance of our proposed approaches with the expensive method of Maus et al. (2010) to find maximin designs for event-related fMRI experiments. We consider designs with $Q = 1, 2$, and 3 stimulus types with their respective design lengths of $N = 255, 242$, and 255. The inter-stimulus interval τ_{ISI} and the time to repetition τ_{TR} are both set to 2 seconds. In this case, $\tau_{\Delta T}$ is also 2 seconds (the greatest time making both $\tau_{ISI}/\tau_{\Delta T}$ and $\tau_{TR}/\tau_{\Delta T}$ integers). For all simulations, the duration of the HRF is set to 32 seconds, the nuisance term $\boldsymbol{S\gamma}$ in models (2.1) and (2.2), is assumed to allow for a second-order polynomial drift in

the fMRI time series. The A-optimality criterion is used for both statistical objectives, namely estimation and detection. We focus on the case where the individual stimulus effects is of interest. We show first the results for the case where the noise is assumed to follow a stationary $AR(2)$ model with uncertain autocorrelation coefficients.

The algorithmic parameters of the genetic algorithm used in the simulations to search for optimal event-related fMRI designs are: G (size of generation) = 20, q (percentage of mutation) = 1%, and N (number of immigrants per generation) = 4. The algorithm is run until a stopping rule is met; i.e., no significant improvement is made. The algorithm keeps track of the best event-related MRI design obtained in each GA generation. We modify the MATLAB program provided by Kao (2009), and combine it with the software package DACE for Kriging approximations (Lophaven, Nielsen, & Søndergaard, 2002). We implement our simulations by using MATLAB on a desktop computer of a 3.7GHz Intel Core i7-8700k 6-core processor with 32GB RAM.

3.4.1 Case I: Estimation

In the first set of case studies, we focus on the study objective of estimating the HRF with model (2.1) where the number of the HRF parameters contained in $\boldsymbol{h}_q$ is $17 = (1 + \lfloor 32/2 \rfloor)$. We compare the following methods:

- Method A (Kriging+Sobol Method): We apply our proposed method as described in Section 3.3 to obtain optimal maximin designs $\boldsymbol{d}^*_{Mm}$. We implement the genetic algorithm of Kao et al. (2009) to generate the locally optimal designs $\boldsymbol{d}^*_{s,\boldsymbol{\phi}}$ that maximize $F_e(\boldsymbol{d};\boldsymbol{\phi}) = m/trace(\boldsymbol{M}^{-1}(\boldsymbol{d};\boldsymbol{\phi}))$ for the 20 selected points in a Sobol sequence from the pre-specified parameter space $\boldsymbol{\Omega}$ of $AR(2)$. We note that, $\boldsymbol{M}(\boldsymbol{d};\boldsymbol{\phi})$ here is defined as in (2.5) and its size is $m = 17Q; Q = 1, 2, 3$. For each candidate design, we then can calculate its RE-values using the 20 locally

optimal designs $\boldsymbol{d}^*_{s,\boldsymbol{\phi}}$. By using the Kriging approximation as in (2.22), these 20 RE-values then help to approximate the surface of the relative efficiencies over the parameter space $\boldsymbol{\Omega}$. This latter approximation is done by using the DACE software package. The default settings of the DACE package are considered. The model has the form of (2.21) and involves a constant mean plus a Gaussian correlation function. The α correlation function parameters in (2.14) are estimated using the maximum likelihood method. The steps described in Section 3.3 are then adopted to search for a maximin design that maximizes the approximated $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d};\boldsymbol{d}^*_{\boldsymbol{\phi}})$.

- Method B (Grid method): Here, we consider the approach of Maus et al. (2010) to achieve maximin designs $\boldsymbol{d}^*_{Mm}$, by considering the $\boldsymbol{\phi}$-values on a fine grid over $\boldsymbol{\Omega}$. For this case, we obtain a locally optimal design $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ for each of the 1797 grid points of $\boldsymbol{\phi} = (\phi_1, \phi_2)$ by implementing the GA of Kao et al. (2009). These locally optimal designs allow to provide an approximated min-RE for each candidate design. We then adapt the GA of Kao et al. (2009) to search for a design $\boldsymbol{d}^*_{Mm}$ that maximizes the minimum RE.

- Method C (Boundary points method): We use the locally optimal designs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ that are optimal for the selected boundary points over the parameter space $\boldsymbol{\Omega}$ of $AR(2)$ as shown in Figure 3.4 (a) to approximate the $minRE(\boldsymbol{d};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ for each candidate design. Then, we adapt the GA of Kao et al. (2009) to search for a design that maximizes the minimum RE.

Table 3.1 represents the time spent to generate the set of the locally optimal designs maximizing $F_e(\boldsymbol{d};\boldsymbol{\phi})$ needed for the three methods. When comparing the maximin designs obtained by Method A and Method C to those obtained by Method B, we evaluate the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by Method A, and Method C

using the 1797 locally optimal designs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ on the fine grid. Table 3.2 provides the $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values of the designs obtained by the three methods for Q=1, 2 and 3.

Table 3.1: CPU Time Spent by the Genetic Algorithm on Maximizing F_e for $Q = 1, 2, 3$ Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(2)$ Model.

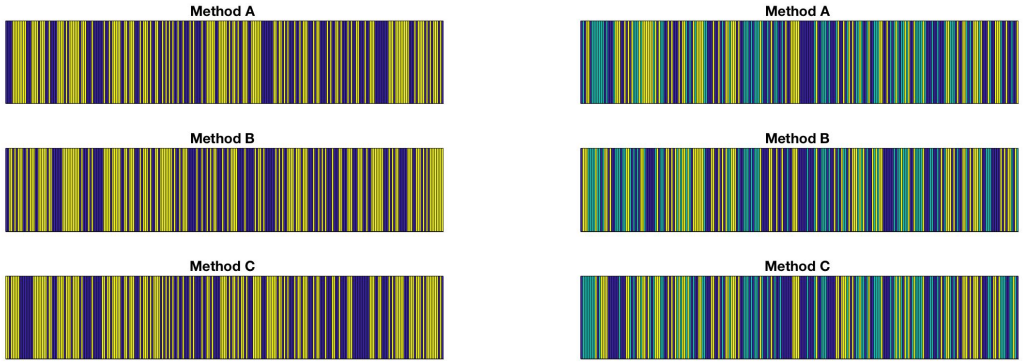
Method	number of LODs	Total time spent (min.): Q=1	Q=2	Q=3
A	20	1.7	3.9	9
B	1797	126	336	786
C	6	0.4	1	1.6

Table 3.2: The Performances of Method A, Method B, and Method C for Estimating the HRF Over the Parameter Space $\boldsymbol{\Omega}$ and the Corresponding (ϕ_1, ϕ_2)-Values That Yield the Maximin Designs $\boldsymbol{d}^*_{Mm}$ for $AR(2)$ Model.

Q	1	2	3
Method A:			
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.9538	0.9546	0.9447
time spent (min.)	1.34	2.2	8.2
(ϕ_1, ϕ_2)	(0,0.5)	(0.5,0)	(0,0.47)
Method B:			
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.9588	0.9593	0.9532
time spent (min.)	156	315	798
(ϕ_1, ϕ_2)	(0.5,0)	(0,0.47)	(0.5,0)
Method C:			
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.9584	0.9552	0.9552
time spent (min.)	0.3	0.5	1.4
(ϕ_1, ϕ_2)	(0,0.5)	(0,0.47)	(0,0.47)

We note that, the reported time in Table 3.2 does not include the time needed for generating the locally optimal designs $\boldsymbol{d}^*_{\phi}$, which can be found in Table 3.1. We observe that the values of $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ for the three methods are quite similar and the obtained designs are very efficient (at least 95% maximin efficiency). From Table

3.2, we also observe a huge reduction in the time needed for generating a maximin design when Method A and Method C are used. In particular, for all cases ($Q = 1, 2, 3$) our proposed methods reduces the CPU time by at least 99%. The proposed methods saved much of the computational resources without sacrificing the efficiency of the obtained designs, hence, they are more efficient than the grid method (Method B). Figure 3.5 represents the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by the three methods for estimating the HRF when $Q = 1$ and 2. Here, different colors represent different event types with dark blue indicating the null or control event.



(a) (Q,N)=(1,255) (b) (Q,N)=(2,242)

Figure 3.5: Maximin designs obtained by Method A, Method B, and Method C for estimating the HRF of $AR(2)$ model.

In addition, deigns obtained by Method A and Method C are compared with some well-known designs in the literature of event-related fMRI studies that have high efficiencies for estimating the HRF under linear models, such as m-sequence and random designs. We generate m-sequence designs following Liu and Frank (2004), and 10 different random designs. The $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.;\boldsymbol{d}^*_{\boldsymbol{\phi}})$ for these deigns are presented in Table 3.3 for each of the following scenarios $(Q, N) = (1, 255), (2, 242)$ and $(3, 255)$. We notice that the maximin designs obtained by Method A (Kriging+Sobol) and the Method C (boundary points method) outperform these traditional designs.

Table 3.3: $minRE_{\phi\in\Omega}(.;\boldsymbol{d}^*_\phi)$ of $\boldsymbol{d}^*_{Mm}$ From Methods A and C Versus Some Traditional fMRI Designs for Estimation Purposes of $AR(2)$ Model.

	Q=1	Q=2	Q=3
$\boldsymbol{d}^*_{Mm,MethodA}$	0.9538	0.9546	0.9447
$\boldsymbol{d}^*_{Mm,MethodC}$	0.9584	0.9552	0.9552
m-sequence-based design	0.8838	0.8241	0.8405
10 random designs	0.6669-0.8711	0.6690-0.7762	0.6785-0.7607

The boxplots in Figure 3.6 display the distribution of the $RE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.;\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values over $\boldsymbol{\Omega}$ for the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by Method A, Method C, an m-sequence design and a random design. Each box gives the 25%, 50%, and 75% percentiles

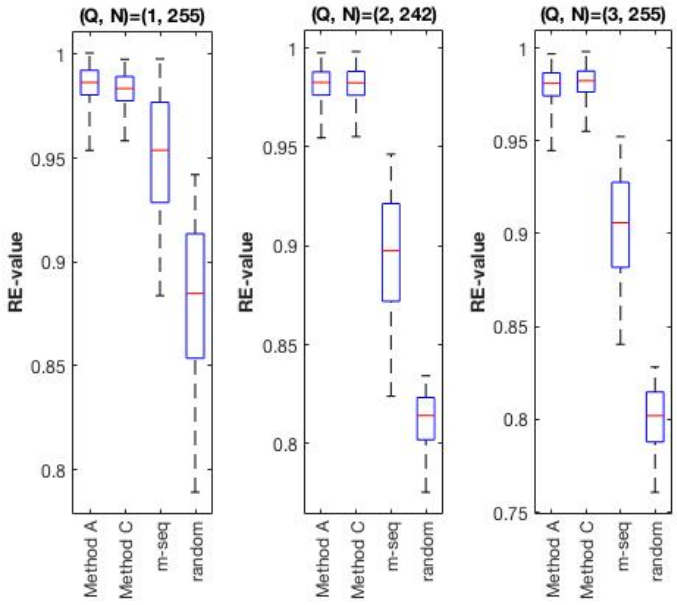


Figure 3.6: Boxplots of RE-Values Over $\boldsymbol{\Omega}$ of $AR(2)$ Model for Designs Obtained by Method A, Method C, an m-Sequence, and a Random Design for Estimating the HRF.

of the 1797 RE-values for each design, and the whiskers represent the highest and lowest values for the relative efficiencies. For instance, the RE-values for $\boldsymbol{d}^*_{Mm,MethodA}$ and $\boldsymbol{d}^*_{Mm,MethodC}$ are ranging from 0.95 to 1 indicating high relative efficiencies for all correlation coefficients. On the other hand, more variation in RE-values occurred

for m-sequence and random designs, with very low RE-values at the worst cases. To conclude, maximin designs are robust against a misspecified correlation coefficient while the traditional designs vary greatly over the parameter space of $AR(2)$ model.

3.4.2 Case II: Detection

In a similar manner as we did for the estimation case, we obtain maximin designs for detecting brain activations by using our proposed methods and compare them to the expensive approach by Maus et al. (2010).

The time spent to generate the locally optimal designs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ that maximize $F_d(\boldsymbol{d}, \boldsymbol{\phi}) = m/trace(\boldsymbol{M}^{-1}(\boldsymbol{d}, \boldsymbol{\phi}))$ where $\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi})$ as in (2.6) and its size $m = Q$; for $Q = 1, 2, 3$, is shown in Table 3.4 for Methods A, B and C. Table 3.5, presents a comparison among

Table 3.4: CPU Time Spent by the Genetic Algorithm on Maximizing F_d for $AR(2)$ Model With $Q = 1, 2, 3$.

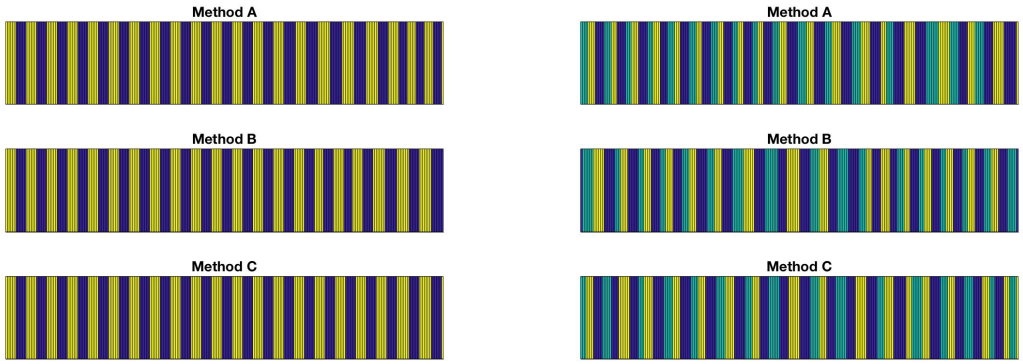
Method	number of LODs	Total time spent (min.): Q=1	Q=2	Q=3
A	20	0.3	1	2.3
B	1797	28.8	87	189
C	6	0.1	0.3	0.5

the obtained maximin designs $\boldsymbol{d}_{Mm,MethodA}$, $\boldsymbol{d}_{Mm,MethodB}$, and $\boldsymbol{d}_{Mm,MethodC}$.

We observe that the value of $minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$ for maximin designs obtained by Method A and Method C are quite similar to the grid method (Method B) with $88\% - 90\%$ maximin efficiencies. We also observe at least 99% reduction in time when Method A and Method C are used. As mentioned before the reported time in table 3.5 does not include the time needed for generating the locally optimal designs which is represented in Table 3.4. Figure 3.7 represents the maximin designs $\boldsymbol{d}^*_{Mm}$ achieved by the three methods, which are in the form of block designs with different block sizes (e.g., $\boldsymbol{d}^*_{Mm}$ designs obtained by Method A have blocks of sizes 6 and 7).

Table 3.5: The Performances of Method A, Method B, and Method C for Detecting Activated Brain Regions Over the Parameter Space $\boldsymbol{\Omega}$ and the Corresponding (ϕ_1, ϕ_2)-Values That Yield the Maximin Designs $\boldsymbol{d}^*_{Mm}$ for $AR(2)$ Model.

Q	1	2	3
Method A:			
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9030	0.8878	0.8910
time spent (min.)	0.3	0.9	0.8
(ϕ_1, ϕ_2)	(0.31,0.35)	(0,0)	(0.31,0.35)
Method B:			
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9067	0.8956	0.8921
time spent (min.)	26	79	111
(ϕ_1, ϕ_2)	(0.31,0.35)	(0,0)	(0,0)
Method C:			
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9090	0.8917	0.8786
time spent (min.)	0.1	0.34	0.3
(ϕ_1, ϕ_2)	(0.31,0.35)	(0.31,0.35)	(0.31,0.35)



(a) (Q,N)=(1,255) (b) (Q,N)=(2,242)

Figure 3.7: Maximin Designs Obtained by Method A, Method B, and Method C for Detection Case of $AR(2)$ Model.

Similarly to the estimation case, we now compare the achieved maximin designs by Method A and Method C with the block designs that are widely used for detecting brain activation; note that the block designs are shown to possess high detection

power when the uncertainty in the autocorrelation is ignored. For demonstration purposes, we generate two block designs with blocks of sizes 8 and 16, for each of the following scenarios $(Q, N) = (1, 255), (2, 242)$ and $(3, 255)$. The $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.; \boldsymbol{d}^*_{\boldsymbol{\phi}})$ for these deigns are reported in Table 3.6. We notice that the maximin designs obtained by our approaches provide high detection power when compared to the two block designs. In Figure 3.8 we compare the RE-values over $\boldsymbol{\Omega}$ of the maximin designs $\boldsymbol{d}^*_{Mm}$

Table 3.6: $minRE_{\phi\in\Omega}(.; \boldsymbol{d}^*_{\phi})$ of $\boldsymbol{d}^*_{Mm}$ From Methods A and C Versus Some Traditional Designs for Detection Case of $AR(2)$ Model.

	Q=1	Q=2	Q=3
$\boldsymbol{d}^*_{Mm,MethodA}$	0.9030	0.8878	0.8910
$\boldsymbol{d}^*_{Mm,MethodC}$	0.9090	0.8917	0.8786
block size8	0.7206	0.6052	0.5630
block size 16	0.4424	0.4013	0.3919

obtained by Method A and Method C, and the block designs. Each box represents the distribution of the 1797 $RE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.; \boldsymbol{d}^*_{\boldsymbol{\phi}})$ values for each design where it can be seen that block designs, which are highly recommended for detecting activated brain regions under linear models, the vary greatly over $\boldsymbol{\Omega}$ comparing to our maximin designs where the variation in RE-values is small. We also notice that block designs provide very low relative efficiencies at the worst cases, such as for block design of size 16 the $minRE$-values are as low as 40% for all scenarios.

From the above-mentioned results, we notice the importance of selecting a good sampling plan of $\boldsymbol{\phi} = (\phi_1, \phi_2)$ that helps to approximate the surface of RE-values over the parameter space $\boldsymbol{\Omega}$ of $AR(2)$ and obtain highly efficient maximin designs. Not all sampling plans give good results. To demonstrate this, we consider three rather extreme sampling plans: (i) the sampled points are aligned on 'diagonal' of $\boldsymbol{\Omega}$, (ii) the sampled points are clustered at the center of $\boldsymbol{\Omega}$, and (iii) the sampled points

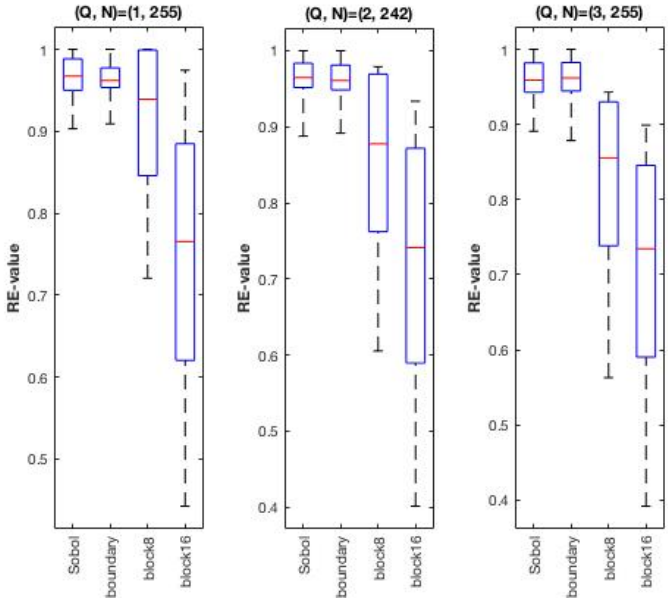
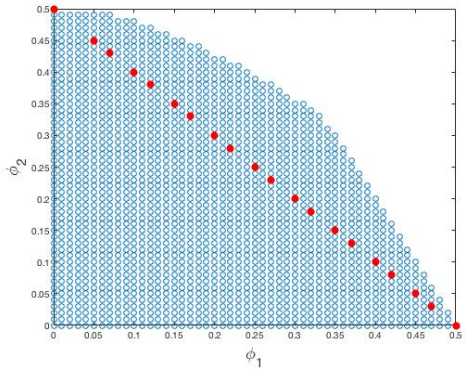


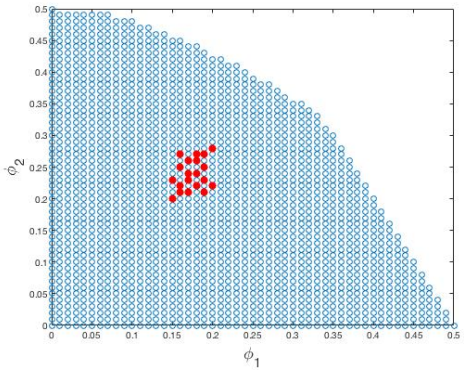
Figure 3.8: Boxplots of RE-Values Over $\boldsymbol{\Omega}$ of $AR(2)$ Model for Designs Obtained by Method A, Method C, and Block Designs of Sizes 8 and 16 for Detection Case of $AR(2)$ Model.

are clustered around the origin $(0, 0)$. These sampling plans are shown in Figure 3.9. Following Method A and Method C, we would like to obtain maximin designs $\boldsymbol{d}_{Mm}$ where we consider the extreme sampling plans and compare their performances to the results obtained by Sobol sampled points and the set of boundary point as shown in Table 3.7 where $\boldsymbol{d}^*_{Mm,cluster1}$ and $\boldsymbol{d}^*_{Mm,cluster2}$ denote the maximin designs when sampled points are clustered at the center of $\boldsymbol{\Omega}$ and around $(0, 0)$, respectively. We observe that maximin designs obtained by the extreme sampling plans for both Method A and Method C perform very poorly when compared to the results of Method A and Method C when Sobol sampled points and boundary points are considered, respectively as represented in Table 3.7.

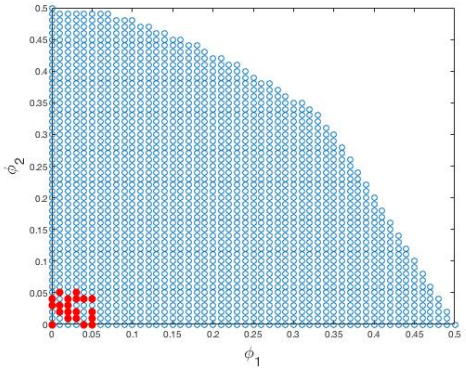
We note that, for Method A we facilitate the Kriging model to approximate the minRE-value when comparing the results obtained by Sobol sampled points to those obtained by the extreme sampling plans, whereas for Method C we no longer need to



(a) Diagonal Points.



(b) Clustered Points at the Center.



(c) Clustered Points Around $(0, 0)$.

Figure 3.9: Examples of Some Extreme Sampling Plans Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(2)$.

Table 3.7: $minRE_{\phi\in\Omega}(.; \boldsymbol{d}^*_\phi)$ of Maximin Designs From Method A and Method C for Some Extreme Sampling Plans Versus Sobol Points and Boundary Points for Detection Case.

	Method A			*Method C*		
Maximin design	Q=1	Q=2	Q=3	Q=1	Q=2	Q=3
$\boldsymbol{d}^*_{Mm}$	0.9030	0.8878	0.8910	0.9090	0.8917	0.8786
$\boldsymbol{d}^*_{Mm,diagonal}$	0.8190	0.8260	0.7888	0.8146	0.8268	0.8113
$\boldsymbol{d}^*_{Mm,cluster1}$	0.8515	0.8401	0.7913	0.8360	0.8375	0.8130
$\boldsymbol{d}^*_{Mm,cluster2}$	0.7212	0.6084	0.5773	0.7212	0.6095	0.5769

use Krigin approximation to compare the results obtained by the selected boundary points to those obtained by the bad samples. In conclusion, the results obtained by

the sampling plans that lack the space-filling property provide low efficiencies of the obtained maximin designs in comparison to our proposed methods.

3.4.3 Third Order Autoregressive Model for the Correlation Coefficients

The results from the previous subsections suggest our proposed efficient 'shortcut' methods are useful for obtaining maximin event-related fMRI designs. In particular, a small set of boundary points with only 6 points selected from the parameter space Ω of $AR(2)$, and the Kriging-based genetic algorithm method can efficiently generate satisfactory maximin designs with a significantly reduced computing time. These shortcut methods are expected to make the search more efficient and faster when the noise is assumed to follow a higher order autoregressive models, such as the $AR(3)$ model. For this case, we assume the same simulation settings as for the $AR(2)$ model except that the error correlation here follows $AR(3)$ model. We aim to find A-optimal maximin event-related designs for estimating the HRF and detecting the activated brain regions. For this case, the sampled points by Sobol sequences are as shown in Figure 3.3(b) with a total number of 30 points, and the selected 7 boundary points are as in Figure 3.4(b). We consider designs with $Q = 1$ and 2 stimulus types corresponding to design lengths of N = 255 and 242 events respectively, where the stimulus plus the control are included in the design sequence. We note that, for $Q = 3$, Method B (grid method), which requires a huge number of LODs, quickly becomes infeasible for $AR(3)$ model.

We discuss now the results for the first case study: the estimation of the HRF. We follow the same steps that was explained in Subsection 3.4.1. to apply Methods A, B, and C. Table 3.8 represents the time spent to generate the LODs maximizing $F_e(\boldsymbol{d}; \boldsymbol{\phi})$ for the selected points by Sobol sequences and boundary points from the parameter space $\boldsymbol{\Omega}$ of $AR(3)$, and the time needed for generating the LODs for the grid points.

When comparing the maximin designs obtained by Method A and Method C to those obtained by Method B, we evaluate the min-RE of the obtained maximin designs $\boldsymbol{d}^*_{Mm}$ designs by considering the 1957 locally optimal designs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ for all $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3) \in \boldsymbol{\Omega}$ in the finer grid. The results of the comparison between the three methods are provided in Table 3.9.

Table 3.8: CPU Time Spent by the Genetic Algorithm on Maximizing F_e for $Q = 1, 2$ Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(3)$ Model.

Method	number of LODs	Total time spent (min.): Q=1	Q=2
A	30	2.5	6
B	1957	167	432
C	7	0.5	1.5

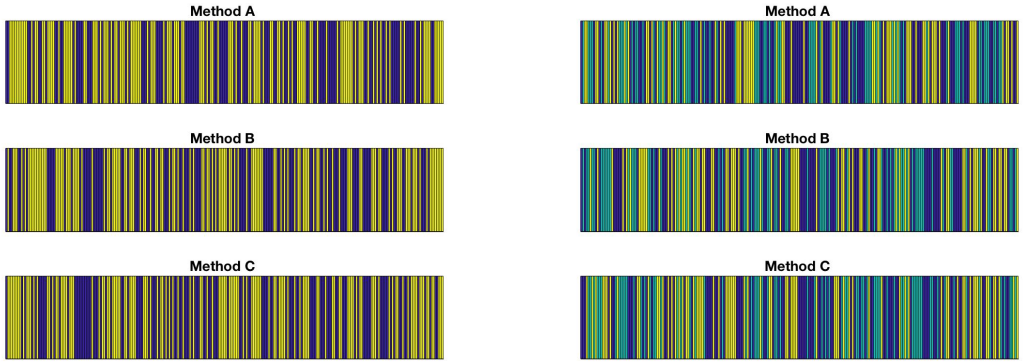
Table 3.9: The Performances of Method A, Method B, and Method C for Estimating the HRF Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(3)$ Model and the Corresponding (ϕ_1, ϕ_2, ϕ_3)-Values That Yield the Maximin Designs $\boldsymbol{d}^*_{Mm}$.

Q	1	2
Method A:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9309	0.9315
time spent (min.)	3	6
(ϕ_1, ϕ_2, ϕ_3)	(0,0,0.5)	(0,0,0.5)
Method B:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9400	0.9417
time spent (min.)	136	391
(ϕ_1, ϕ_2, ϕ_3)	(0,0.38,0.38)	(0,0,0.5)
Method C:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9389	0.9421
time spent (min.)	0.3	0.7
(ϕ_1, ϕ_2, ϕ_3)	(0.5,0,0)	(0,0,0.5)

We note that, the reported time in Table 3.9 is omitting the time needed for generating the locally optimal designs. We observe that the values of $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\boldsymbol{\phi}})$

for the three methods are quite similar and the obtained designs are very efficient (at least 93% maximin efficiency). We also observe a huge reduction in CPU time needed for generating maximin designs using boundary points method and Method A by 99% compared to the time spent by Method B. In conclusion, our proposed methods save much of the computational resources without sacrificing the efficiency of the obtained designs.

Figure 3.10 represents the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by the three methods for the case of estimating the HRF, when $(Q = 1, 2)$. Bars in dark blue indicate the null or control event and yellow bars represent the event type. In the same



(a) (Q,N)=(1,255) (b) (Q,N)=(2,242)

Figure 3.10: Maximin Designs Obtained by Methods A, B, and C for Estimating the HRF When $Q = 1, 2$ and the Error Correlation Follows $AR(3)$ Model.

manner as in subsection 3.4.1, we compare the performance of the maximin deigns obtained by the method of boundary points and Method A to some traditional designs of event-related fMRI studies that are known to have high estimation efficiency, such as m-sequence and random designs. As was mentioned before, we generate m-sequence designs following Liu and Frank (2004), and 10 different random designs. The $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.; \boldsymbol{d}^*_{\boldsymbol{\phi}})$ for these deigns are presented in Table 3.10.

Table 3.10: Comparison Between Maximin Designs $\boldsymbol{d}^*_{Mm}$ Obtained by Method A, Method C, and Some Traditional fMRI Designs for Estimation Purposes for $AR(3)$.

	Q=1	Q=2
$\boldsymbol{d}^*_{Mm,MethodA}$	0.9309	0.9315
$\boldsymbol{d}^*_{Mm,MethodC}$	0.9389	0.9421
m-sequence-based design	0.8837	0.8176
10 random designs	0.6664-0.8462	0.6702-0.7766

We notice that the maximin designs obtained by Method A (Kriging+Sobol) and the boundary points method outperform those traditional designs. The boxplots in Figure 3.11 display the distribution of the $RE_{\phi\in\boldsymbol{\Omega}}(.;\boldsymbol{d}^*_{\phi})$ values for the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by Method A, boundary points method, an m-sequence design and a random design. Each box in Figure 3.11 gives the median and 25% and 75%

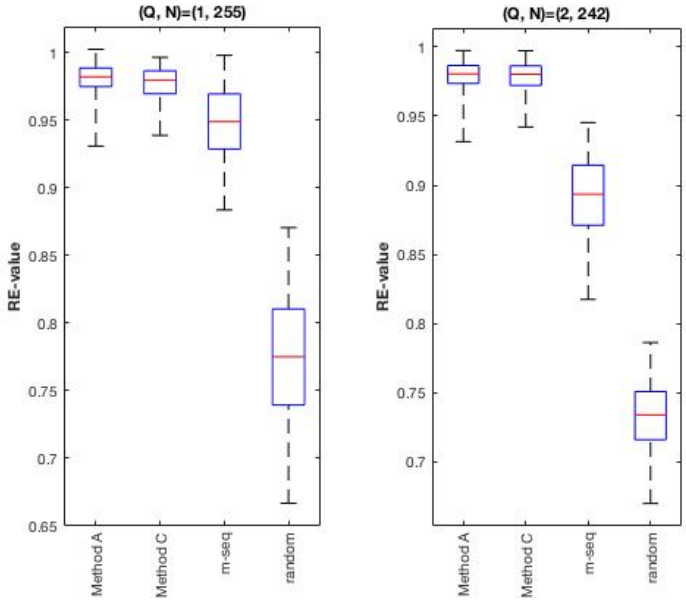


Figure 3.11: Boxplots of Relative Efficiencies for Designs Obtained by Sobol Points, Boundary Points, and Traditional Designs for Estimating the HRF When the Error Correlation Follows $AR(3)$ Model.

percentiles of the 1957 RE-values for each design, and the whiskers represent the highest and lowest values for the relative efficiencies. For instance, the RE-values

for $\boldsymbol{d}^*_{Mm,Sobol}$ and $\boldsymbol{d}^*_{Mm,boundary}$ are ranging from 0.93 to 1 indicating high relative efficiencies for all correlation coefficients. On the other hand, more variation in RE-values occurred for m-sequence and random designs. It can also be seen that the selected random design performs poorly when compared to other designs as it provides the lowest relative efficiency values.

Now we discuss the obtained results for detection case of $AR(3)$ model. The sample points consisting of correlation parameters at the boundaries of the parameter space $\boldsymbol{\Omega}$ and Method A provide efficient results compared to Method B as shown in Table 3.11. We notice that the minimum relative efficiencies are high for all maximin

Table 3.11: The Performances of Methods A and C Compared to Method B for Detection Case When the Error Correlation Follows $AR(3)$ Model.

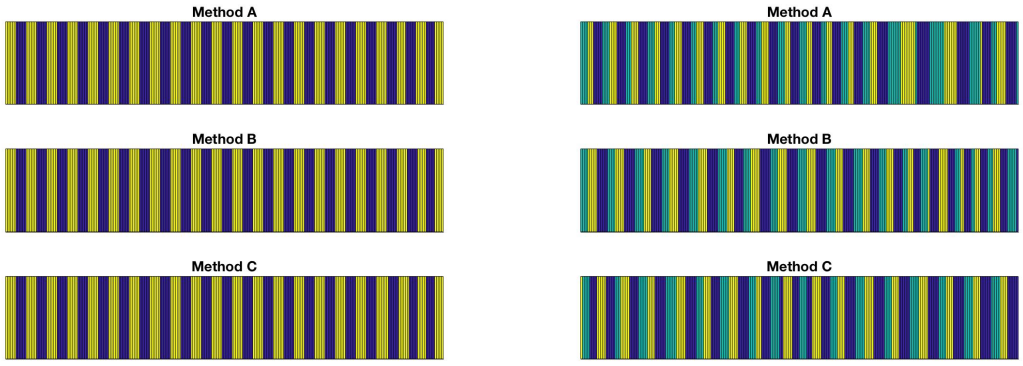
Q	1	2
Method A:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.9030	0.8757
time spent (min.)	0.3	1
(ϕ_1,ϕ_2,ϕ_3)	(0,0,0)	(0.09,0.34,0.31)
Method B:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.9030	0.8805
time spent (min.)	24	84
(ϕ_1,ϕ_2,ϕ_3)	(0,0,0)	(0.09,0.34,0.31)
Method C:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.9022	0.8649
time spent (min.)	0.1	0.2
(ϕ_1,ϕ_2,ϕ_3)	(0,0,0)	(0,0,0)

designs obtained by the three methods. The value of $minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$ for maximin design obtained by the two proposed methods are quite similar to the results of the grid method (Method B) with $86\% - 90\%$ maximin efficiencies. We also observe a 99% reduction in time using the shortcut methods. Again, the reported time in

table 3.11 does not include the time needed for generating the locally optimal designs which is represented in Table 3.12. The maximin designs $\boldsymbol{d}^*_{Mm}$ for the detection case which are in the form of block designs are presented in Figure 3.12.

Table 3.12: CPU Time Spent by the Genetic Algorithm on Maximizing F_d When $Q = 1, 2$ for $AR(3)$ Model.

Method	number of LODs	Total time spent (min.): Q=1	Q=2
A	30	0.5	1.5
B	1957	36	98
C	7	0.1	0.3



(a) (Q,N)=(1,255) (b) (Q,N)=(2,242)

Figure 3.12: Maximin Designs Obtained by Methods A, B, and C for Detection Case When $Q = 1, 2$ and the Error Correlation Follows $AR(3)$ Model.

To compare the performance of the achieved maximin designs to some known high detection power designs, we use the same block designs of sizes 8 and 16 that were used for detection comparison in Subsection 3.4.2. The $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.;\boldsymbol{d}^*_{\boldsymbol{\phi}})$ for these deigns are reported in Table 3.13. We notice that the maximin designs obtained by our proposed approaches outperform the efficiencies of the traditional fMRI designs. The boxplots of RE-values for maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by boundary points method and Method A versus traditional block designs are displayed in Figure 3.13,

Table 3.13: $minRE_{\phi\in\Omega}(.;\boldsymbol{d}^*_\phi)$ of $\boldsymbol{d}^*_{Mm}$ From Methods A and C Versus Some Traditional Designs for Detection Case of $AR(3)$ Model.

	Q=1	Q=2
$\boldsymbol{d}^*_{Mm,MethodA}$	0.9030	0.8757
$\boldsymbol{d}^*_{Mm,MethodC}$	0.9022	0.8649
block design of size8	0.6629	0.5223
block design of size 16	0.3261	0.2768

where each box represents the distribution of the 1957 $RE_{\boldsymbol{\phi\in\Omega}}(.;\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values for each of the above-mentioned designs when only one stimulus type is presented during the event-related fMRI experiment.

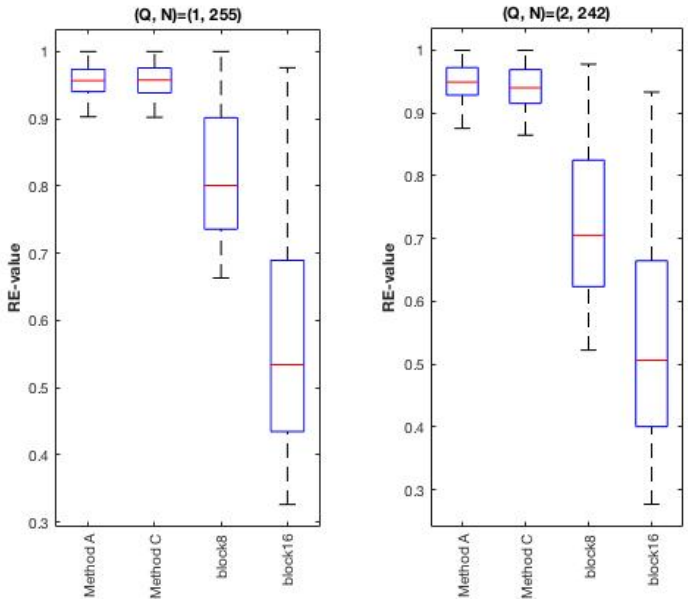


Figure 3.13: Boxplots of RE-Values Over $\boldsymbol{\Omega}$ of $AR(3)$ Model for Designs Obtained by Method A, Method C, and Block Designs of Sizes 8 and 16 for Detection Case.

We notice that the lowest RE-values for $\boldsymbol{d}^*_{Mm,MethodA}$ and $\boldsymbol{d}^*_{Mm,MethodC}$ is at 90% and 88% for $Q = 1$ and 2, respectively which indicates high relative efficiencies for all correlation coefficients. On the contrary, the RE-values for the block designs vary greatly over the parameter space as the size of the blocks increases.

3.5 Discussion

In this chapter, we proposed two efficient approaches to find maximin designs for event-related fMRI experiments with uncertain error correlation. The first approach is a combination of a Kriging approximation and a knowledge- based genetic algorithm where Sobol sequences are used to sample over the parameter space of the correlation coefficients. The second approach was inspired by the performance of the points that are located on the boundaries of the specified parameter space. Moreover, to give an idea of how could the sampling method affects the efficiencies of the obtained maximin designs, we provided some examples of bad sampling schemes where the correlation coefficients are clustered at the center or at the corner of the assumed parameter space then compared the results to the ones obtained by the sampling methods that have a space-filling property. For all simulations in this chapter, we considered experiments with simple trials where stimuli of one component (e.g., a picture) are presented to the experimental subject and assume the noise to follow higher orders of autoregressive models such as $AR(2)$ and $AR(3)$ models. The performances of our proposed approaches were compared to the grid method by Maus et al. (2010) for the objectives of estimating the HRF and detecting active brain regions. From case studies, we concluded that maximin designs obtained by our methods are as efficient as those obtained by the grid method, but the computing time required by our methods is very short.

Chapter 4

COMPOUND TRIALS EVENT-RELATED FMRI EXPERIMENTS

4.1 Introduction

For neuroimaging studies, the use of compound stimuli (trials) that consist of two or more mental tasks is not uncommon in practice (e.g., Müller, Bartelt, Donner, Villringer, & Brandt, 2003; Müller & Kleinschmidt, 2003, 2004; Pochon et al., 2001; Ress, Backus, & Heeger, 2000; Serences, 2004; Silver, Ress, & Heeger, 2007). In this chapter, we extend our work to the case where each stimulus consists of more than one component. For simplicity, we assume that each stimulus involves two components although this assumption is not essential, and our method can be easily extended to cases where a stimulus is formed by three or more components or different numbers of tasks are considered for different stimulus types. As an example, each stimulus may consist of a brief cue and a mental task for the subject to complete after some specific time interval following the cue. In this case, an event-related fMRI design can be written in the same way as designs for simple stimuli such as a design with two stimulus types $\boldsymbol{d} = \{11022....0\}$, but 1 here represents the event that a cue is followed by a stimulus of the first type, 2 denotes the second type of stimulus following a cue, and 0 means no stimulus presentation at the corresponding time point. We note that, the time between onsets of consecutive events (i.e. elements in $\boldsymbol{d}$) is assumed to be a pre-specified constant denoted by τ_{ICI}; τ_{ICI} is thus the counterpart of τ_{ISI}. The time between consecutive fMRI scans is denoted by τ_{TR} as mentioned in Subsection 2.1.1. The time interval between components of each trial (i.e., the time from the cue to the mental task) is denoted by τ_{CTSI}. For this time interval, we consider two cases. For

Case (1), we assume that the time between the different components of a compound trial remains fixed for all trials. In this case, the primary interest lies in finding designs suitable for detecting active brain voxels only. With this assumption, the objective of estimating the HRF evoked by the individual component is not possible because we normally have nonidentifiable HRF parameters between the two components. Thus the HRF is not estimable. For Case (2), we allow some variation in τ_{CTSI} across the events. For this case, we aim to obtain efficient maximin designs for both detection and estimation purposes.

As an example for Case (1), we consider the above-mentioned design $\boldsymbol{d}$, and assume $\tau_{ICI} = 8$ seconds and $\tau_{CTSI} = 4$ seconds is fixed throughout the experiments. The first component of type-1 trial will be presented at 0 s, 8 s, and so on. The second component of type-1 trial will be presented at 4 seconds, 12 seconds, and so on. Furthermore, the first component of type-2 trial will be presented at 24 seconds, 32 seconds and so on, while the second component of type-2 trial will be presented at 28 and 36 seconds. As another example for Case (2) when τ_{CTSI} varies, we assume $\tau_{CTSI} = \{43022...0\}$ which is a sequence that has the same length as the design $\boldsymbol{d}$ to indicate the τ_{CTSI} time that corresponds to each stimulus in $\boldsymbol{d}$. We also consider the same $\tau_{ICI}(= 8$ seconds) as for the previous example. To explain this, the experimental subject will receive the second component of type-1 trial at 4 and 11 seconds following the presentation of the cue (at 0s and 8s, respectively). We simply set τ_{CTSI} to 0 when $d_n = 0$. The second component of type-2 trial will be shown to the subject at 26 and 34 seconds. In a similar manner as the simple trials case, we assume that at an activated brain voxel, each component evokes a change in the event-elated fMRI signal, which is described by the hemodynamic response function HRF. Additionally, components of the same type evoke the same HRF throughout the experiment, and the heights of overlapping HRFs sum linearly. To satisfy the assumption of additive

HRFs, specifically for varying τ_{CTSI} case, the possible values that τ_{CTSI} could take are restricted to the following condition: $2 \leq \tau_{CTSI} \leq \tau_{ICI} - 2$; otherwise, the assumption will likely to be violated.

For experiments involving compound trials, the interest might be on examining how the subject's brain reacts to the mental tasks rather than the joint response to both the cue and the task. We put our focus on this type of studies, but note that, for cases where the interest is in the joint response, one may consider to treat each compound stimulus as a simple one, and adapt/adopt the design for simple stimulus cases. Our main goal is to find high-quality experimental designs by taking into account the uncertainty of the correlation coefficients at the design stage. However, this is a very challenging task because of the increased number of components in each trial type which leads to an expansion in the design space of all possible event-related fMRI. Clearly, an exhaustive search over this space is impossible. Therefore, an efficient approach is needed for obtaining designs with high efficiencies to allow valid and precise statistical inferences.

4.2 Statistical Models for Event-Related fMRI with Compound Trials

In Chapter 2, we discussed the commonly used general linear models for estimation and detection objectives as in (2.1) and (2.2), respectively, for experiments involving simple trials. We generalize those models to accommodate the compound trials case. In particular, the general linear model for estimating the HRF can be written as:

$$\boldsymbol{y} = \sum_{q=1}^{Q} (\boldsymbol{X}_{q,1}\boldsymbol{h}_{q,1} + \boldsymbol{X}_{q,2}\boldsymbol{h}_{q,2}) + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{4.1}$$

where $\boldsymbol{y} = ((y_t))_{t=1,...,T}$ is a $T \times 1$ vector representing the BOLD time series from a brain voxel, Q represents the total number of trial types, and $\boldsymbol{h}_{q,1} = (h_{q,1,1}, ..., h_{q,1,k_1})'$ is the HRF parameter vector evoked by the first component of the qth trial type.

Here, each $h_{q,1,j}, j = 1, ..., k_1$, corresponds to the jth height of the HRF for the first component of the qth trial type at time $((j-1)\tau_{\Delta T})$ and k_1 is defined as $1+\lfloor H/\tau_{\Delta T} \rfloor$; H represents the duration of the HRF evoked by the first component (32 seconds). $\boldsymbol{h}_{q,2} = (h_{q,2,1}, ..., h_{q,2,k_2})'$ is the HRF parameter vector evoked by the second component of the qth trial type, where each $h_{q,2,j}, j = 1, ..., k_2$, corresponds to the jth height of the HRF for the second component of the qth trial type at time $((j-1)\tau_{\Delta T} + \xi)$ where $\xi = \tau_{CTSI} - c\tau_{\Delta T}$ and $0 \leq \xi < \tau_{\Delta T}$, and $c = \lfloor \tau_{CTSI}/\tau_{\Delta T} \rfloor$. k_2 is defined as $1 + \lfloor (H - \xi)/\tau_{\Delta T} \rfloor$. We assume that both HRFs have a duration of 32 seconds. But, the lengths of the HRF parameters for these two components might not be the same as the onset times of the second component might not be synchronized with the scanning times (i.e. when $\xi \neq 0$). $\tau_{\Delta T}$ is the discretization interval, which is the greatest time making both $m\tau_{ICI} = \tau_{ICI}/\tau_{\Delta T}$ and $m\tau_{TR} = \tau_{TR}/\tau_{\Delta T}$ integers. We note that $\tau_{\Delta T}$ may or may not be an integer. $\boldsymbol{X}_{q,i}$ is the $T \times k_i$, where $i = 1, 2$, design matrix for the ith component of the qth trial type, where the $(t, j)th$ element of $\boldsymbol{X}_{q,i}$ is 1 when $h_{q,i,j}$ contributes to y_t, and 0 otherwise. The remaining terms are as in (2.1). For demonstration purposes, the noise is assumed to follow a stationary $AR(2)$ process with unknown correlation coefficients parameters.

The design matrix $\boldsymbol{X}_{q,1}$ is constructed in a similar way as for the simple trials case as described in Subsection 2.1.2 by using τ_{ICI} to replace τ_{ISI} and each column of $\boldsymbol{X}_{q,1}$ corresponds to a parameter in $\boldsymbol{h}_{q,1}$. The design matrix for the second component $\boldsymbol{X}_{q,2}$ can be constructed in a similar fashion as the construction of $\boldsymbol{X}_{q,1}$; but, now the value of ξ needs to be taken into account. Each column of $\boldsymbol{X}_{q,2}$ corresponds to a parameter in $\boldsymbol{h}_{q,2}$. For Case (1), $\boldsymbol{X}_{q,2}$ is sometimes a shift of $\boldsymbol{X}_{q,1}$ as seen in the following example.

Example 4.1: Let the design $\boldsymbol{d} = \{11010....0\}, Q = 1$, and assume $\tau_{ICI} = 4$ seconds, $\tau_{TR} = 2$ seconds and $\tau_{\Delta T} = 2$ seconds. The time between components of each trial

is set to $\tau_{CTSI} = 2$ seconds which is fixed throughout the experiment. We can write $\boldsymbol{X}_{1,1}$ and $\boldsymbol{X}_{1,2}$ as:

$$\boldsymbol{X}_{1,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 1 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}; \boldsymbol{X}_{1,2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 1 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}.$$

Here, $\boldsymbol{X}_{1,2}$ is a shifted version of $\boldsymbol{X}_{1,1}$. We note that the first row of $\boldsymbol{X}_{1,2}$ is a row of zeros as the second component does not contribute to the fMRI signal acquired at the first scan.

The general form of writing the design matrix $\boldsymbol{X}$ for the compound trials case is: $\boldsymbol{X} = [\boldsymbol{X}_{1,1}, \boldsymbol{X}_{2,1}, ..., \boldsymbol{X}_{Q,1}, \boldsymbol{X}_{1,2}, \boldsymbol{X}_{2,2}, ..., \boldsymbol{X}_{Q,2}]$. Considering the design $\boldsymbol{d}$ in **Example 4.1**, we only have one trial type, so the design matrix $\boldsymbol{X}$ can be written as the concatenation of the two matrices $\boldsymbol{X}_{1,1}$ and $\boldsymbol{X}_{1,2}$. For the case of more than one trial type, the construction of the design matrix should also be straightforward. For Case (2), the construction of the design matrix $\boldsymbol{X}_{q,2}$ can alternatively be done with the steps explained below. Specifically, we further set $\xi_r = \tau_{CTSI_r} - c_r\tau_{\Delta T}$, where $c_r = \lfloor \tau_{CTSI_r}/\tau_{\Delta T} \rfloor$ and $0 \leq \xi_r < \tau_{\Delta T}$, and $r = 1, 2, ...N$; N represents the number of different τ_{CTSI} under the condition $2 \leq \tau_{CTSI_r} \leq \tau_{ICI} - 2$ to allow ad-

ditivity of overlapping HRFs. To construct $\boldsymbol{X}_{q,2}$, we may first construct the $T \times j$, 0-1 design matrices $\boldsymbol{X}_{q,2,1}, \boldsymbol{X}_{q,2,2}..., \boldsymbol{X}_{q,2,N}$ for the N τ_{CTSI_r}'s, respectively. Each column of $\boldsymbol{X}_{q,2,r}$ still corresponds to a parameter in $\boldsymbol{h}_{q,2}$ and each row of these design matrices corresponds to the scanning time τ_{TR}; its element is 1 if, under the given τ_{CTIS_r}, the corresponding HRF height contributes to the fMRI signal. $\boldsymbol{X}_{q,2}$ is then $\boldsymbol{X}_{q,2,1} + \cdots + \boldsymbol{X}_{q,2,N}$.

Example 4.2: For simplicity, we consider a design with one trial type $\boldsymbol{d} = \{1010...0\}$ and assume $\tau_{ICI} = 6$ seconds, $\tau_{TR} = 2$ seconds and $\tau_{\Delta T} = 2$ seconds. The varying time interval between the two components τ_{CTSI} can be written as a sequence that has the same length as $\boldsymbol{d}$, and this sequence is assumed to be $\{4020...0\}$. Here, we also say that τ_{CTSI} can either be 2 seconds or 4 seconds. In the current setting, $\xi_r = 0$ for all r, and $\boldsymbol{h}_{1,2}$ contains the HRF heights of the second components evaluated every 2 seconds from the onset of the second component to the complete return of the HRF. In this example, the design matrix for the cues is:

$$\boldsymbol{X}_{1,1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}.$$

The individual design matrices for the second components $\boldsymbol{X}_{1,2,1}$ and $\boldsymbol{X}_{1,2,2}$ are:

$$\boldsymbol{X}_{1,2,1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}; \boldsymbol{X}_{1,2,2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}.$$

Specifically, $X_{1,2,1}$ is for $\tau_{CTSI} = 2$ seconds, and $X_{1,2,2}$ is for $\tau_{CTSI} = 4$ seconds. Then the design matrix $\boldsymbol{X}_{1,2} = \boldsymbol{X}_{1,2,1} + \boldsymbol{X}_{1,2,2}$:

$$\boldsymbol{X}_{1,2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 1 & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}.$$

As mentioned earlier, by concatenating the two matrices $\boldsymbol{X}_{1,1}$ and $\boldsymbol{X}_{1,2}$ we obtain the final form for the design matrix $\boldsymbol{X}$.

The general linear model for the objective of detecting the activated regions in the brain for experiments with compound trials is in the following form:

$$\boldsymbol{y} = \sum_{q=1}^{Q} (\boldsymbol{X}_{q,1}\boldsymbol{h}_1^*\theta_{q,1} + \boldsymbol{X}_{q,2}\boldsymbol{h}_2^*\theta_{q,2}) + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{4.2}$$

where $\boldsymbol{X}_{q,i}, q = 1, .., Q$ and $i = 1, 2$ is the design matrix of 0-1 with 1 corresponding to the heights of the HRF ($\boldsymbol{h}_i^*\theta_{q,i}$) that contribute to the response y_t. $\boldsymbol{h}_i^*$ is defined as $\boldsymbol{h}_i^* = (h_{i,1}^*, ..., h_{i,k_i}^*)'$ is a $k_i \times 1$ vector representing the assumed shape of the HRF for the ith-component, which is commonly set to the double gamma function of the SPM software package as defined in (2.4), with $h_{1,j}^* = g((j-1)\tau_{\Delta T})/max_s g(s), 1 \leq j \leq \lfloor H/\tau_{\Delta T} \rfloor$, and $h_{2,j}^* = g((j-1)\tau_{\Delta T} + \xi)/max_s g(s), 1 \leq j \leq \lfloor (H-\xi)/\tau_{\Delta T} \rfloor$, where $0 \leq \xi < \tau_{\Delta T}$. $\theta_{q,i}$ represents the unknown response amplitude of the HRF for the ith component of the qth trial type. The remaining terms in (4.2) are defined as in (2.2).

For simplicity, we only explain the model formulation for Case (1), but this can easily be extended to Case (2). Both models could also be expanded easily for the case when more than two components are considered. We also note that models (4.1) and (4.2) are for cases where different cues are used for different types of stimuli. When the same cue is used, we may assume that the same HRF is evoked by the cue, and the corresponding design matrix for it is the sum of the $X_{q,1}$'s.

4.2.1 Design Selection Criteria for Compound Trials

For simple trials studies in Chapter 2, we discussed the importance of finding designs that yield the most precise parameter estimates by some functions of the variance-covariance matrix (or the information matrix). For the compound trials studies, to estimate the HRF parameter vector $\boldsymbol{h} = (\boldsymbol{h}'_{1,1}, ..., \boldsymbol{h}'_{Q,1}, \boldsymbol{h}'_{1,2}, ..., \boldsymbol{h}'_{Q,2})'$, we aim to find a design $\boldsymbol{d}$ that maximizes the following information matrix:

$$\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi}) = \boldsymbol{X}'[\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}]\boldsymbol{X}. \tag{4.3}$$

Here, the design matrix $\boldsymbol{X}$ is defined as a concatenation of the matrices of the cues and the matrices of the mental stimuli following the cues as described in the previous Section. The remaining terms are defined as in (2.5).

For the objective of detecting the activated brain voxels, we extend (2.6) to accommodate the compound case by setting $\boldsymbol{\theta} = (\theta_{1,1}, ..., \theta_{Q,1}, \theta_{1,2}, ..., \theta_{Q,2})'$ to represent the vector of unknown HRFs amplitudes. The aim is to find a design $\boldsymbol{d}$ that maximizes the following information matrix to precisely estimate $\boldsymbol{\theta}$:

$$\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi}) = \boldsymbol{H}'\boldsymbol{X}'[\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}\boldsymbol{S})^{-1}\boldsymbol{S}'\boldsymbol{\Sigma}(\boldsymbol{\phi})^{-1}]\boldsymbol{X}\boldsymbol{H}, \tag{4.4}$$

where $\boldsymbol{H} = (\boldsymbol{I}_Q \otimes \boldsymbol{h}_1^*) \oplus (\boldsymbol{I}_Q \otimes \boldsymbol{h}_2^*) = diag((\boldsymbol{I}_Q \otimes \boldsymbol{h}_1^*), (\boldsymbol{I}_Q \otimes \boldsymbol{h}_2^*))$. The remaining terms are as in (2.6).

In order to evaluate the efficiencies of competing designs for the estimation purposes for compound trials experiments, both A- and D- optimality criteria could be applied and defined as:

$$F_e(\boldsymbol{d};\boldsymbol{\phi}) = \begin{cases} m_h/trace(\boldsymbol{C}_h\boldsymbol{M}^{-1}(\boldsymbol{d};\boldsymbol{\phi})\boldsymbol{C}_h'), & forA-optimality \\ det(\boldsymbol{C}_h\boldsymbol{M}(\boldsymbol{d};\boldsymbol{\phi})\boldsymbol{C}_h')^{1/m_h}, & forD-optimality \end{cases}$$

where m_h is the number of rows of $\boldsymbol{C}_h$ which is the matrix of linear combinations of the HRF heights and can be defined as $\boldsymbol{C}_h = [\boldsymbol{O}_{Qk}, \boldsymbol{I}_{Qk}]$ for the case when the interest is on examining the effects of the second component of each trial type, $\boldsymbol{O}_{Qk}$ is the Qk by Qk matrix of zeros, $\boldsymbol{I}_{Qk}$ is the Qk by Qk identity matrix, Q represents the total number of compound trial types, and k is the total number of HRF heights.

For detection power, the design selection criteria are defined as:

$$F_d(\boldsymbol{d};\boldsymbol{\phi}) = \begin{cases} m_\theta/trace(\boldsymbol{C}_\theta\boldsymbol{M}^{-1}(\boldsymbol{d};\boldsymbol{\phi})\boldsymbol{C}_\theta'), & forA-optimality \\ det(\boldsymbol{C}_\theta\boldsymbol{M}(\boldsymbol{d};\boldsymbol{\phi})\boldsymbol{C}_\theta')^{1/m_\theta}, & forD-optimality \end{cases}$$

where m_θ is the number of rows of $\boldsymbol{C}_\theta$ which is the matrix of linear combinations of the response amplitudes of the second component of the compound trial and can be defined as $\boldsymbol{C}_\theta = [\boldsymbol{O}_Q, \boldsymbol{I}_Q]$ where $\boldsymbol{O}_Q$ and $\boldsymbol{I}_Q$ are zeros and identity matrices of dimension Q. Both $F_e(\boldsymbol{d};\boldsymbol{\phi})$ and $F_d(\boldsymbol{d};\boldsymbol{\phi})$ are considered as the higher the better criteria.

4.3 Case Studies for Compound Trials Event-Related fMRI Experiments

For all simulations, we assume that a compound trial consists of a brief cue followed by a mental task after some specific time interval τ_{CTSI}. The interest is on examining how the experimental subject reacts to the second component of the compound trial. For evaluating the goodness of the competing designs for both estimation

and detection, we consider the A-optimality criterion as defined in the previous Subsection and focusing on the case where the individual stimulus effects is of interest. The duration of the HRF is set to 32 seconds, the nuisance term $\boldsymbol{S\gamma}$ in models (4.1) and (4.2), is assumed to allow for a second-order polynomial drift in the fMRI time series. The noise is assumed to follow a stationary $AR(2)$ process with unknown correlation coefficients parameters. We consider the same settings for the algorithmic parameters of the genetic algorithm as in Section 3.4 for simple trials experiments. To search for optimal event-related fMRI designs with compound trials, we modify the MATLAB program provided by Kao (2009), and combine it with the software package DACE for Kriging approximations.

For all case studies, we consider designs with $Q = 1$ and 2 trial types with their respective design lengths of $N = 255$, and 242. The time interval between cues τ_{ICI} is set to 8 seconds and the time to repetition τ_{TR} is 2 seconds, thus, $\tau_{\Delta T}$ is also 2 seconds (the greatest time making both $\tau_{ICI}/\tau_{\Delta T}$ and $\tau_{TR}/\tau_{\Delta T}$ integers). We assume different cues for each trial type. For Case (1), the time interval between components of the same trial is fixed, and is set to be $\tau_{CTSI} = 4$ seconds. For Case (2) when a varying time interval is allowed between components of the same trial, each individual τ_{CTSI_ν} is allowed to be $2, 4$, or 6 seconds. All simulations are implemented on the same desktop computer as described in Section 3.4. We study the performance of our proposed approach of combining the maximin criterion and the Kriging approximation model considering a sampled correlation parameters by Sobol sequences (Method A) and the boundary points approach (Method C) as described in Chapter 3. For comparison purposes, we also apply the expensive method of Maus et al. (2010) (Method B) to find maximin designs for compound trials experiments. We note that, for Method A, we use the same set of 20 $\boldsymbol{\phi} = (\phi_1, \phi_2)$ points sampled from the parameter space $\boldsymbol{\Omega}$ of $AR(2)$ model and the same set of 1797 $\boldsymbol{\phi}$-values on the fine

grid over $\boldsymbol{\Omega}$ for Method B as shown in Figure 3.3(a). For Method C, we consider the same set of boundary points sampled from the parameter space $\boldsymbol{\Omega}$ of $AR(2)$ as shown in Figure 3.4 (a). We aim to find optimal or near optimal designs for the objective of detecting activated brain regions for Case (1), and for detection and estimation purposes for Case (2).

4.3.1 Case (1): Fixed Time Between Components of the Compound Trials.

Our aim is to find maximin event-related fMRI designs for detecting brain activations for compound trials experiment when the time interval between components of the compound trials is assumed to be fixed. As mentioned earlier, the HRF parameters are not estimable in Case (1). Following the same steps in Subsection 3.4.1, we compare the performance of Method A, Method B and Method C in obtaining maximin designs with some modifications to accommodate the case of the compound trials. Specifically, when generating the LODs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ for the selected correlation parameters, which then can be used to calculate the $\underset{\boldsymbol{\phi}\in\boldsymbol{\Omega}}{min} RE(\boldsymbol{d};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values as defined in (3.1), we consider the A-optimality criteria for detection purposes $F_d(\boldsymbol{d};\boldsymbol{\phi})$ as defined in Subsection 4.2.1.

Table 4.1 represents the time spent in hours to generate the LODs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ for Methods A, B and C, that maximize $F_d(\boldsymbol{d},\boldsymbol{\phi}) = m_\theta/trace(\boldsymbol{C}_\theta \boldsymbol{M}^{-1}(\boldsymbol{d},\boldsymbol{\phi})\boldsymbol{C}'_\theta)$ where $\boldsymbol{M}(\boldsymbol{d};\boldsymbol{\phi})$ as in (4.4).

Table 4.1: CPU Time Spent by the Genetic Algorithm on Maximizing $F_d(\boldsymbol{d},\boldsymbol{\phi})$ of $AR(2)$ Model With $Q = 1$and 2 for Experiments With Fixed Time Between Components of Compound Trials.

Method	number of LODs	Total time spent (hrs.): Q=1	Q=2
A	20	0.03	0.1
B	1797	2.5	9
C	6	0.01	0.03

The efficiencies of the maximin designs obtained by our proposed methods compared to Method B (Grid method) for Case (1) are reported in Table 4.2 with the correlation parameters values where the $minRE$-value has occurred. For experiments

Table 4.2: The Performances of Method A, Method B, and Method C for Detecting Activated Brain Regions Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(2)$ Model and the Corresponding (ϕ_1, ϕ_2)-Values That Yield the Maximin Designs $\boldsymbol{d}^*_{Mm}$ for Experiments With Fixed Time Between Components of Compound Trials.

	Q= 1	Q= 2
Method A:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9667	0.9900
time spent (hrs.)	0.1	0.1
(ϕ_1, ϕ_2)	(0.23,0.41)	(0.18,0.43)
Method B:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9681	0.9929
time spent (hrs.)	2	6
(ϕ_1, ϕ_2)	(0.11,0.01)	(0.18,0.43)
Method C:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9647	0.9912
time spent (hrs.)	0.02	0.01
(ϕ_1, ϕ_2)	(0.23,0.41)	(0.04,0.01)

with one trial type the efficiencies of the obtained designs by all three methods are at least 96% and 99% for both scenarios. The proposed approaches have done a good job in finding maximin designs that are as efficient as the one obtained by the expensive method (Method B) with a huge reduction in the CPU time by at least 99.9%. The reported time in table 4.2 does not include the time needed for generating the locally optimal designs which is in Table 4.1.

In addition, the performances of designs obtained by Methods A and C are compared to a set of different traditional designs. As a result, our proposed approaches outperformed the traditional ones for all cases. To demonstrate that, we consider

experiments with one and two trial types and generate two block designs with blocks of size 4 for both scenarios. We also include m-sequence based designs for each of the considered scenarios. We generate three mixed designs which are combinations of the previously generated block and m-sequence designs. The first mixed design has 25% block design and 75% m-sequence, the second mixed design has 50% block design and 50% m-sequence, and the third mixed design has 75% block design and 25% m-sequence. A total of 20 randomly generated designs for both scenarios were also included to the set. The $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.;\boldsymbol{d}^*_{\boldsymbol{\phi}})$ for these deigns are reported in Table 4.3.

Table 4.3: $minRE_{\phi\in\Omega}(.;\boldsymbol{d}^*_{\phi})$ of $\boldsymbol{d}^*_{Mm}$ From Methods A and C Versus a Block Design, an m-Sequence Design, 3 Mixed Design and 10 Random Designs for Detection Purposes of $AR(2)$ Model for Experiments With Fixed Time Between Components of Compound Trials.

	Q=1	Q=2
$\boldsymbol{d}^*_{Mm,MethodA}$	0.9648	0.9906
$\boldsymbol{d}^*_{Mm,MethodC}$	0.9667	0.9882
block design	0.4971	0.6862
m-sequence design	0.6839	0.7856
mixed 1 design	0.6491	0.7640
mixed 2 design	0.5922	0.7398
mixed 3 design	0.5383	0.7138
10 random designs	0.5054-0.6854	0.7572-0.8258

We observe that block designs are the worst for detection purposes whereas m-sequence and random designs outperform the block design for compound trial experiments, which is in contrast to simple trial experiments where block designs are highly recommended for detection and both m-sequence and random designs known for poor detection power. For mixed designs, we notice that the efficiencies improve as the fraction of the design is made from an m-sequence designs increases (or as the frac-

tion of the design made from a block design decreases). Overall the maximin designs obtained by our proposed approaches provide high detection power when compared to the other designs.

Figure 4.1 represents a comparison between the distribution of the RE-values over $\boldsymbol{\Omega}$ of $AR(2)$ model for maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by Method A and Method C versus the other traditional designs. Each box represents the distribution of the 1797 $RE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.;\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values for each of the competing designs. We observe there are more variation in RE- values occurred for traditional designs when compared to the distribution of the RE-values for Methods A and C. The performance of the block designs was the worst among all designs with the largest variation in RE-values over the parameter space of $AR(2)$ model. In conclusion, maximin designs obtained by Methods A and C show high relative efficiencies for all correlation coefficients.

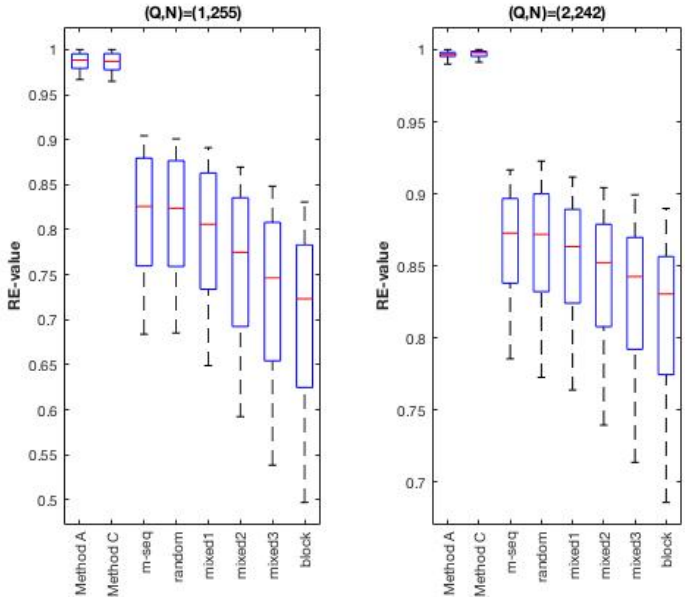


Figure 4.1: Boxplots of RE-Values Over $\boldsymbol{\Omega}$ of $AR(2)$ Model for Designs Obtained by Method A, Method C, an m-Sequence, a Random Design, 3 Mixed Designs, and a Block Design for Detection Purposes for Experiments With the Fixed Time Between Components of Compound Trials.

4.3.2 *Case (2): Varying Time Between Components of the Compound Trials.*

As mentioned earlier for this particular case, both estimation and detection are allowed here. We start by discussing the results of the estimation purposes. We aim to precisely estimate the HRF parameters considering model (4.1). Since the focus is on the effect of the second component rather than the joint response from both the cue and the task, the number of the HRF parameters to be estimated for the second component when $Q = 1$ is 17 and that with $Q = 2$ is 34, assuming that the HRF duration is 32 seconds. We follow the same steps in Subsection 3.4.1 to study the performances of Methods A, B and C in obtaining the desired maximin designs. Here, we consider the A-optimality criteria for estimation purposes $F_e(\boldsymbol{d}; \boldsymbol{\phi})$ as defined in Subsection 4.2.1 to generate the LODs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ for the selected correlation parameters, which then can be used to calculate the minRE-value for each candidate design. Note that, for Case (2), in additional to the presentation times of the compound stimuli, the search algorithm also selects the τ_{CTSI} for each stimulus to yield an optimal design. Table 4.4 represents the time spent to generate the set of the LODs maximizing $F_e(\boldsymbol{d}; \boldsymbol{\phi}) = m_h / trace(\boldsymbol{C}_h \boldsymbol{M}^{-1}(\boldsymbol{d}; \boldsymbol{\phi}) \boldsymbol{C}_h')$ where $\boldsymbol{M}(\boldsymbol{d}; \boldsymbol{\phi})$ as in (4.3).

Table 4.4: CPU Time Spent by the Genetic Algorithm on Maximizing $F_e(\boldsymbol{d}; \boldsymbol{\phi})$ for $Q = 1$ and 2 Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(2)$ Model for Experiments With Varying Time Between Components of Compound Trials.

Method	number of LODs	Total time spent (hrs.):	
		Q=1	Q=2
A	20	0.3	0.6
B	1797	23	63
C	6	0.1	0.2

When comparing the maximin designs obtained by Method A and Method C to those obtained by Method B, we evaluate the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by

Method A, and Method C using the 1797 LODs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ on the fine grid. Table 4.5 provides the $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values of the designs obtained by the three methods for $Q = 1$ and 2.

Table 4.5: The Performances of Method A, Method B, and Method C for Estimating the HRF Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(2)$ Model and the Corresponding (ϕ_1, ϕ_2)-Values That Yield the Maximin Designs $\boldsymbol{d}^*_{Mm}$ for Experiments With Varying Time Between Components of Compound Trials.

Q	1	2
Method A:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.8935	0.9210
time spent (hrs.)	0.2	0.3
(ϕ_1, ϕ_2)	(0.01,0.07)	(0.01,0.04)
Method B:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.8856	0.9280
time spent (hrs.)	19	48
(ϕ_1, ϕ_2)	(0.19,0.42)	(0.01,0.04)
Method C:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\phi})$	0.8866	0.9276
time spent (hrs.)	0.1	0.1
(ϕ_1, ϕ_2)	(0.19,0.24)	(0.33,0.25)

We note that, the reported time in Table 4.5 does not include the time needed for generating the locally optimal designs $\boldsymbol{d}^*_{\phi}$, which can be found in Table 4.4. We observe that the values of $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(\boldsymbol{d}^*_{Mm};\boldsymbol{d}^*_{\boldsymbol{\phi}})$ for the three methods are quite similar with a huge reduction in the time needed for generating a maximin design when Method A and Method C are used. In particular, for all scenarios ($Q = 1, 2$) our proposed methods reduce the CPU time by at least 99.9%. Figure 4.2 represents the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by the three methods for estimating the HRF when $Q = 1$. Here, the control event is represented by the dark blue color while each of the remaining 3 colors represents the first trial type with the time interval τ_{CTSI} is either 2, 4 or 6 seconds.

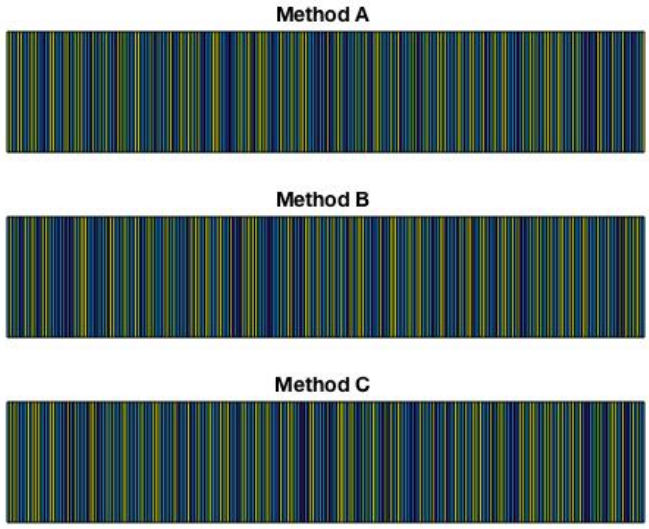


Figure 4.2: Maximin Designs Obtained by Method A, Method B and Method C for Estimating the HRF of $AR(2)$ Model for Experiments With Varying Time Between Components of Compound Trials When $(Q, N) = (1, 255)$.

We compare the efficiencies of the deigns obtained by our proposed methods to some well-known designs in the literature that are recommended for estimating the HRF under linear models for simple trials experiments. We generate m-sequence and 10 random designs for each of the following scenarios $(Q, N) = (1, 255), (2, 242)$. The m-sequence designs are generated following Liu and Frank (2004) with some modifications to accommodate the compound trials case. The minRE-values for these deigns are reported in Table 4.6. We notice that the maximin designs obtained by Methods A and C outperform other designs. The boxplots in Figure 4.3 display the distribution of the RE-values over $\boldsymbol{\Omega}$ for the maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by Method A, Method C, an m-sequence design and a random design. We notice that the RE-values for $\boldsymbol{d}^*_{Mm,MethodA}$ and $\boldsymbol{d}^*_{Mm,MethodC}$ are ranging from 0.9 to 1 which implies high relative efficiencies for most of the correlation coefficients. However, the RE-values for m-sequence and random designs are very low at the worst cases.

Table 4.6: $minRE_{\phi\in\Omega}(.;\boldsymbol{d}^*_\phi)$ of $\boldsymbol{d}^*_{Mm}$ From Methods A and C Versus Some Traditional fMRI Designs for Estimation Purpose of $AR(2)$ Model for Experiments With Varying Time Between Components of Compound Trials.

	Q=1	Q=2
$\boldsymbol{d}^*_{Mm,MethodA}$	0.8935	0.9210
$\boldsymbol{d}^*_{Mm,MethodC}$	0.8866	0.9276
m-sequence-based design	0.5828	0.6426
10 random designs	0.5178-0.6681	0.5416-0.6429

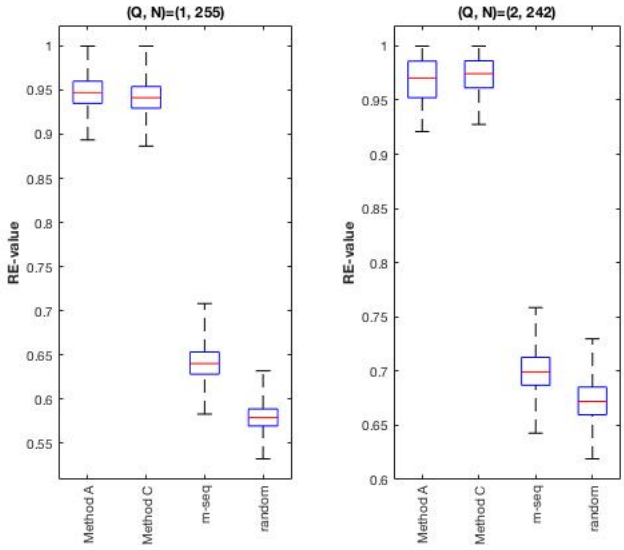


Figure 4.3: Boxplots of RE-Values Over $\boldsymbol{\Omega}$ of $AR(2)$ Model for Designs Obtained by Method A, Method C, an m-Sequence Design, and a Random Design for Estimating the HRF in Compound Trial Experiments With Varying Time Between the Components.

In a similar manner as we did for the estimation case, we obtain maximin designs for detecting brain activations by using our proposed methods and compare them to the grid method by Maus et al. (2010). Once again, τ_{CTSI} is allowed to equal $2, 4$, or 6 seconds for the simulations for detection purposes in Case (2). Table 4.7 represents the time spent in hours to generate the LODs $\boldsymbol{d}^*_{\boldsymbol{\phi}}$ for Methods A, B and C, that maximize $F_d(\boldsymbol{d},\boldsymbol{\phi}) = m_\theta/trace(\boldsymbol{C}_\theta\boldsymbol{M}^{-1}(\boldsymbol{d},\boldsymbol{\phi})\boldsymbol{C}'_\theta)$ where $\boldsymbol{M}(\boldsymbol{d};\boldsymbol{\phi})$ as in (4.4).

The efficiencies of the maximin designs obtained by our proposed methods compared to Method B for Case (2) are reported in Table 4.8 with the correlation parameters values where the $minRE$-value has occurred.

Table 4.7: CPU Time Spent by the Genetic Algorithm on Maximizing $F_d(\boldsymbol{d}, \boldsymbol{\phi})$ for $AR(2)$ Model With $Q = 1$and 2 for Experiments With Varying Time Between Components of Compound Trials.

Method	number of LODs	Total time spent (hrs.): Q=1	Q=2
A	20	0.1	0.3
B	1797	7	23
C	6	0.03	0.1

Table 4.8: The Performances of Method A, Method B, and Method C for Detecting Activated Brain Regions Over the Parameter Space $\boldsymbol{\Omega}$ of $AR(2)$ Model and the Corresponding (ϕ_1, ϕ_2)-Values That Yield the Maximin Designs $\boldsymbol{d}^*_{Mm}$ for Experiments With Varying Time Between Components of Compound Trials.

	Q= 1	Q= 2
Method A:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9450	0.9497
time spent (hrs.)	0.03	0.1
(ϕ_1, ϕ_2)	(0.01,0)	(0.07,0.04)
Method B:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9691	0.9745
time spent (hrs.)	5	4.5
(ϕ_1, ϕ_2)	(0.2,0.42)	(0.07,0.04)
Method C:		
$minRE_{\phi\in\Omega}(\boldsymbol{d}^*_{Mm}; \boldsymbol{d}^*_{\phi})$	0.9608	0.9509
time spent (hrs.)	0.02	0.03
(ϕ_1, ϕ_2)	(0.01,0)	(0.07,0.04)

For both scenarios, the efficiencies of the obtained designs by all three methods are at least 95%. The proposed approaches have done a good job in finding maximin designs that are as efficient as the one obtained by the expensive method (Method B)

with a huge reduction in the CPU time by at least 99.9%. Once again, the reported time in table 4.8 does not include the time needed for generating the locally optimal designs which is in Table 4.7. The obtained maximin designs $\boldsymbol{d}^*_{Mm}$ for Case (1) and Case (2) are presented in Figure 4.4 for experiment with a single trial type and Method A applied. We notice that for both cases, the obtained designs are not in the form of block designs which is in contrast to the case of simple trials as shown previously in Figure 3.7. As mentioned before, the control event is represented by the dark blue bars while the other colors represent the compound trials with different τ_{CTSI} intervals between the two components. For instance, the yellow bars in the design of Case (1) represents the trial type where the mental task is presented after 4 seconds following the cue. The maximin design for Case (2) has less rest (control) events than Case (1) design. We compare the detection power of the maximin designs

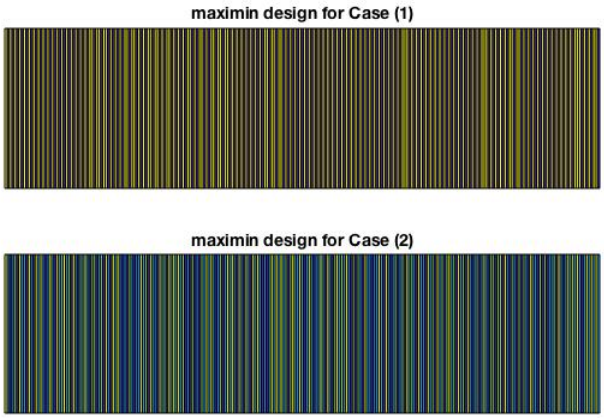


Figure 4.4: Maximin Designs for Detection Purposes Obtained by Method A for Cases When $\tau_{CTSI} = 4$ Seconds and $\tau_{CTSI} = \{2, 4, 6\}$ Seconds of $AR(2)$ Model for Compound Trials Experiments When $(Q, N) = (1, 255)$.

obtained by the three method for both cases by evaluating them over the parameter space Ω of $AR(2)$. We observe that Case (2) designs have higher detection power than those for Case (1) as shown in Figure 4.5.

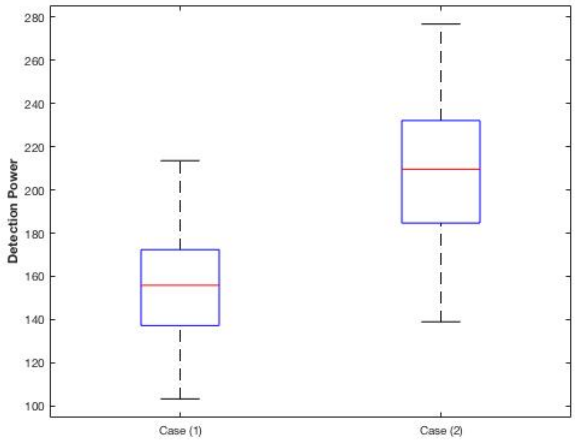


Figure 4.5: Detection Power of the Maximin Designs Obtained by Method A for Cases When $\tau_{CTSI} = 4$ Seconds and $\tau_{CTSI} = \{2, 4, 6\}$ Seconds of $AR(2)$ Model for Compound Trials Experiments When $(Q, N) = (1, 255)$.

Furthermore, the performances of designs obtained by Methods A and C are compared to a set of different traditional designs. We note that, each of the traditional design is generated by taking into account that the time between components of a compound trial is varying. We generated two block designs with blocks of size 4 for $Q = 1$ and of size 6 for $Q = 2$. We also included m-sequence based designs, three mixed designs considering the same combinations as described in Subsection 4.3.1 and a total of 20 randomly generated designs for each of the considered scenarios. The $minRE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.; \boldsymbol{d}^*_{\boldsymbol{\phi}})$ for these deigns are reported in Table 4.9. We observe that block designs are the worst for detection purposes when experiments include compound trials whereas m-sequence and random designs are notably outperformed the block design here. Once again, the efficiencies of the mixed designs are improving as the fraction of the design is made from an m-sequence designs increases (or as the fraction of the design made from a block design decreases). Overall the maximin designs obtained by our proposed approaches outperform the other designs for detection purposes.

Table 4.9: $minRE_{\phi\in\Omega}(.;\boldsymbol{d}^*_\phi)$ of $\boldsymbol{d}^*_{Mm}$ From Methods A and C Versus a Block Design, an m-Sequence Design, 3 Mixed Designs and 10 Random Designs for Detection Purposes When $(Q,N)=(1,255)$ and $(2,242)$ of $AR(2)$ Model for Experiments With Varying Time Between Components of Compound Trials.

	Q=1	Q=2
$\boldsymbol{d}^*_{Mm,MethodA}$	0.9450	0.9497
$\boldsymbol{d}^*_{Mm,MethodC}$	0.9608	0.9509
block design	0.5461	0.5516
m-sequence design	0.7176	0.7635
mixed 1 design	0.6501	0.7109
mixed 2 design	0.6264	0.6631
mixed 3 design	0.5768	0.5915
10 random designs	0.6752-0.7525	0.7248-0.7679

Figure 4.6 represents a comparison between the distribution of the RE-values over $\boldsymbol{\Omega}$ of $AR(2)$ model for maximin designs $\boldsymbol{d}^*_{Mm}$ obtained by Method A and Method C versus the other traditional designs. Each box in Figure 4.6 represents the distribution

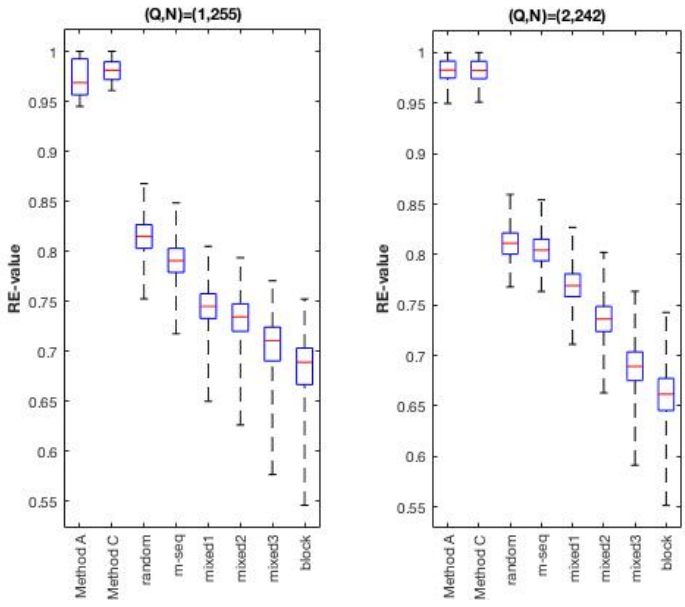


Figure 4.6: Boxplots of RE-Values Over $\boldsymbol{\Omega}$ of $AR(2)$ Model for Designs Obtained by Method A, Method C, a Block Designs, an m-Sequence Design, a Mixed Design and a Random Design for Detection Case for Experiments With Varying Time Between Components of Compound Trials.

of the 1797 $RE_{\boldsymbol{\phi}\in\boldsymbol{\Omega}}(.;\boldsymbol{d}^*_{\boldsymbol{\phi}})$ values for each of the competing designs. We observe that more variation in RE- values occurred for traditional designs when compared to the distribution of REs for Methods A and C. The block designs provide the lowest RE-values at the worst cases. In conclusion, maximin designs obtained by Methods A and C show high relative efficiencies for all correlation coefficients.

4.4 Discussion

In this chapter, we considered more complicated structure of event-related fMRI experiments, where each trial consists of multiple components rather than one as was discussed in Chapter 3. When dealing with such experiments, there are two cases that should be taking into account, namely (1) the case with a fixed time between components of the compound trial and (2) the case with varying time between components. The construction of the design matrix for both cases was explained by using examples.

We aim to find optimal designs that are robust against the misspecification of the error correlation when compound trials are in use. To do this, we applied our proposed approaches that were first introduced in Section 3.3 and Subsection 3.3.2. The results were obtained for the objective of estimating the HRF and detecting activated brain regions. Our methods has shown to be the ones with the highest efficiencies when compared to other designs. However, after running the code multiple times for estimation purposes with varying τ_{CTSI} (i.e., Case (2)), we noticed the efficiencies of the obtained maximin designs were slightly fluctuated after each run, especially for when $Q = 2$ as shown in Table 4.10. This has only occurred when compound trials are involved for Case (2) in the experiment. We have not observed similar behavior for simple trial experiments. This could be due to the complexity and the size of the problem when compound trials are used in the experiments.

Table 4.10: The Results of Multiple Runs for Estimating the HRF When the Time Interval Is Allowed to Vary Between Components of Compound Trials.

Method A	Method B	Method C
0.9180	0.9280	0.9276
0.8866	0.9315	0.9484
0.9331	0.8980	0.9641

To overcome this issue, we tried to maximize the size of the first generation in the genetic algorithm from 20 to 40 designs. In addition, we considered the first stopping rule provided in Kao (2009), where the algorithm will be terminated after $10,000$ iterations. Both settings did not make a significant change in the obtained results; besides that, the computing time needed to achieve these results is almost double the time spent when a smaller population size and the second stopping rule are considered.

Chapter 5

SUMMARY AND CONCLUSION

In this dissertation, we propose two efficient approaches in order to reduce the computational resource needed for obtaining maximin event-related fMRI designs with uncertain autocorrelation coefficients. In the first approach, we combine the Kriging approximation method with a knowledge-based genetic algorithm. This approach helps to approximate the surface of the RE-values of each design over the parameter space of the correlation coefficients by using a small subset of the parameter space. We consider a sampling plan that has a space-filling property, such as the well-known Sobol sequences, to sample the small set of correlation parameters from the specified region of interest.

Two common statistical objectives in the analysis of event-related fMRI experiments were discussed in this dissertation, which are the estimation of the HRF and the detection of activated brain regions. For these two common objectives, our proposed method can obtain maximin designs that perform similarly to designs obtained by the grid method considered by Maus et al. (2010) with a significant reduction in time considering our method. When approximating the objective function to be optimized, we considered the ordinary Kriging model. This selection leads to results that are quite satisfactory for both estimation and detection purposes. There might be other approximation methods that can be applied to the current situation and may provide slightly better results. For our simulations, we adopted the standard settings for Kriging model that were used in the literature which provided good results for our cases. A possible extension is to consider other settings and explore their performance in obtaining maximin designs. Our proposed method has been applied to the case of

ultra-fast fMRI experiments that allow a relatively high temporal resolution where the sampling rate τ_{TR} is considered to be as small as tens or hundreds of milliseconds (Alghamdi, Alrumayh, & Kao, 2019). Our method has shown to be very effective in achieving high efficient maximin designs for this pioneering brain mapping technique.

Another efficient shortcut method is proposed in this dissertation, where points located at the boundaries of the parameter space of interest are used without the use of the Kriging model. This method helps to achieve high-quality maximin designs in a very short time. A future research of interest is to consider to work on higher order $AR(p), p > 3$, models to test the usefulness of the boundary points method in obtaining good maximin event-related fMRI designs. The two proposed methods seem to be useful for fast fMRI studies.

In the first part of this study, we consider simple trial experiments of 8 to 9 minutes for $Q = 1, 2$ and 3. Our proposed methods provide superior results in comparison to some traditional designs that are recommended for estimation or detection purposes. For this case the noise is assumed to follow high order autoregressive models such as $AR(2)$ and $AR(3)$.

In the second part of this dissertation, experiments with compound trials were introduced. Our proposed methods are applied to this case to find high-quality maximin event-related designs. In our case studies, we consider experiments of 32 to 34 minutes for compound trials with one stimulus type and two stimulus types, respectively. For estimation, the HRF parameters can all be identifiable only when the time interval between the components of the compound trial is allowed to vary. To satisfy the assumption of additive HRFs, the time interval between the two components should not be less than 2 seconds and not more than $\tau_{ICI} - 2$ seconds. This is to ensure that components of compound trials are not too close to each other. For detection, both fixed and varied time intervals are allowed. We observe that maximin

designs obtained for Case (1) have lower detection power than Case (2) designs. Our recommendation is to consider varying time interval between components of the compound trails for the purpose of detecting the activated brain regions. Furthermore, the performances of our proposed approaches for these complicated type of experiments outperform those for other designs that are known in the literature. Here, the noise is assumed to follow $AR(2)$ model. Some recommendations for future research are (1) to consider another search algorithm that overcomes the randomness issue of the genetic algorithm and provides more consistent results, especially for the objective of estimating the HRF, and (2) another possibility is to relax the assumption of the restricted variation on the possible time between the components and allow it to vary throughout the experiment.

REFERENCES

Alghamdi, R., Alrumayh, A., & Kao, M.-H. (2019). Maximin designs for ultra-fast functional brain imaging. In *Contemporary biostatistics with biopharmaceutical applications* (pp. 115–128). Springer.

Amaro, E., Brammer, M., Williams, S., Andrew, C., Curtis, V., Ahmad, F., ... McGuire, P. (1999). Event-related fmri without scanner acoustic noise. In *Schizophrenia research* (Vol. 36, pp. 217–217).

Antonov, I. A., & Saleev, V. (1979). An economic method of computing lpτ-sequences. *USSR Computational Mathematics and Mathematical Physics*, *19*(1), 252–256.

Berger, M. P., & Tan, F. E. (2004). Robust designs for linear mixed effects models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *53*(4), 569–581.

Bratley, P., & Fox, B. L. (1988). Algorithm 659: Implementing sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software (TOMS)*, *14*(1), 88–100.

Buckner, R. L. (1998). Event-related fmri and the hemodynamic response. *Human brain mapping*, *6*(5-6), 373–377.

Buračas, G. T., & Boynton, G. M. (2002). Efficient design of event-related fmri experiments using m-sequences. *Neuroimage*, *16*(3), 801–813.

Buxton, R., Liu, T., Martinez, A., Frank, L., Luh, W.-M., & Wong, E. (2000). Sorting out event-related paradigms in fmri: The distinction between detecting an activation and estimating the hemodynamic response. *NeuroImage*, *5*(11), S457.

Chi, H., & Mascagni, M. (2004). Scrambled quasirandom sequences and their applications. *The florida state university, college of arts and science*.

Dale, A. M. (1999). Optimal experimental design for event-related fmri. *Human brain mapping*, *8*(2-3), 109–114.

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

D'Esposito, M., Zarahn, E., & Aguirre, G. K. (1999). Event-related functional mri: implications for cognitive psychology. *Psychological bulletin*, *125*(1), 155.

Eck, J., Kaas, A. L., & Goebel, R. (2013). Crossmodal interactions of haptic and visual texture information in early sensory cortex. *Neuroimage*, *75*, 123–135.

Fang, K.-T. (1980). Uniform design: application of number-theoretic methods in experimental design. *Acta Math. Appl. Sin.*, *3*, 363–372.

Fang, K.-T., Li, R., & Sudjianto, A. (2006). *Design and modeling for computer experiments*. CRC Press.

Fang, K.-T., Lin, D. K., Winker, P., & Zhang, Y. (2000). Uniform design: theory and application. *Technometrics*, *42*(3), 237–248.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998). Event-related fmri: characterizing differential responses. *Neuroimage*, *7*(1), 30–40.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fmri. *Neuroimage*, *10*(5), 607–619.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, *57*(298), 369–375.

Henson, R. (2003). Analysis of fmri time series: Linear time-invariant models, event-related fmri and optimal experimental design. Elsevier.

Henson, R., & Friston, K. (2007). Convolution models for fmri. *Statistical parametric mapping: The analysis of functional brain images*, 178–192.

Holland, J. H. (1975). Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.

Huettel, S. A. (2012). Event-related fmri in cognition. *Neuroimage*, *62*(2), 1152–1156.

Joe, S., & Kuo, F. Y. (2008). Notes on generating sobol sequences. *ACM Transactions on Mathematical Software (TOMS)*, *29*(1), 49–57.

Johnson, M. E., Moore, L. M., & Ylvisaker, D. (1990). Minimax and maximin distance designs. *Journal of statistical planning and inference*, *26*(2), 131–148.

Jones, B., & Johnson, R. T. (2009). Design and analysis for the gaussian process model. *Quality and Reliability Engineering International*, *25*(5), 515–524.

Josephs, O., Turner, R., & Friston, K. (1997). Event-related f mri. *Human brain mapping*, *5*(4), 243–248.

Kao, M.-H. (2009). Multi-objective optimal experimental designs for er-fmri using matlab. *Journal of Statistical Software*, *30*(11), 1–13.

Kao, M.-H. (2014). A new type of experimental designs for event-related fmri via hadamard matrices. *Statistics & Probability Letters*, *84*, 108–112.

Kao, M.-H., Majumdar, D., Mandal, A., & Stufken, J. (2013). Maximin and maximin-efficient event-related fmri designs under a nonlinear model. *The Annals of Applied Statistics*, *7*(4), 1940–1959.

Kao, M.-H., Mandal, A., Lazar, N., & Stufken, J. (2009). Multi-objective optimal experimental designs for event-related fmri studies. *NeuroImage*, *44*(3), 849–856.

Kao, M.-H., Mandal, A., & Stufken, J. (2008). Optimal design for event-related functional magnetic resonance imaging considering both individual stimulus effects and pairwise contrasts. *Statistics and applications*, *6*, 235–256.

Kao, M.-H., & Mittelmann, H. D. (2014). A fast algorithm for constructing efficient event-related functional magnetic resonance imaging designs. *Journal of Statistical Computation and Simulation*, *84*(11), 2391–2407.

Kao, M.-H., & Zhou, L. (2017). Optimal experimental designs for fmri when the model matrix is uncertain. *NeuroImage*, *155*, 594–604.

Kashou, N. H. (2014). A practical guide to an fmri experiment.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 272–319.

Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2011). Emergence of perceptual gestalts in the human visual cortex: The case of the configural-superiority effect. *Psychological science*, *22*(10), 1296–1303.

Kucherenko, S., Albrecht, D., & Saltelli, A. (2015). Exploring multi-dimensional spaces: A comparison of latin hypercube and quasi monte carlo sampling techniques. *arXiv preprint arXiv:1505.02350*.

Lenoski, B., Baxter, L. C., Karam, L. J., Maisog, J., & Debbins, J. (2008). On the performance of autocorrelation estimation algorithms for fmri analysis. *IEEE Journal of Selected Topics in Signal Processing*, *2*(6), 828–838.

Lindquist, M. A. (2008). The statistical analysis of fmri data. *Statistical science*, *23*(4), 439–464.

Liu, T. T., & Frank, L. R. (2004). Efficiency, power, and entropy in event-related fmri with multiple trial types: Part i: Theory. *NeuroImage*, *21*(1), 387–400.

Locascio, J. J., Jennings, P. J., Moore, C. I., & Corkin, S. (1997). Time series analysis in the time domain and resampling methods for studies of functional magnetic

resonance brain imaging. *Human brain mapping*, *5*(3), 168–193.

Lophaven, S. N., Nielsen, H. B., & Søndergaard, J. (2002). *Dace: a matlab kriging toolbox* (Vol. 2). Citeseer.

Martin, J. D., & Simpson, T. W. (2005). Use of kriging models to approximate deterministic computer models. *AIAA journal*, *43*(4), 853–863.

Matheron, G. (1963). Principles of geostatistics. *Economic geology*, *58*(8), 1246–1266.

Maus, B., Van Breukelen, G. J., Goebel, R., & Berger, M. P. (2010). Robustness of optimal design of fmri experiments with application of a genetic algorithm. *NeuroImage*, *49*(3), 2433–2443.

Maus, B., van Breukelen, G. J., Goebel, R., & Berger, M. P. (2012). Optimal design for nonlinear estimation of the hemodynamic response function. *Human brain mapping*, *33*(6), 1253–1267.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239–245.

Morris, M. D., & Mitchell, T. J. (1995). Exploratory designs for computational experiments. *Journal of statistical planning and inference*, *43*(3), 381–402.

Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A physiological correlate of the zoom lens of visual attention. *Journal of Neuroscience*, *23*(9), 3561–3565.

Müller, N. G., & Kleinschmidt, A. (2003). Dynamic interaction of object-and space-based attention in retinotopic visual areas. *Journal of Neuroscience*, *23*(30), 9812–9816.

Müller, N. G., & Kleinschmidt, A. (2004). The attentional spotlight's penumbra: center-surround modulation in striate cortex. *Neuroreport*, *15*(6), 977–980.

Niederreiter, H. (1988). Low-discrepancy and low-dispersion sequences. *Journal of number theory*, *30*(1), 51–70.

Niederreiter, H. (1992). *Random number generation and quasi-monte carlo methods* (Vol. 63). Siam.

Pochon, J.-B., Levy, R., Poline, J.-B., Crozier, S., Lehéricy, S., Pillon, B., ... Dubois, B. (2001). The role of dorsolateral prefrontal cortex in the preparation of forthcoming actions: an fmri study. *Cerebral Cortex*, *11*(3), 260–266.

Purdon, P. L., Solo, V., Weisskoff, R. M., & Brown, E. N. (2001). Locally regularized spatiotemporal modeling and model comparison for functional mri. *NeuroImage*, *14*(4), 912–923.

Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nature neuroscience*, *3*(9), 940.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, 409–423.

Saleh, M., Kao, M.-H., & Pan, R. (2017). Design d-optimal event-related functional magnetic resonance imaging experiments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *66*(1), 73–91.

Santner, T. J., Williams, B. J., & Notz, W. (2003). *The design and analysis of computer experiments*. Springer Science & Business Media.

Serences, J. T. (2004). A comparison of methods for characterizing the event-related bold timeseries in rapid fmri. *Neuroimage*, *21*(4), 1690–1700.

Silver, M. A., Ress, D., & Heeger, D. J. (2007). Neural correlates of sustained spatial attention in human early visual cortex. *Journal of neurophysiology*, *97*(1), 229–237.

Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, *7*(4), 784–802.

Sobol, I. M. (1976). Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, *16*(5), 236–242.

Stinstra, E., den Hertog, D., Stehouwer, P., & Vestjens, A. (2003). Constrained maximin designs for computer experiments. *Technometrics*, *45*(4), 340–346.

Trosset, M. W. (1999). Approximate maximin distance designs. In *Proceedings of the section on physical and engineering sciences* (pp. 223–227).

Wager, T. D., & Nichols, T. E. (2003). Optimization of experimental design in fmri: a general framework using a genetic algorithm. *Neuroimage*, *18*(2), 293–309.

Worsley, K. J., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., & Evans, A. (2002). A general statistical analysis for fmri data. *Neuroimage*, *15*(1), 1–15.

APPENDIX A

GENETIC ALGORITHM

Genetic Algorithms are very common to use in solving optimization problems. They became very popular after the work proposed by Holland (1975). The idea is to use good solutions (parents) to generate better ones (offsprings). Following the genetic algorithm of Kao (2009) which makes use of well known event-related fMRI designs, such as block designs, m-sequence, random designs, and mixed designs, the outline of the algorithm is as follows:

Step 1: Generate $2G$ initial designs (e.g., m-sequences, block designs, and mixed designs) that form the first generation. The fitness of these designs is then evaluated through the objective function, which can be the selected optimality criterion.

Step 2: With probability proportional to the fitness, G pairs of distinct designs are selected with replacement to generate offspring designs via the crossover operator. That is, to select a random cut-point and exchange the corresponding subsequences of the paired designs.

Step 3: The mutation operator then randomly selects $q\%$ of the elements of the $2G$ offspring designs, and randomly perturbs these elements. Then, obtain the fitness of the resulting designs.

Step 4: Add to the population another N immigrant drawn from random and block designs, and their combinations, and obtain their fitness.

Step 5: Create a pool to include the designs of the current generation, offspring designs, and immigrants. According to their fitness, keep the best $2G$ designs in the pool to form the next generation, and discard the others.

Step 6: The process repeats until a stopping rule is met; for example, until no significant improvement is made. The algorithm keeps track of the design with the best fit over all generations.

The Stopping Rule: We consider the second stopping rule provided in Kao (2009), where the algorithm will be terminated if there is no significant improvement. For instance, the algorithm performs a check every g generations, then it will compute the improvement in overall efficiency from generation 1 to generation g, and then for each set of g generations. We assume $g = 200$, if the improvement in the value of the objective function in the last 200 generations is no more than 10^{-7} of that of the first 200 generations, the search is stopped (Kao & Mittelmann, 2014).