fMRI Design under Autoregressive Model with One Type of Stimulus

by

Chuntao Chen

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2017 by the
Graduate Supervisory Committee:

John Stufken, Chair
Mark Reiser
Ioannis Kamarianakis

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

Functional magnetic resonance imaging (fMRI) is used to study brain activity due to stimuli presented to subjects in a scanner. It is important to conduct statistical inference on such time series fMRI data obtained. It is also important to select optimal designs for practical experiments. Design selection under autoregressive models have not been thoroughly discussed before. This paper derives general information matrices for orthogonal designs under autoregressive model with an arbitrary number of correlation coefficients. We further provide the minimum trace of orthogonal circulant designs under AR(1) model, which is used as a criterion to compare practical designs such as M-sequence designs and circulant (almost) orthogonal array designs. We also explore optimal designs under AR(2) model. In practice, types of stimuli can be more than one, but in this paper we only consider the simplest situation with only one type of stimuli.

ACKNOWLEDGMENTS

Lots of thanks and best wishes to my supervisor, John Stufken.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

Functional magnetic resonance imaging (fMRI) is a technique to illustrate brain activity. The magnetic resonance (MR) scanner will scan the brain when a series of mental stimuli (such as sound and picture) is presented. A change of an oxy to deoxy concentration at some brain voxels (small cuboid of size 3x3x5 mm^3, Lazar (2008)) will be caught by the scanner. A fMRI experiment can contain several types of stimuli. This paper discusses the simplest case, one type of stimuli. This stimuli will be presented to the human subject repeatedly with rest time between two stimuli onsets when no stimuli is given. Thus, the design of a fMRI experiment will be a sequence 1001010010011110..., of which "0" represents rest time and "1" represents that stimuli is given at that time. The length of series (run size) of stimuli can be several hundred.

The main objective of a fMRI experiment is to study the response of the brain to mental stimuli. The response of every voxel will lead to a time series from the brain. This response is described by a smooth function over time called Hemodynamic Response Function (HRF). If there were only one stimuli, HRF would just have one peak and return to the baseline after some seconds. Since lots of stimuli are presented, the accumulated HRF will look like a long wave with lots of peaks. The linear model is popularly used to describe this function (Worsley and Friston (1995);Dale (1999)). In the linear model, the errors are usually assumed to be uncorrelated. This paper will discuss the autoregressive model where error terms are correlated.

There are several standard ways to obtain fMRI design. M-sequence is widely used to construct fMRI designs. This type of design has good mathematical properties

(Liu (2004) and Jansma *et al.* (2013)). A good property of an M-sequence is any non-zero 2-tuples, 11, 10, 01 (2 is specifically for one type of stimuli) occurs equally frequently. The main drawback is that the length (run size n) of M-sequences is restricted to $n = (p+1)^l - 1$, where l is an positive integer, $p+1$ is prime and p is the number of type of stimuli, which is 1 in this paper. For example, 0111001 is an M-sequence. Thus, the gap between adjacent run sizes can be large. In order to make run size flexible, an extended m-sequence (Kao (2013)) and CAOA (circulant almost orthogonal array)(Lin *et al.* (2016)) are used. An extended M-sequence is to insert one additional 0 such that any 2-tuple (including 00) occurs equally. For example, the M-sequence 0111001 will change to an extended M-sequence 00111001 if another 0 is added at the beginning of the sequence.

CAOA designs in Lin *et al.* (2016) have more flexible run sizes such as $n \equiv 1(mod 4)$ and $n \equiv 3(mod 4)$. This type of design has a small deviation from a special type of design, orthogonal design, which will be discussed later. Another popular way to find an optimal design is using the computational method. Genetic algorithm is used to generate optimal designs (Kao (2009)). Kao's genetic algorithm allows generation of optimal designs under the autoregressive(AR) model with one correlation coefficient. This process is done by using a whitening matrix to make the error uncorrelated. However, there is not much theoretical work on AR model. In the next section we will evaluate performance of these different types of designs under AR models.

Chapter 2

RESULT

1. Information matrices under autocorrelation models and orthogonal designs

Let y be a n by 1 vector of bold signals collected by a fMRI scanner on a single voxel of the brain. It is assumed that once a stimuli occurs, after k time periods (for example, 4 seconds), the HRF returns to the baseline if no other stimuli follow. k=9 and 17 are commonly selected in practice (Cheng *et al.* (2017)).

Consider the autocorrelation model as follows: y $=1_n\mu + X\tau + \epsilon$, where X is a n by k design matrix with entries 1 or 0. The reason that X has k columns is that after k time periods, since the HRF will completely drop down to baseline, there is no need to estimate parameters besides k time periods. The first column of X is the sequence of stimuli presented to subject. Every other column is a permutation of first column. Such design is called circulant design.

$$X = \begin{bmatrix} x_1 & x_n & x_{n-1} & \dots & x_{n-k+2} \\ x_2 & x_1 & x_n & \dots & x_{n-k+3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & x_{n-1} & x_{n-2} & \dots & x_{n-k+1} \end{bmatrix}$$

Under the case of one stimuli, design of which any 2 tuples (combination of 1 and 0: 11,10,01,00) occurs equally, is called orthogonal design. Orthogonal designs have properties such as $\Sigma x_i^2 = \Sigma x_i = n/2$ and $\Sigma x_i x_{i-l} = n/4$, $1 \leq l \leq k-1$. We also denote that $x_{-1} = x_n$ and $x_{-j} = x_{n-j+1}$. Assuming ϵ are autocorrelated, we need to estimate

τ. Starting from simplest AR model, AR(1) model with correlation coefficient λ_1, the whitening matrix is as follows:

$$H = \begin{bmatrix} 1 & 0 & 0 & ... & 0 & -\lambda_1 \\ -\lambda_1 & 1 & 0 & ... & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & \ddots & 0 & 0 \\ 0 & 0 & ... & -\lambda_1 & 1 & 0 \\ 0 & 0 & ... & 0 & -\lambda_1 & 1 \end{bmatrix}$$

The reason to choose this whitening matrix is that we assumed any two adjacent data points are correlated with coefficient λ_1. It is mathematically simple to arrange the whitening matrix like this. However, the first data point and last data point are not correlated in practice.

Under least squares estimation, the variance of τ is:
$\sigma^2[X^T H^T H X - (1/(1_n^T H^T H 1_n)) X^T H^T H 1_n 1_n^T H^T H X]^{-1}$. Thus, the information matrix of τ is the subtraction of two matrices:
$C = X^T H^T H X - (1/(1_n^T H^T H 1_n)) X^T H^T H 1_n 1_n^T H^T H X$. Observe that:

$$X^T H^T H X = \begin{bmatrix} d & e_1 & e_2 & ... & e_j & ... & ... & .. & ... & e_{k-2} & e_{k-1} \\ e_1 & d & e_1 & e_2 & ... & e_j & ... & ... & ... & ... & e_{k-2} \\ e_2 & e_1 & d & e_1 & e_2 & ... & ... & ... & ... & ... & ... \\ ... & e_2 & e_1 & d & e_1 & e_2 & ... & e_j & ... & ... & ... \\ e_j & ... & e_2 & e_1 & d & e_1 & ... & ... & e_j & ... & ... \\ ... & e_j & ... & e_2 & e_1 & d & ... & ... & ... & e_j & ... \\ ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\ ... & ... & ... & e_j & ... & ... & ... & d & e_1 & e_2 & ... \\ ... & ... & ... & ... & e_j & ... & ... & e_1 & d & e_1 & e_2 \\ e_{k-2} & ... & ... & ... & ... & e_j & ... & e_2 & e_1 & d & e_1 \\ e_{k-1} & e_{k-2} & ... & ... & ... & ... & ... & ... & e_2 & e_1 & d \end{bmatrix}$$

where $d = \Sigma_{i=1}^n (x_j - \lambda_1 x_{i-1})^2$ and $e_j = \Sigma_{i=1}^n (x_i - \lambda_1 x_{i-1})(x_{i-j} - \lambda_1 x_{i-1-j}) =$

$(1+\lambda_1^2)\Sigma_{i-1}^n x_i x_{i-j} - \lambda_1 \Sigma_{i=1}^n x_i x_{i-j+1} - \lambda_1 \Sigma_{i=1}^n x_i x_{i-j-1}$.

For an orthogonal design,

$d = (1+\lambda_1^2)\frac{n}{4} - \lambda_1\frac{n}{4} - \lambda_1\frac{n}{4} = (1-\lambda_1+\lambda_1^2)\frac{n}{4}$. For the e_j, there are three cases,

Case 1: j=1

$e_1 = (1+\lambda_1^2)\Sigma_{i-1}^n x_i x_{i-1} - \lambda_1 \Sigma_{i=1}^n x_i^2 - \lambda_1 \Sigma_{i=1}^n x_i x_{i-2} = (1+\lambda_1^2)\frac{n}{4} - \lambda_1\frac{n}{2} - \lambda_1\frac{n}{4} = (1-3\lambda_1+\lambda_1^2)\frac{n}{4}$.

Case 2: $2 \le j \le k-2$

In this case, $i-1 \le i-j+1$ and $i-j-1 \ge i-k+1$. Hence, $\Sigma x_i x_{i-j+1} = \Sigma x_i x_{i-j-1} = \frac{n}{4}$ for an orthogonal design. Therefore, $e_j = (1+\lambda_1^2)\frac{n}{4} - \lambda_1\frac{n}{4} - \lambda_1\frac{n}{4} = (1-\lambda_1)^2\frac{n}{4}$.

Case 3: j=k-1

$e_{k-1} = (1-\lambda_1+\lambda_1^2)\frac{n}{4} - \lambda_1 \Sigma x_i x_{i-k}$.

We also observe that $H1_n = (1-\lambda_1)1_n$ and $1_n^T H^T H 1_n = n(1-\lambda_1)^2$. Thus, for orthogonal design,

$$X^T H^T H 1_n = (1-\lambda_1)^2 \frac{n}{2} 1_k.$$

The matrix after minus sign of the information matrix is

$(1/(1_n^T H^T H 1_n)) X^T H^T H 1_n 1_n^T H^T H X = n(1-\lambda_1)^{-2}[(1-\lambda_1)^2\frac{n}{2}]^2 1_k 1_k^T = (1-\lambda_1)^2 \frac{n}{4} 1_k 1_k^T$, of which every entry equals $(1-\lambda_1)^2\frac{n}{4}$, the same quantity in Case 2.

The information matrix of orthogonal design under AR(1) model is shown as below.

Following the same derivation of that of AR(1) model, the general form of information matrix under AR(p) model, where $p \le (k-1)/2$, is as in **Table 2.2**. This matrix is also symmetric for any k. In this k by k matrix, the diagonal entries are $d = (\Sigma_{i=1}^p \lambda_i^2 + 1)n/4$. Some off-diagonal entries are $e_j = (-\lambda_j + \lambda_{j+1}\lambda_1 + \lambda_{j+2}\lambda_2 + ... + \lambda_{p-j}\lambda_p)n/4$ for $j = 1, 2, ..., p$ (specially, $e_p = -\lambda_p n/4$). For those entries below zero's on the lower left of the matrix, they depend on different designs. e_{k-1} will contain p uncertainties, $\Sigma x_i x_{i-k}$, $\Sigma x_i x_{i-k-1}$, $\Sigma x_i x_{i-k-2}$,..., $\Sigma x_i x_{i-k-p+1}$. e_{k-2} will contain $p-1$ uncertainties, $\Sigma x_i x_{i-k}$, $\Sigma x_i x_{i-k-1}$, $\Sigma x_i x_{i-k-2}$,..., $\Sigma x_i x_{i-k-p+2}$. e_{k-p} is equal to

Table 2.1: Information Matrix of Orthogonal Design under AR(1) Model

$$C = \begin{bmatrix} (1+\lambda_1^2)n/4 & \frac{-\lambda_1 n}{4} & 0 & ... & 0 & \lambda_1[(n/4) - \Sigma x_i x_{i-k}] \\ \frac{-\lambda_1 n}{4} & (1+\lambda_1^2)n/4 & \frac{-\lambda_1 n}{4} & ... & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ddots & \ddots & \frac{-\lambda_1 n}{4} & 0 \\ 0 & 0 & ... & \frac{-\lambda_1 n}{4} & (1+\lambda_1^2)n/4 & \frac{-\lambda_1 n}{4} \\ \lambda_1[(n/4) - \Sigma x_i x_{i-k}] & 0 & ... & 0 & \frac{-\lambda_1 n}{4} & (1+\lambda_1^2)n/4 \end{bmatrix}$$

Table 2.2: Information Matrix of Orthogonal Designs under AR(p) Model

$$C = \begin{bmatrix} d & e_1 & e_2 & ... & e_j & ... & e_p & \vec{0}'_{k-2p-1} & ... & e_{k-2} & e_{k-1} \\ e_1 & d & e_1 & e_2 & ... & ... & ... & ... & ... & ... & e_{k-2} \\ e_2 & e_1 & d & e_1 & ... & ... & ... & ... & ... & ... & ... \\ ... & e_2 & e_1 & d & e_1 & ... & ... & ... & ... & e_p & ... \\ e_j & ... & e_2 & e_1 & d & e_1 & ... & ... & e_j & ... & e_p \\ ... & e_j & ... & e_2 & e_1 & d & ... & ... & ... & e_j & ... \\ e_p & ... & e_j & ... & e_2 & e_1 & ... & ... & ... & ... & e_j \\ \vec{0}_{k-2p-1} & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ ... & ... & ... & ... & ... & ... & ... & ... & ... & ... & ... \\ e_{k-2} & ... & ... & ... & ... & ... & ... & ... & e_1 & d & e_1 \\ e_{k-1} & e_{k-2} & ... & ... & e_p & ... & e_j & ... & e_2 & e_1 & d \end{bmatrix}$$

$\lambda_p[(n/4) - \Sigma x_i x_{i-k}]$.

2. AR(1) model

A-optimality is used in this paper. To get an optimal design, we need to minimize the variance of τ. But its variance is a matrix. Evaluating trace of this variance/covariance matrix is called A-optimality criterion (Dale (1999) and Friston *et al.* (1999)). Hence, the goal is to find a design which minimizes the trace of C^{-1}.

In this section, we explore the trace of C^{-1} in AR(1) model using calculus. Since

$\Sigma x_i x_{i-k}$ is the only uncertainty in the information matrix, it is convenient to set $\lambda_1[(n/4) - \Sigma x_i x_{i-k}] = (n/4)\lambda_1 b_1$, i.e., $\Sigma x_i x_{i-k} = (1-b_1)(n/4)$. Note that $b_1 \in [-1, 1]$ because $n/4 \le \Sigma x_i x_{i-k} \le n/2$.

The table below shows the trace of inverse information matrix computed by Maple, for $k = 3, 4, 5, 6, 7$ and 8. Correlation coefficient λ_1 is denoted as a in this table while b_1 is noted as b.

k	the trace of inverse information matrix
3	$\frac{3+3a^4+(-b^2+4)a^2}{(a^4+a^3b+ab+1)(a^2-ab+1)}$
4	$\frac{4+4a^6+(-2b^2+6)a^4+(-2b^2+6)a^2}{1+a^8+(-b^2+1)a^6+(-b^2+2b+1)a^4+(-b^2+1)a^2}$
5	$\frac{5+5a^8+(-3b^2+8)a^6+(-4b^2+9)a^4+(-3b^2+8)a^2}{(a^6+a^5b+ab+1)(a^4-a^3b+a^2-ab+1)}$
6	$\frac{6+6a^{10}+(-4b^2+10)a^8+(-6b^2+12)a^6+(-6b^2+12)a^4+(-4b^2+10)a^2}{1+a^{12}+(-b^2+1)a^{10}+(-b^2+1)a^8+(-b^2+2b+1)a^6+(-b^2+1)a^4+(-b^2+1)a^2}$
7	$\frac{7+7a^{12}+(-5b^2+12)a^{10}+(-8b^2+15)a^8+(-9b^2+16)a^6+(-8b^2+15)a^4+(-5b^2+12)a^2}{(a^8+a^7b+ab+1)(a^6-a^5b+a^4-a^3b+a^2-ab+1)}$
8	$\frac{8+8a^{14}+(-6b^2+14)a^{12}+(-10b^2+18)a^{10}+(-12b^2+20)a^8+(-12b^2+20)a^6+(-10b^2+18)a^4+(-6b^2+14)a^2}{1+a^{16}+(-b^2+1)a^{14}+(-b^2+1)a^{12}+(-b^2+1)a^{10}+(-b^2+2b+1)a^8+(-b^2+1)a^6+(-b^2+1)a^4+(-b^2+1)a^2}$

From the table above, the trace of C^{-1} follows a pattern. We assume that trace for any k will follow this pattern. Thus, trace of C^{-1} will be in the form of $\frac{num}{den}$, where $num = c_1 - c_2 b_1^2$ and $den = c_3 + 2c_4 b_1 - c_5 b_1^2$. This denotes that $c_1 = \Sigma_{i=1}^{k}(\lambda_1^{2(k-i)})(i(k-i+1))$, $c_2 = \Sigma_{i=1}^{k}(\lambda_1^{2(k-i)})((i-1))(k-i))$, $c_3 = \Sigma_{i=0}^{k}\lambda_1^{2i}$, $c_4 = \lambda_1^k$, $c_5 = \Sigma_{i=0}^{k-1}\lambda_1^{2i}$.

By taking the derivative of trace with respect to b_1, and set it zero, we can get a root $b_1 = r_1 = \frac{d_1 - \sqrt{d_2}}{2d_3}$ where $d_1 = c_1c_5 - c_2c_3$, $d_2 = d_1^2 - 4c_1c_2c_4^2$. It is proved in the appendix that the trace obtains a global minimum at this r_1. r_1 is larger than zero and is very close to zero when λ_1 is small. **Table 2.3** shows efficiency of extended orthogonal designs (i.e. designs when $b_1 = 0$) compared to optimal designs when $b_1 = r_1$. For some specific run sizes (such as extended M-sequence discussed later), extended orthogonal designs exist, but it is usually hard to construct optimal designs (they might not exist) when $b_1 = r_1$. As shown, extended orthogonal designs are already very efficient.

The minimum trace at $b_1 = r_1$ might not be obtained by any orthogonal designs in practice. Therefore, the minimum of trace formed above only serves as a criterion to evaluate other practical designs in this paper. There might be a way to construct or search optimal orthogonal design at $b_1 = r_1$, but it will be very difficult.

a. **Extended orthogonal design;**

We define extended orthogonal designs as designs that $\Sigma_{i=1}^{n} x_i x_{i-l} = n/4$ for arbitrary integer $l \leq k$. It follows that $b_1 = 0$ for extended orthogonal design. Designs generated from extended M-sequence have the property that $\Sigma_{i=1}^{n} x_i x_{i-l} = n/4$ for arbitrary integer $l \leq n$. Hence, under AR(1) model, extended M-sequence design will make b_1 zero. The efficiencies of extended orthogonal designs are shown in **Table 2.3**. Therefore, extended M-sequence designs are highly efficient, but run size n is limited to specific numbers such as 32, 64, 128, i.e. 2^m, $m \in \mathbb{N}$ ($p = 1$ in this case due to one type of stimuli).

b. **Circulant-almost-orthogonal array(CAOA);**

We also consider the non-orthogonal designs explained in Lin *et al.* (2016). **Table 2.4** shows the efficiency of some non-orthogonal designs. One design is generated by M-sequence with run size $n = 7$. By exploring the non-orthogonal designs, we can see some efficiencies are greater than 1. Thus, orthogonal designs are not necessarily optimal under AR(1) model. Some non-orthogonal designs perform better than optimal orthogonal design. There are several types of designs in the paper of Lin *et al.* (2016), where T2 is a special design where $n \equiv 2(mod 4)$. T2 designs are claimed to be very efficient under error-uncorrelated model. But under AR(1) model, T2 designs are not efficient.

3. AR(2) model

From previous empirical study(Lenoski *et al.* (2008)), AR(2) model fits the result better. Thus, we also explore the AR(2) model a little bit. The information matrix of orthogonal design under AR(2) model is shown in **Table 2.5**.

Since there are two design dependent uncertainties $\Sigma x_i x_{i-k}$ and $\Sigma x_i x_{i-(k+1)}$ for the information matrix of of τ under orthogonal design and the AR(2) model, the trace will be a bi-variate function of $\Sigma x_i x_{i-k}$ and $\Sigma x_i x_{i-(k+1)}$. It is hard to compute a general form of minimum trace, like what is done for the AR(1) model. We only consider the case when $\lambda_1 = 0.3$ and $\lambda_2 = 0.2$. We chose a smaller λ_2 while correlation between two adjacent errors should be stronger than that between nonadjacent errors. We use fmincon function with 1000x1000 starting points in Matlab to search the minimum trace in the range of $b_1 \in [-1, 1]$ and $b_2 \in [-1, 1]$. The optimal b_1 and b_2 are shown in **Table 2.6**. From this table, the optimal b_1 and b_2 are approaching zero from two directions when k increases. Therefore, the extended orthogonal (i.e. $b_1 = b_2 = 0$) designs obtain higher efficiency with larger k. It is also true that extended orthogonal design (i.e. $b_1 = b_2 = 0$) for some run size might not exist. We use this minimum trace to compare practical designs for run size 240, which is a commonly used run size in experiments. CAOA (circulant orthogonal (almost-) array) designs were not considered since we did not find a CAOA design of run size $n = 240$. We also compute the efficiencies of designs from genetic algorithm for run size $n = 240$. All of these designs are non-orthogonal. The efficiencies are very high for $k = 16, ..., 20$. Some efficiencies are larger than 1. Thus, non-orthogonal designs can be better than orthogonal design in terms of A-optimality.

Table 2.3: Efficiency of Extended Orthogonal Design under AR(1) Model

k	$\lambda_1 = 0.1$	$\lambda_1 = 0.3$	$\lambda_1 = 0.5$	$\lambda_1 = 0.8$
3	0.999	0.990421009	0.951164055	0.875960998
4	1	0.998849113	0.984453376	0.923333999
5	1	0.999869495	0.995142309	0.948984005
6	1	0.999985813	0.998525345	0.9648162
7	1	0.999998503	0.999564187	0.975308658
8	1	0.999999845	0.99987403	0.982539416
9	1	0.999999984	0.99996423	0.987626607
10	1	0.999999998	0.999989986	0.991240995
11	1	1	0.999997228	0.993817469
12	1	1	0.99999924	0.995653008
13	1	1	0.999999793	0.996956959
14	1	1	0.999999944	0.997879464
15	1	1	0.999999985	0.998529007
16	1	1	0.999999996	0.998984071
17	1	1	0.999999999	0.999301292
18	1	1	1	0.999521353
19	1	1	1	0.999673309
20	1	1	1	0.999777783

Table 2.4: Efficiency of CAOA(Circulant (Almost) Orthogonal Array) Design under AR(1) Model

k	n	$\lambda_1 = 0.1$	$\lambda_1 = 0.3$	$\lambda_1 = 0.5$	Type of design	Generating vector of CAOA design
6	17	0.9897	1.0017	1.0098	Not T2	1110100001001110
7	14	$**$	0.5344	$**$	$T2$	1001111101000
8	21	0.9913	1.0013	1.0121	Not T2	10101101111100100000
9	25	0.9941	1.006575	1.0222	Not T2	001110101111011000100100
10	30	0.8824	0.9971	1.0177	Not T2	00000011100110111110101101000 1
11	29	$**$	1.005156	$**$	Not T2	0001000100111100111111 0100101
12	33	$**$	1.0057	$**$	Not T2	00010000111101100111110101101 0001
13	37	$**$	1.1478	$**$	Not T2	001000010100010011000111110101101 1111
14	41	$**$	1.1432	$**$	Not T2	0101110000101101110111011110010110001 0000
15	39	$**$	0.9596	$**$	Not T2	00011101010011001011100111101011011 0000
16	45	$**$	1.1351	$**$	Not T2	01101011010000010100010011001111001111110100
17	49	$**$	1.1328	$**$	Not T2	000000010110010010111101001110011111010101110 0110
18	42	$**$	0.681	$**$	$T2$	00101100001010111011001111000010001101110 1
19	38	$**$	0.4802	$**$	$T2$	0110000101011110010011000010101111001 1
23	46	$**$	0.4749	$**$	$T2$	000010100110011010111100000101001100110101111 1

Table 2.5: Information Matrix of Orthogonal Design under AR(2) Model

$$C = \begin{bmatrix} (1+\lambda_1^2+\lambda_2^2)n/4 & (-\lambda_1+\lambda_1\lambda_2)n/4 & -\lambda_2 n/4 & \vec{0}'_{k-5} & \lambda_2[(n/4)-\Sigma x_i x_{i-k}] & \lambda_1[(n/4)-\Sigma x_i x_{i-k}]+\lambda_1\lambda_2(\Sigma x_i x_{i-k-1}-\frac{n}{4}) \\ (-\lambda_1+\lambda_1\lambda_2)n/4 & (1+\lambda_1^2+\lambda_2^2)n/4 & ... & ... & 0 & \lambda_2[(n/4)-\Sigma x_i x_{i-k}] \\ -\lambda_2 n/4 & \ddots & \ddots & \ddots & \vdots & \vec{0}_{k-5} \\ \vec{0}_{k-5} & 0 & \ddots & \ddots & (-\lambda_1+\lambda_1\lambda_2)n/4 & -\lambda_2 n/4 \\ \lambda_2[(n/4)-\Sigma x_i x_{i-k}] & 0 & ... & ... & (1+\lambda_1^2+\lambda_2^2)n/4 & (-\lambda_1+\lambda_1\lambda_2)n/4 \\ \lambda_1[(n/4)-\Sigma x_i x_{i-k}]+\lambda_1\lambda_2(\Sigma x_i x_{i-k-1}-\frac{n}{4}) & \lambda_2[(n/4)-\Sigma x_i x_{i-k}] & \vec{0}'_{k-5} & -\lambda_2 n/4 & (-\lambda_1+\lambda_1\lambda_2)n/4 & (1+\lambda_1^2+\lambda_2^2)n/4 \end{bmatrix}$$

Table 2.6: Optimal b_1 and b_2 and Efficiencies of Selected Design under AR(2) Model (run size n=240, $\rho_1 = 0.3$ and $\rho_2 = 0.2$)

k	b_1	b_2	Extended Orthogonal designs	Genetic Algorithm designs
9	0.2377	−0.3436	0.998714513	**
10	0.1673	−0.2446	0.999434854	**
11	0.1153	−0.1692	0.999760697	**
12	0.0791	−0.1166	0.999890897	**
13	0.0537	−0.0793	0.999959894	**
14	0.0362	−0.0537	0.99998145	**
15	0.0243	−0.0362	0.999988496	**
16	0.0162	−0.0241	0.999994623	0.990245426
17	0.0107	−0.016	1	0.9921242
18	0.0071	−0.0106	1	1.002165974
19	0.0047	−0.007	1	0.991557547
20	0.0031	−0.0045	1	1.017984105

Chapter 3

CONCLUSION

Under AR(1) model, orthogonal designs are not necessarily optimal for arbitrary correlation coefficient $\lambda_1 \in (-1, 1)$. Designs of CAOA can be more statistically efficient than orthogonal design. Under AR(2), it is hard to determine a minimum trace as in AR(1) model, but we can see from the typical case listed above, extended orthogonal designs are very efficient for common k's used in practice. Non-orthogonal designs generated by genetic algorithm are also very efficient. Non-orthogonal designs can be better than orthogonal designs under AR model. It is no surprise to see that extended M-sequence design, CAOA design and designs from genetic algorithm which perform well under uncorrelated error model, retain good performance under autoregressive model.

This paper only discussed designs with one type of stimuli. More types of stimuli should be considered in the future. Further discussion on AR model with three or more correlation coefficients might not be practical.

REFERENCES

Cheng, C., M.-H. Kao and F. Phoa, "Optimal and efficient designs for functional brain imaging experiments", Journal of Statistical Planning and Inference **181**, 71–80 (2017).

Dale, A., "Optimal experimental design for event-related fmri", Human Brain Maping **8**, 109–114 (1999).

Friston, K., E. Zarahn, O. Josephs, R. N. Henson and A. Dale, "Stochastic designs in event-related fmri", Neurolmage **10**, 607–619 (1999).

Jansma, J. M., J. A. de Zwart, P. van Gelderen, J. H. Duyn, W. C. Drevets and M. L. Furey, "In vivo evaluation of the effect of stimulus distribution on fir statistical efficiency in event-related fmri", Journal of neuroscience methods **215(2)**, 190–195 (2013).

Kao, M.-H., "Multi-objective optimal experimental designs for ER-fMRI using matlab", Journal of Statistical Software **30**, p.1–13 (2009).

Kao, M.-H., "On the optimality of extended maximal length linear feedback shift register sequences", Statistics and Probability Letters **83(6)**, 1479–1483 (2013).

Lazar, N., "The statistical analysis of functional MRI data, statistics for biology and health", Statistics for Biology and Health, New York: Springer (2008).

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

Lin, Y.-L., F. K. H. Phoa and M.-H. Kao, "Optimal design of fMRI experiments using cirsulant (almost-) orthogonal arrays", (submitted) (2016).

Liu, T. T., "Efficiency, power, and entropy in event-related fmri with multiple trial types: Part ii: design of experiments", NeuroImage **21(1)**, 401–413 (2004).

Worsley, K. and K. Friston, "Analysis of fmri time-series revisitedagain", NeuroImage **2**, 173181 (1995).

APPENDIX A

PROOF OF MINIMUM OF TRACE UNDER AR(1) MODEL

The general form is the pattern we found when listing the trace of information matrix for orthogonal designs in cases k=3,4,...,20. They are the general form for $k \leq 20$. In practice, k will not be larger than 20. We make a conjecture that it is the general form for all integer $k \geq 3$. We assume that this form is true for any k. To simplify notation, denote the correlation coefficient as a, and b_1 as b. The trace of C^{-1} will be in the form of $\frac{num}{den}$, where $num = c_1 - c_2 b^2$ and $den = c_3 + 2c_4 b - c_5 b^2$. This denotes that $c_1 = \Sigma_{i=1}^{k}(a^{2(k-i)})(i(k-i+1))$, $c_2 = \Sigma_{i=1}^{k}(a^{2(k-i)})\Big((i-1))(k-i)\Big)$, $c_3 = \Sigma_{i=0}^{k} a^{2i}$, $c_4 = a^k$, $c_5 = \Sigma_{i=0}^{k-1} a^{2i}$.

Differentiating this expression with respect to b and setting it to be zero gives two roots: $\frac{d_1 \pm \sqrt{d_2}}{2d_3}$ where $d_1 = c_1 c_5 - c_2 c_3$, $d_2 = d_1^2 - 4c_1 c_2 c_4^2$, and $d_3 = c_2 \times c_4$. Since $d_2 \leq d_1^2$ and consequently $d_1 - \sqrt{d_2} \geq 0$, both roots are greater than zero. However, both roots exist if and only if $d_2 \geq 0$. Denote the smaller root as r_1 and the other root r_2. The first step to prove $d_2 \geq 0$ is to find the general pattern of d_2. After a lot of trial and error, we found a general form of d_2 . d_2 can be expressed as a product of two non-negative terms, t_1 and t_2 where
$t_1 = 2\,a^4 \left(\sum_{i=1}^{k-2}\left(\sum_{j=1}^{i} j\,(j+1)\,a^{2\,i-2}\right) + \sum_{i=1}^{k-1}\left(\sum_{j=1}^{i} j\,(j+1)\,a^{4\,k-6-2\,i}\right)\right)$, and
$t_2 = \sum_{i=1}^{k} i a^{2\,i-2} + \sum_{i=1}^{k-1} i a^{4\,k-2-2\,i}$, i.e.

$$d_2 = \left(2\,a^4 \left(\sum_{i=1}^{k-2}\left(\sum_{j=1}^{i} j\,(j+1)\,a^{2\,i-2}\right) + \sum_{i=1}^{k-1}\left(\sum_{j=1}^{i} j\,(j+1)\,a^{4\,k-6-2\,i}\right)\right)\right)\times \quad \text{(A.1)}$$

$$\left(\sum_{i=1}^{k} i a^{2\,i-2} + \sum_{i=1}^{k-1} i a^{4\,k-2-2\,i}\right) \quad \text{(A.2)}$$

Following the same way as finding d_2, the general form of d_1 is:

$$d_1 = \Sigma_{i=1}^{k-1}\Big((i(i+1))a^{4k-2-2i} + (i(i+1))a^{2i}\Big) \quad \text{(A.3)}$$

Obviously, t_1 and t_2 are summations of some non-negative terms. Under the help of Maple, we verified that $t_1 * t_2 = d_1^2 - 4c_1 c_2 c_4^2$. As indicated above, $d_2 = t_1 * t_2$, t_1, t_2 ≥ 0, hence, $d_2 \geq 0$. Actually, under AR(1) model, $a \neq 0$, t_1 and t_2 are both positive. Therefore, $d_2 > 0$. Now we can draw the conclusion that both roots exist and are positive.

The next step is to prove that the trace at the smaller root r_1 is the global minimum. The derivative of the trace with respective to b, equals $\frac{-(b^2 c_1 c_4 - b c_1 c_5 + b c_2 c_3 + c_1 c_4)}{(b^2 c_5 - 2b c_4 - c_3)^2}$. Observe that the denominator is greater than zero and the numerator is a polynomial of b of degree two. From -1 ($b \in [-1, 1]$) to r_1, the derivative is negative, and becomes positive between the range of r_1 and r_2. In other words, the trace decreases from -1 to r_1 and increases between r_1 and r_2. Hence, the trace reaches local minimum at r_1. On the other hand, r_1 is smaller than 1 for $k = 3, 4, ..., 20$ The other possible minimum occurs at the boundary $b = 1$.

Now consider r_2 is a function of a. r_2 is always larger than 1 in the range of $a \in (0, 1)$. This is because the numerator of r_2 has terms of degree 2 and the denominator $2d_3$ is of degree $3k - 4$. $2d_3 = c_2 \times c_4$, does not have constant terms. Thus, r_2 will rise greatly near zero and reach its minimum near 1 because r_2 is a function of a of negative powers. All the traces at $b = 1$, for $k = 3, 4, ..., 20$ are checked, and they are all larger than 1. On the other hand, all the plots of r_1 versus a are checked, the maximum of r_1's are smaller than 1. Therefore, 1 is always between r_1 and r_2. Hence, the trace obtained at $b = 1$ is larger than trace at r_1. We draw the conclusion that trace has global minimum at r_1 for any $a \in (-1, 1)$.