Optimal Experimental Designs for Event-Related Functional Magnetic
Resonance Imaging

by

Ming-Hung Kao

(Under the direction of John Stufken and Abhyuday Mandal)

Abstract

The main foci of this dissertation are 1) developing efficient computational approaches for finding optimal experimental designs for event-related functional magnetic resonance imaging (ER-fMRI) and 2) studying the characteristics of optimal ER-fMRI designs obtained. Taking into account both statistical efficiencies and practical constraints, we develop an approach that includes rigorously formulated models, well defined design criteria, and an efficient, versatile search algorithm. Our algorithm incorporates knowledge about the performance of well known ER-fMRI designs, and can accommodate a variety of experimental settings. Through simulations, we show that our approach is more efficient than other methods. Designs found by our approach outperform other designs currently in use by fMRI researchers.

Under the popular linear model framework, we adopt our approach to study cases where both individual stimulus effects and pairwise contrasts of stimulus types are of interest; these two effects are common interests of fMRI researchers. A practical situation where a long scanning session is divided into multiple short scanning sessions is also investigated. We also take into account the warm-up period of an MR scanner when finding optimal ER-fMRI designs. These studies indicate that our approach can work reliably well for a variety of practical situations encountered in ER-fMRI experiments.

A nonlinear model is also considered in our study. While previous studies use two linear models for the two common statistical objectives, namely estimating the hemodynamic response function and detecting brain activation, the nonlinear model approach that we propose provides a natural, unified setting for these two objectives. In addition to finding locally optimal designs and pseudo-Bayesian designs, we also adopt techniques for solving multi-objective optimization problems to obtain a set of designs for researchers to choose from based on their goals and needs.

INDEX WORDS: Boltzmann transformation, compound design criterion, design efficiency, individual stimulus effect, genetic algorithms, multi-objective optimization, NSGA-II, pairwise contrasts, pseudo-Bayesian design

OPTIMAL EXPERIMENTAL DESIGNS FOR EVENT-RELATED FUNCTIONAL MAGNETIC

RESONANCE IMAGING

by

MING-HUNG KAO

B.S., National Central University, Taiwan, 1997

M.S., National Central University, Taiwan, 1999

M.S., The University of Georgia, 2006

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2009

Optimal Experimental Designs for Event-Related Functional Magnetic
Resonance Imaging

by

Ming-Hung Kao

Approved:

Major Professors: John Stufken
                    Abhyuday Mandal

Committee: Nicole Lazar
                    Dibyen Majumdar
                    Lynne Seymour

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2009

DEDICATION

To my parents and Yi-Chun

TABLE OF CONTENTS

List of Tables

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Event-related functional magnetic resonance imaging (ER-fMRI) is one of the leading techniques for studying brain activity in response to mental stimuli (e.g. looking at pictures or tapping fingers). It is an important advance in neuroscience (Culham, 2006), and is a popular brain mapping technique (Josephs and Henson, 1999; Rosen et al., 1998). The design for ER-fMRI can be viewed as a sequence of finite numbers; each number indicates a stimulus or control (rest or fixation). The elements, which can be stimuli or control, of a design sequence are referred to as events, and are presented sequentially to an experimental subject while an MR scanner scans the subject's brain every TR (time to repetition) seconds to collect fMRI time series. One design issue is to find the best design sequence to help collect informative data for making inference about the subject's brain activity. Our main focus is on this important design issue.

Due to the popularity and high cost of ER-fMRI experiments, optimal experimental designs that help render precise and valid statistical inference are crucial. However, finding these designs is an arduous task. One reason is that the design space consisting of all possible sequences is enormous. It can have $2^{255} \cong 5.8 \times 10^{76}$ designs for an experiment involving only one stimulus type (i.e., two event types) and design length of 255. Experiments with larger scales are also not uncommon and the corresponding design spaces are even larger. To search over such huge spaces for good designs, we need an efficient search algorithm. In addition, the experimental setting varies with the experimenter's interests. We would like a versatile search algorithm that can accommodate various experimental conditions and help experimenters find designs best suited to their needs.

Moreover, an ER-fMRI experiment typically involves conflicting objectives. These objectives include statistical issues and practical considerations. Two common statistical objectives are estimation and detection (e.g. Birn et al., 2002). Estimation refers to the estimation of the hemodynamic response function (or HRF), a function of time describing an effect of a single, brief stimulus. An HRF might look like the curve in Figure 1.1. The alteration of this curve is linked to the change in the ratio of oxygenated to deoxygenated hemoglobin (e.g. Lazar, 2008; Cabeza and Kingstone, 2006); it typically starts to increase in about one or two seconds following a stimulus onset, reaches the peak in about five to seven seconds and then falls back to be below baseline before completely returning to baseline. The shape of the HRF might vary with brain voxels (equal sized volume elements of the brain), experimental subjects, and scanning sessions; see, e.g., Menz et al. (2006) and Handwerker et al. (2004). It is argued that studying the HRF may help us to understand the underlying neuronal activity (Lindquist et al., 2009; Henson et al., 2002).
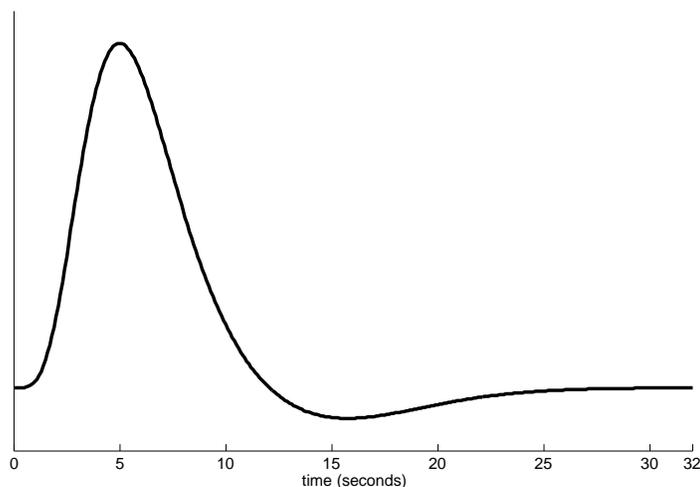
Figure 1.1: The hemodynamic response function

The goal of detection is to identify brain voxels that are activated by the stimuli. This objective is intrinsically linked to the estimation of the HRF (Makni et al., 2008), and

considering both objectives in one study is not uncommon. However, there is a trade-off relationship between these two statistical objectives (e.g. Liu et al., 2001). Good multi-objective designs achieving advantageous compromises between the two dimensions are thus called for; see also Wager and Nichols (2003) and references therein.

In addition to statistical efficiency, we also need to take into account psychological constraints. When a design sequence is patterned or easy to predict, psychological effects such as habituation or anticipation can occur to contaminate the data (Liu et al., 2001; Dale, 1999). A good design should help avoid these confounds while retaining high statistical efficiencies. To achieve such designs, we propose an approach that can accommodate a various experimental settings, take into account the competing objectives and search efficiently over the huge design space for such good designs. Our approach is built upon existing knowledge of the performance of well known ER-fMRI designs studied in previous researches.

Previous studies on ER-fMRI designs (e.g. Liu and Frank, 2004; Liu, 2004; Wager and Nichols, 2003) mainly focus on the popular linear model framework (Friston et al., 1995a; Worsley and Friston, 1995; Friston et al., 1995). Under this framework, two linear models are considered, one for estimation, and the other for detection. The model for estimation uses a finite number of parameters to represent the values of the HRF at equally spaced time points following a stimulus onset. The HRF is estimated through estimating these parameters. On the other hand, the model for detection assumes that the HRFs are formed by a basis function (e.g., the canonical HRF of SPM, http://www.fil.ion.ucl.ac.uk/spm) with unknown amplitudes. The aim of detection problems lies at studying the amplitude; the larger the amplitude, the more the voxel is activated by the stimuli.

Design criteria are developed with respect to these two models, and a common goal is to obtain designs achieving advantageous compromises between these two competing statistical objectives while avoiding psychological constraints. Liu (2004) finds designs having intermediate efficiencies between the two statistical objectives and evaluates the ability of designs to circumvent psychological constraints. Wager and Nichols (2003) define multi-objective

design criteria that involve both statistical and psychological considerations. They propose a genetic algorithm (GA) to search for designs optimizing their multi-objective design criteria.

Building upon these pioneering studies, we propose an efficient approach for finding optimal multi-objective ER-fMRI designs. Our approach involves rigorously formulated models, well defined design criteria and an efficient, knowledge-based genetic algorithm for searching for optimal designs. In our model formulation, we follow Dale (1999) to discretize the HRF via a given discretization interval, which facilitates the use of a finite number of HRF parameters in representing the continuous HRF. In contrast to previous studies, we consider both the TR and ISI (inter-stimulus interval; a given, fixed time interval between consecutive event onsets in the design sequence) when specifying the discretization interval; the HRF parameters in our models become interpretable.

In addition, the design criteria that we define provide clear targets for our search algorithm. By contrast, values of the design criteria of Wager and Nichols (2003) depend largely on the composition of each generation of their GA. Since the composition changes over GA generations, their criteria are moving targets. Furthermore, our search algorithm incorporates knowledge about ER-fMRI designs and is demonstrated, in Chapter 2, to be more efficient than other methods.

To help researchers easily find designs suited for their goals and needs, we develop a computer program using MATLAB that implements our proposed approach. This program allows user-specified experimental settings and is described in Chapter 3.

We also conduct studies on practical design issues that can be encountered in ER-fMRI experiments. Chapter 4 investigates the case where both individual stimulus effects and pairwise contrasts between stimulus types are considered in one experiment; these two effects are of great interest to ER-fMRI researchers (e.g. Amaro and Barker, 2006; Donaldson and Buckner, 2002). We search for optimal designs and observe that designs efficient for individual stimulus effects are more robust with respect to a change in interest than designs for pairwise contrasts. In addition, the frequency of the stimuli in the designs that we obtain increases

when more weights are assigned to pairwise contrasts. On the other hand, the frequency of the control increases when researchers are more interested in individual stimulus effects. The frequencies of stimuli of our designs concur with the approximated optimal stimulus frequencies derived by Liu and Frank (2004).

In Chapter 5, a study on designs for multiple scanning sessions is conducted. Dividing a long scanning session into several short ones is a widely used strategy, especially when the subject cannot maintain her/his attention throughout a long session. To find optimal designs suited for this situation, we consider models that take into account the session effect. Design criteria are defined and we compare six approaches for finding designs optimizing the design criteria. The first approach finds an efficient short design for the first scanning session, and uses this short design for other sessions, too. The second approach finds different efficient short designs for different sessions. The third approach obtains an efficient short design and then cyclically permutes stimulus types in this design to create designs for subsequent sessions. Our simulations show that the third approach performs better than the first two. In addition, we also consider approaches that find efficient long designs for the entire experiment; these designs are divided into short ones to form short designs for different sessions. The forth approach searches for optimal long designs under models without session effects, and the fifth approach finds long designs for models with session effects. Comparing these two approaches reveals that failing to take into account session effects can result in inefficient designs. Furthermore, we also search for efficient long design over a restricted design space. Each design in this smaller space is formed by juxtaposing a short design and its cyclically permuted designs. This last approach is different from the third approach. The third approach uses a short design, that is efficient for one scanning session, to create a long design which is not necessarily efficient for the entire experiment. On the other hand, the sixth approach finds an efficient long design formed by a short design that may be inefficient when only one scanning session is considered. Our simulation results indicate that the fifth and sixth approaches outperform the others.

In Chapter 6, we investigate the performance of several hybrid algorithms, that are developed by combining concepts of simulated annealing and GAs, for finding optimal designs for studies with a warm-up period of an MR machine. While data obtained during this period are usually discarded, the effects of stimuli that are present in this period can still be carried over to the following scans. We therefore take into account such carry-over effects in our model and search for optimal designs using six different algorithms. The first algorithm is the knowledge-based GA proposed in Chapter 2, and the other algorithms are its variants. To develop the variant algorithms, we consider two modifications. One modification involves the way of selecting designs into a mating pool for offspring reproduction. The other modification pertains to the way of selecting designs to survive to the next GA generation.

In the knowledge-based GA, designs are selected into a mating pool with probabilities proportional to fitness (value of the design criterion). In addition to this approach, we consider selecting designs based on Boltzmann transformed fitness. This transformation is indexed by a parameter called temperature, and allows a variety of selective pressure which is the ratio of the probability of selecting the best design to the average selection probability of all designs in the population. A high temperature results in a weak selective pressure, and the selection is close to a random selection. On the other hand, a low temperature corresponds to a strong selective pressure, and the GA search finds good designs quickly but with a higher risk of being prematurely converged. We consider two schedules to change the temperature during a GA search; the cooling schedule gradually decreases the temperature, whereas the warming schedule alters the temperature in the reverse direction.

As for selecting designs to survive to the next generation, the knowledge-based GA is elitist-based, and offspring designs compete with their parents for survival. Another possibility is to consider age-based algorithms where designs of the previous generation are removed from the population; the next generation is formed only by offspring designs. Our simulation results show that the elitist-based algorithms outperform the age-based ones. In addition, elitist-based algorithms with Boltzmann transformed fitness do not perform better

than the the original knowledge-base GA (the elitist-base algorithm without Boltzmann transformation). We therefore advocate the use of the original knowledge-based GA.

When conducting the studies described previously, we follow the line of research in which estimation and detection are tackled using different linear models. The "dual model" approaches stem from the seemingly contradictory results of Dale (1999) and Friston et al. (1999). Dale (1999) suggests to use designs with stimuli rapidly alternating among types; these designs are random designs. On the other hand, Friston et al. (1999) show that block designs, where stimuli of the same type are clustered, perform the best among the designs of their investigation which include random designs. Buxton et al. (2000) argue that the two problems need to be distinguished. Dale (1999) considers estimation, whereas Friston et al. (1999) focus on detection. Two different models are considered for these two problems and they lead to different optimal designs.

The study of Buxton et al. (2000) demonstrates that random designs have high efficiencies in estimating the HRF, but these designs are not as efficient as block designs in detection problems. On the other hand, block designs does not perform well for estimation. A combination of these two types of designs can have an intermediate efficiency between the two competing goals. Liu et al. (2001) and Liu and Frank (2004) further investigate these two objectives and conclude a trade-off relationship. When both objectives are of interest, the dual model approaches are considered for finding designs achieving advantageous compromises between the two dimensions; see also Wager and Nichols (2003), and Liu (2004).

However, the dual model approaches have a limitation. We need to assume the shape of the HRF in the model for detection. This might be difficult since researchers usually do not have full knowledge about the HRF; as described previously, the HRF may change with brain voxels, subjects and sessions. To address this issue, we consider a nonlinear model that provides a unified setting for both estimation and detection. The nonlinear model that we consider is a natural extension of the popular linear models, and is highly related to the one used by Wager et al. (2005). It involves free parameters describing temporal attributes

and amplitudes of the HRF. When focusing on detection or the amplitude parameter, the uncertainty in the HRF's temporal attributes is also taken into account.

Finding optimal designs under nonlinear models is challenging because the result depends on unknown parameters. One approach is to search for locally optimal designs, which are optimal at guessed parameter values (Chernoff, 1953). However, locally optimal designs may be inefficient when guessed parameter values are far away from the truth. To take into account the uncertainty in the parameter values, pseudo-Bayesian designs (Chaloner and Larntz, 1989) which are obtained under a pre-specified distribution of the parameters are also considered in our study. In addition, we propose a novel approach for design selection. Adopted from techniques of multi-objective optimization problems, our proposed approach finds a class of designs for researchers to choose from based on their goals and needs. This study on designs for a nonlinear model is described in Chapter 7.

Conclusions of our studies is provided in Chapter 8 along with discussions of future researches.

1.1 REFERENCES

Amaro, E. J. and Barker, G. J. (2006), "Study Design in MRI: Basic Principles," *Brain and Cognition*, 60, 220–232.

Birn, R. M., Cox, R. W., and Bandettini, P. A. (2002), "Detection Versus Estimation in Event-Related fMRI: Choosing the Optimal Stimulus Timing," *NeuroImage*, 15, 252–264.

Buxton, R. B., Liu, T. T., Martinez, A., Frank, L. R., Luh, W. M., and Wong, E. C. (2000), "Sorting out Event-Related Paradigms in fMRI: The Distinction between Detecting an Activation and Estimating the Hemodynamic Response," *NeuroImage*, 11, S457.

Cabeza, R. and Kingstone, A. (2006), *Handbook of Functional Neuroimaging of Cognition*, Cognitive Neuroscience, Cambridge, Mass.: MIT Press, 2nd ed.

Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Planning and Inference*, 21, 191–208.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 24, 586–602.

Culham, J. C. (2006), "Functional Neuroimaging: Experimental Design and Analysis," in *Handbook of Functional Neuroimaging of Cognition*, eds. Cabeza, R. and Kingstone, A., Cambridge, Massachusetts: MIT Press, 2nd ed., pp. 53–82.

Dale, A. M. (1999), "Optimal Experimental Design for Event-Related fMRI," *Humman Brain Mapping*, 8, 109–114.

Donaldson, D. I. and Buckner, R. L. (2002), "Effective Paradigm Design," in *Functional MRI: An Introdcution to Methods*, eds. Jezzard, P., Matthews, P. M., and Smith, S. M., New York: Oxford University Press, pp. 177–195.

Friston, K. J., Frith, C. D., Turner, R., and Frackowiak, R. S. J. (1995a), "Characterizing Evoked Hemodynamics with fMRI," *NeuroImage*, 2, 157–165.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. (1995b), "Analysis of fMRI Time-Series Revisited," *NeuroImage*, 2, 45–53.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., and Dale, A. M. (1999), "Stochastic Designs in Event-Related fMRI," *NeuroImage*, 10, 607–619.

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004), "Variation of BOLD Hemodynamic Responses across Subjects and Brain Regions and Their Effects on Statistical Analyses," *Neuroimage*, 21, 1639–1651.

Henson, R. N. A., Price, C. J., Rugg, M. D., Turner, R., and Friston, K. J. (2002), "Detecting Latency Differences in Event-Related BOLD Responses: Application to Words Versus Nonwords and Initial Versus Repeated Face Presentations," *Neuroimage*, 15, 83–97.

Josephs, O. and Henson, R. N. A. (1999), "Event-Related Functional Magnetic Resonance Imaging: Modelling, Inference and Optimization," *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 354, 1215–1228.

Lazar, N. A. (2008), *The Statistical Analysis of Functional MRI Data*, Statistics for Biology and Health, New York: Springer.

Lindquist, M. A., Loh, J. M., Atlas, L. Y., and Wager, T. D. (2009), "Modeling the Hemodynamic Response Function in fMRI: Efficiency, Bias and Mis-Modeling," Academic Press Inc Elsevier Science, pp. S187–S198.

Liu, T. T. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part II: Design of Experiments," *NeuroImage*, 21, 401–413.

Liu, T. T. and Frank, L. R. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part I: Theory," *NeuroImage*, 21, 387–400.

Liu, T. T., Frank, L. R., Wong, E. C., and Buxton, R. B. (2001), "Detection Power, Estimation Efficiency, and Predictability in Event-Related fMRI," *NeuroImage*, 13, 759–773.

Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., and Ciuciu, P. (2008), "A Fully Bayesian Approach to the Parcel-Based Detection-Estimation of Brain Activity in fMRI," *NeuroImage*, 41, 941–969.

Menz, M. M., Neumann, J., Muller, K., and Zysset, S. (2006), "Variability of the BOLD Response over Time: An Examination of within-Session Differences," *NeuroImage*, 32, 1185–1194.

Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998), "Event-Related Functional MRI: Past, Present, and Future," *PNAS*, 95, 773–780.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of Experimental Design in fMRI: A General Framework Using a Genetic Algorithm," *NeuroImage*, 18, 293–309.

Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. (2005), "Accounting for Nonlinear BOLD Effects in fMRI: Parameter Estimates and a Model for Prediction in Rapid Event-Related Studies," *NeuroImage*, 25, 206–218.

Worsley, K. J. and Friston, K. J. (1995), "Analysis of fMRI Time-Series Revisited–Again," *NeuroImage*, 2, 173–181.

CHAPTER 2

MULTI-OBJECTIVE OPTIMAL EXPERIMENTAL DESIGNS FOR EVENT-RELATED FMRI

STUDIES[1]

2.1 Introduction

ER-fMRI is one of the leading technologies for studying human brain activity in response to mental stimuli (Josephs et al., 1997; Rosen et al., 1998; Dale, 1999; Bandettini and Cox, 2000). Before conducting an ER-fMRI experiment, a design sequence consisting of stimuli of one or more types interlaced with rests is prepared. This sequence is presented to an experimental subject, while the MR scanner measures changes in the subject's blood oxygenation level dependent (BOLD) response for the end purpose of statistical inference. The design issue here is to best allocate the stimuli so that inference is precise and valid.

Two common statistical goals in ER-fMRI are to estimate the HRF (the noise-free BOLD time series triggered by a single, brief stimulus), and to detect brain activation; see also Buxton et al. (2000) and Birn et al. (2002). Considering both goals in one experiment is not uncommon, but it requires a good multi-objective design that simultaneously achieves high efficiencies on both dimensions. However, statistical efficiency is not the only concern for planning ER-fMRI design sequences. Psychology plays an important, even crucial, role. When a design sequence is patterned or easy to predict, psychological effects such as habituation or anticipation may occur to confound stimulus effects (Dale, 1999). Therefore, a good design should provide safeguards against the psychological confounds while retaining a high efficiency for statistical inference. Moreover, customized requirements such as a required frequency for each stimulus type might also arise to further complicate the design problem. As a consequence, the search for a good, multi-objective design is inevitable and a well-defined multi-objective design criterion (or MO-criterion for short) is needed to evaluate competing designs. In addition, the design space, consisting of all possible ER-fMRI designs, is enormous and irregular (Buračas and Boynton, 2002; Liu, 2004). Searching over this huge space for an optimal design is an arduous task, thus an efficient search algorithm is as well crucial.

Wager and Nichols (2003), referred to as WN henceforward, propose a framework for finding multi-objective optimal ER-fMRI designs. They formulate the MO-criterion as a weighted average of the design criteria for the individual objectives of interest. A modified

genetic algorithm (or WN's GA) is also introduced to search for optimal or near-optimal multi-objective designs. This trailblazing work has been applied in many studies over the last few years (e.g., Callan et al., 2006; Ramautar et al., 2006; Summerfield et al., 2006; Wang et al., 2007).

Inspired by WN's pioneering work, we develop an efficient approach to search for optimal multi-objective designs. Our approach has two major advantages. First, we incorporate well-known fMRI designs in our algorithm to facilitate the search. Second, we define a family of MO-criteria that allows consistent design comparisons. While crucial to the success of a search algorithm, WN's criteria do not always achieve this. Furthermore, our algorithm is simple and easy to implement, yet effective.

The efficiency and effectiveness of our approach are demonstrated through simulations under two popular cases, one focuses on individual stimulus effects and the other on pairwise contrasts. We also discuss the case when both cases are simultaneously of interest. While taking less computation time than WN's approach, our algorithm achieves designs with significantly higher efficiencies. We also demonstrate that our designs form an advantageous trade-off between estimation efficiency and detection power, and we find designs yielding higher estimation efficiencies than m-sequences. Moreover, under the model with white noise and a constant nuisance parameter, the stimulus frequencies of the designs we obtained are in good agreement with the optimal stimulus frequencies derived by Liu and Frank (2004).

In this chapter, our proposed algorithm is introduced, and its performance is demonstrated via simulations. Other details and additional simulations are presented in Kao et al. (2007). The rest of the chapter is organized as follows. Section 2.2 presents our proposed approach. Simulations are provided in Section 2.3. Conclusions and a discussion are in Section 2.4.

2.2 Methodology

We propose an efficient and effective approach to search for optimal multi-objective designs for ER-fMRI. Four objectives are considered: 1) estimating the HRF, 2) detecting brain activation, 3) avoiding psychological confounds, and 4) maintaining the desired stimulus frequency in the design sequence. By assigning weights to these objectives based on the researcher's discretion, our algorithm finds a design best suited to the researcher's needs. We briefly introduce our approach in this section. The approach is general enough that other objectives, beside the four listed above, could be accommodated as well.

2.2.1 Underlying Model and Design Criteria

To find an optimal design, we need to specify the underlying model for the two primary statistical objectives, namely estimation and detection. As in WN and Liu and Frank (2004), two popular linear models are considered (Friston et al., 1995; Worsley and Friston, 1995; Dale, 1999):

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{e}, \text{ and} \tag{2.1}$$

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{2.2}$$

where $\boldsymbol{Y}$ is the voxel-wise BOLD time series, $\boldsymbol{h} = (\boldsymbol{h}_1', ..., \boldsymbol{h}_Q')'$ is the parameter vector for the HRFs of the Q stimulus types, $\boldsymbol{X} = [\boldsymbol{X}_1 \cdots \boldsymbol{X}_Q]$ is the design matrix, $\boldsymbol{\theta} = (\theta_1, ..., \theta_Q)'$ represents the response amplitudes, $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{h}_0$ is the convolution of stimuli with an assumed basis, $\boldsymbol{h}_0$, of the HRF, $\boldsymbol{S}\boldsymbol{\gamma}$ is a nuisance term describing the trend or drift of $\boldsymbol{Y}$, and $\boldsymbol{e}$ and $\boldsymbol{\eta}$ are noise. Following WN, we assume a known whitening matrix, $\boldsymbol{V}$, such that $\boldsymbol{V}\boldsymbol{e}$ and $\boldsymbol{V}\boldsymbol{\eta}$ are white noise.

Model (2.1) is typically used for estimating the HRF and model (2.2) for detecting activation. Under these models, the A- or D-optimal design criteria can be applied to evaluate competing designs with respect to the objectives of estimation and detection. Both of these

criteria are widely accepted and the choice between A- and D-optimality depends on individual preference. A-optimality aims at minimizing the average variance of estimators of parametric functions. In our simulations, these will be individual stimulus effects, or pairwise contrasts. On the other hand, a D-optimal design minimizes the generalized variance of estimators of linearly independent parametric functions, or, under normality, it minimizes the volume of simultaneous elliptical confidence regions for these parametric functions at any specified confidence level. For our problem, these parametric functions will be either individual stimulus effects, or $(Q - 1)$ linearly independent pairwise contrasts. For further details, see Atkinson et al. (2007).

For technical reasons, we formulate these design criteria as "larger-the-better" criteria, and designs maximizing them help to optimize statistically meaningful functions of the parameter estimators as previously described. The value of the design criterion for estimation, referred to as "estimation efficiency", is denoted by F_e. Likewise, the term "detection power" and the notation F_d are used to indicate the value of the design criterion for detection. These two criteria are defined to have one of the following two forms:

$$
F_i = \begin{cases} r_c/trace(\boldsymbol{M}), & \text{for } A\text{-optimality}; \\ det(\boldsymbol{M})^{-1/r_c}, & \text{for } D\text{-optimality}, \end{cases} \tag{2.3}
$$

where $\boldsymbol{M} = \boldsymbol{C}[\boldsymbol{W}'\boldsymbol{V}'(\boldsymbol{I} - P_{\boldsymbol{Vs}})\boldsymbol{VW}]^{-}\boldsymbol{C}'$, $\boldsymbol{W} \equiv \boldsymbol{X}$ for F_e, $\boldsymbol{W} \equiv \boldsymbol{Z}$ for F_d, $\boldsymbol{I}$ is an identity matrix, $P_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'$ is the orthogonal projection on the vector space spanned by the column vectors of $\boldsymbol{A}$, $\boldsymbol{A}^{-}$ is a generalized inverse matrix of $\boldsymbol{A}$, $\boldsymbol{C}$ is a matrix of linear combinations of the parameters, and r_c is the number of rows of $\boldsymbol{C}$.

The third objective is to avoid psychological confounds. We would like a sequence that makes it difficult for a subject to anticipate future stimuli based on past stimuli. To achieve this, the Rth order counterbalancing property of WN is considered, where R is a given integer. This property is defined on a sub-design of the original design obtained by keeping only the stimuli but deleting all rests. For any $r \in \{1, ..., R\}$, we count the pairs of stimuli that appear in positions $(t, t + r)$ in the sub-design, $t = 1, .., (n - r)$; n is the length of the

sub-design. The Rth order counterbalancing aims at having each pair appear a number of times that is proportional to the product of the specified proportions for the stimuli. The corresponding design criterion can be written as:

$$F_c = \sum_{r=1}^{R}\sum_{i=1}^{Q}\sum_{j=1}^{Q}\lfloor|n_{ij}^{(r)} - (n-r)P_iP_j|\rfloor,$$

where $n_{ij}^{(r)}$ is the number of occurrences of a type-i stimulus being the tth element and a type-j stimulus being the $(t+r)$th element, $t = 1, ..., (n-r)$, P_i is the specified proportion for the type-i stimulus in the sub-design which may be taken as $1/Q$ if there is no preference, and $\lfloor|a|\rfloor$ is the integer part of the absolute value of a. This criterion measures the departure from counterbalancing and is a "smaller-the-better" criterion.

The fourth design criterion is also defined on the sub-design. It is $F_f = \sum_{i=1}^{Q}\lfloor|n_i - nP_i|\rfloor$, where n_i is the number of the type-i stimulus in the sub-design. This criterion helps to maintain the desired stimulus frequency and is a "smaller-the-better" criterion.

An MO-criterion is defined as a convex combination of the above four individual criteria. To ensure comparability, they are standardized before combining. We use the following standardization:

$$F_i^* = \begin{cases} \frac{F_i - min(F_i)}{max(F_i) - min(F_i)}, & i = d, e; \\ 1 - \frac{F_i - min(F_i)}{max(F_i) - min(F_i)}, & i = c, f. \end{cases}$$

Our family of MO-criteria is then defined as $\{F^* = w_cF_c^* + w_dF_d^* + w_eF_e^* + w_fF_f^* : w_i \geq 0, i = c, d, e, f; \sum_i w_i = 1\}$; w_is are weights selected based on the researcher's emphasis in a given study.

By contrast, WN standardize each F_i by its mean and standard deviation over designs within the current generation of their GA. Since designs change with successive generations, so do these means and standard deviations. The resulting MO-criteria are moving targets during the search. Thus, fair, consistent design comparisons may not be achieved. Our MO-criteria are free from this drawback.

With the MO-criterion for evaluating the "goodness" of competing designs, we propose a GA-based algorithm to search for the optimal ER-fMRI design.

2.2.2 SEARCH ALGORITHM

GAs (Holland, 1975; 1992) are popular for solving optimization problems, in which good solutions (parents) are used to generate better ones (offsprings). To efficiently apply this technique, we take advantage of well-known results about good fMRI designs so that our search over the huge design space can be carried out more efficiently. The outline of our algorithm is as follows:

Step 1. (Initial designs) Generate G initial designs consisting of random designs, an m-sequence-based design, a block design and their combinations. Use the objective function to evaluate the fitness of each initial design.

Step 2. (Crossover) With probability proportional to fitness, draw with replacement $G/2$ pairs of designs to crossover — select a random cut-point and exchange the corresponding fractions of "genetic material" in paired designs. See Wager and Nichols (2003) for a nice graphical presentation. Here, the "genetic material" is the design sequence.

Step 3. (Mutation) Randomly select $q\%$ of the events from the G offspring designs. Replace these events by randomly generated ones. Here, an event is a stimulus or a rest.

Step 4. (Immigration) Add to the population another I designs drawn from random designs, block designs and their combinations.

Step 5. (Fitness) Obtain the fitness scores of the offsprings and immigrants.

Step 6. (Natural selection) Keep the best G designs according to their fitness scores to form the parents of the next generation. Discard the others.

Step 7. (Stop) Repeat steps 2 through 6 until a stopping rule is met (e.g., after M_g generations). Keep track of the best design over generations.

We describe below some details of our GA. MATLAB code implementing this algorithm can be found at http://www.stat.uga.edu/∼amandal.

INITIAL DESIGNS AND IMMIGRANTS

In Step 1, m-sequence-based designs or m-sequences are generated following Liu (2004); see also Buračas and Boynton (2002). These designs are well-known for their high estimation efficiencies. Since they are not always available, concatenations or truncations of the existing ones are also considered. We include the one yielding the highest estimation efficiency as one of the initial designs.

The initial block design has the highest detection power among designs of differing numbers of blocks and of two different patterns. In this pool of candidate block designs, the number of blocks for each stimulus type ranges among one to five, 10, 15, 20, 25, 30, and 40. The two patterns include repetitions of NABC and NANBNC, where N is a block of rests and A, B and C represent blocks of stimuli of different types. In addition to the initial block design, immigrants in Step 4 ensure a steady supply of blocks of different sizes.

The combination of a block design with an m-sequence-based design or a random design is obtained through crossover. These mixed designs constitute a portion, e.g., one-third, of the initial designs. The remaining initial designs are formed by random designs.

OBJECTIVE FUNCTION

The objective function used in Step 1 and Step 5 of our GA evaluates the fitness or "goodness" of the designs. Based on the goal of the search, the objective function can be taken as a single F_i or as an MO-criterion with weights selected by the researcher's interest. Note that the extreme values of the F_is are required to use our MO-criteria.

Theoretical values of $\max(F_e)$ and $\max(F_d)$ are generally not available. They can be approximated by performing a "pre-run" of our GA using the non-standardized function F_e (or F_d) as the objective function. The values of $\min(F_e)$ and $\min(F_d)$ are set to zero, corresponding to designs for which the parameters of interest are non-estimable. Both $\min(F_c)$ and $\min(F_f)$ are zero. Their maximal values are attained by the design containing only the stimulus type with the smallest specified proportion P_i. With these extreme values and given

weights, an MO-criterion F^* is well-defined and serves as the objective function for finding optimal multi-objective designs.

2.3 SIMULATIONS

In the following illustrative simulations, we consider designs with three stimulus types ($Q = 3$) and $L = 255$ events. The ISI (inter-stimulus interval, time between consecutive event onsets) and the TR (time to repetition, or sampling rate) are both set to two seconds.

For F_e and F_d, we use the A-optimality criterion and consider two popular situations, namely individual stimulus effects and pairwise contrasts. For the former situation, the C matrix described after (2.3) is the identity matrix. For the latter case, the rows of C correspond to the $Q(Q-1)/2$ pairwise contrasts between stimulus types. The canonical HRF, a combination of two Gamma distributions (SPM2, http://www.fil.ion.ucl.ac.uk/spm), is used as h_0 in model (2.2). In the first two simulations, the drift, described by $S\gamma$, is assumed to be a second-order Legendre polynomial, and the noise follows a stationary AR(1) process with a correlation coefficient of 0.3. In the last simulation, white noise is assumed and S is taken to be a vector of ones. As for F_c and F_f, we require a third-order counterbalancing property ($R = 3$) and equal frequencies for the three stimulus types; i.e., $P_i = 1/3$, $i = 1, 2, 3$.

Unless otherwise specified, the algorithmic parameters are G (size of population) $= 20$, q (percentage of mutation) $= 1\%$, I (number of immigrants) $= 4$ and M_g (number of generations) $= 10,000$. A larger value of M_g does not seem to lead to significantly better designs. The simulations are performed on a Pentium Dual 3.20/3.19 GHz computer with 3.5 Gb of RAM.

SIMULATION 1

We first consider three weighting schemes, namely (A) $w_c = w_d = w_e = w_f = 0.25$, (B) $w_e = 1$, and (C) $w_d = 1$, with the C matrix being the identity matrix. The first weighting scheme finds a multi-objective design, whereas the latter two schemes search for the best

designs for estimation and detection, respectively. The achieved values of the design criterion over the 10,000 GA generations are presented in Figure 2.1. For weighting scheme (A), the value of the MO-criterion is presented. The estimation efficiency, F_e, and detection power, F_d, are reported for weighting schemes (B) and (C), respectively.
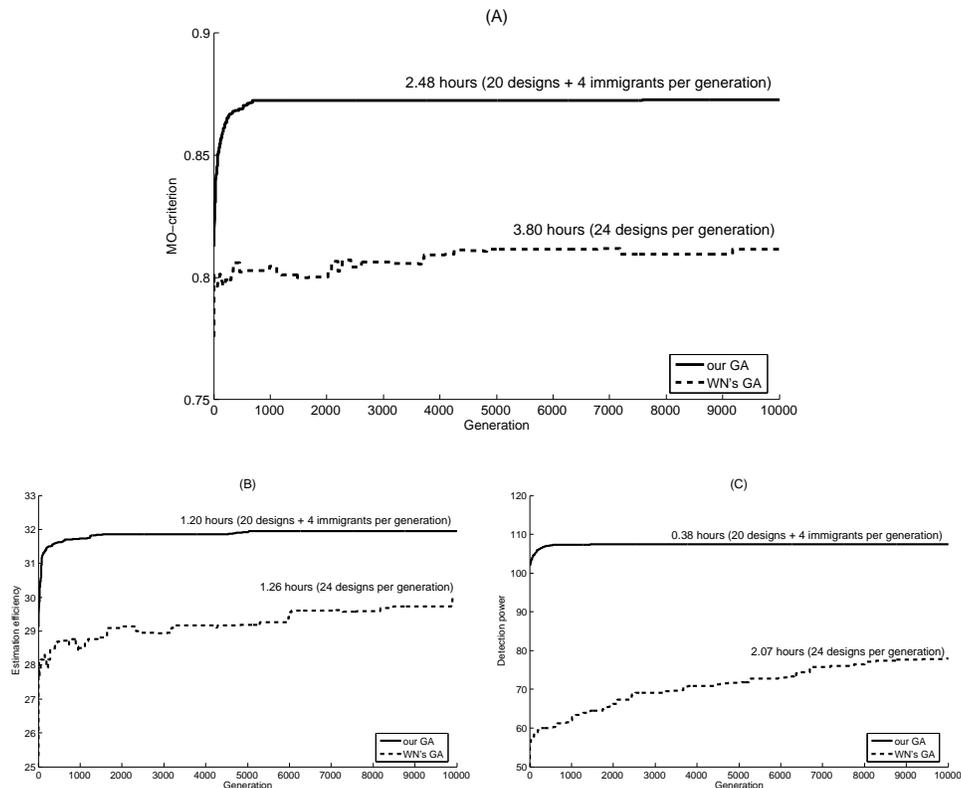
Figure 2.1: Achieved values of design efficiency vs. generation for (A) $w_c = w_d = w_e = w_f$, (B) $w_e = 1$, and (C) $w_d = 1$. CPU times for completing 10,000 generations are presented.

Our GA is compared to WN's GA. For comparison, we include 24 designs in each generation of their GA since they do not allow immigration. As shown in Figure 2.1 (A), our GA achieves a value of the MO-criterion of 0.873 while WN's GA attains 0.812. In addition, our algorithm uses less CPU time than their GA. Significant improvements made by our GA are also observed in Figure 2.1 for the other two weighting schemes. Note that the efficiency curve for the MO-criterion in WN's GA is not monotone, a result of the inconsistency of their normalization method that was pointed out in Subsection 2.2.

Under weighting scheme (B), our GA finds a design yielding a higher estimation efficiency than the m-sequence-based design. The estimation efficiency is 31.96 for our design compared to 29.12 for the m-sequence-based design. Our design, featuring small off-diagonal elements in the information matrix (not shown), possesses a property similar to the "decorrelation" property described in Buračas and Boynton (2002). In their paper, random designs with this property are observed to yield higher estimation efficiencies than m-sequence-based designs when correlated noise is assumed. In addition to correlated noise, our algorithm can also take into account the second-order polynomial drift, $\boldsymbol{S\gamma}$.

Another feature held by our design pertains to the stimulus frequency. While the stimulus proportion of the m-sequence-based design is always $1/(Q+1)$, that of our design concurs with the approximated optimal proportion of Liu and Frank (2004). The relative frequencies of the three stimulus types in our design are 0.21, 0.22, and 0.22, and the approximated optimal proportion is 0.21. As shown in Simulation 3, this agreement is reached consistently.

When focusing on weighting scheme (C), our GA finds a design close to a block design; see Figure 2.4 (A) in the Appendix. Although this design looks similar to the initial block design, our algorithm does not always yield designs that are similar to the initial ones. For example, when considering the pairwise contrasts between stimulus types, the design found by our GA contains only blocks of stimuli while the initial block design includes also rests; see Figure 2.4 (B) in the Appendix. Our GA tends to converge to a block design when detecting activation is the only concern. The design parameters, including block sizes, number of blocks and the design pattern, are tuned to yield a high efficiency along the evolution of our GA.

Simulation 2

This simulation focuses on the two statistical objectives — detection and estimation. By letting w_d increase from 0 to 1 in steps of 0.05 and keeping $w_c = w_f = 0$ (w_e decreases accordingly), our GA finds designs providing an advantageous trade-off between estimation efficiency and detection power. Figure 2.2 (A) provides the result for individual stimulus
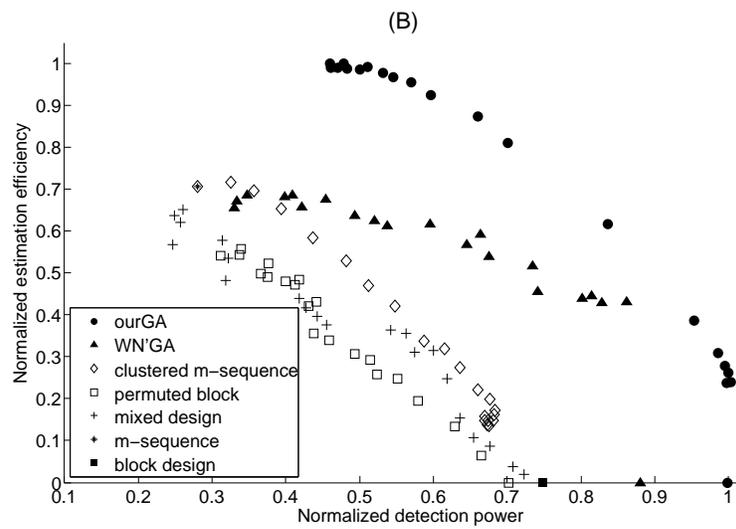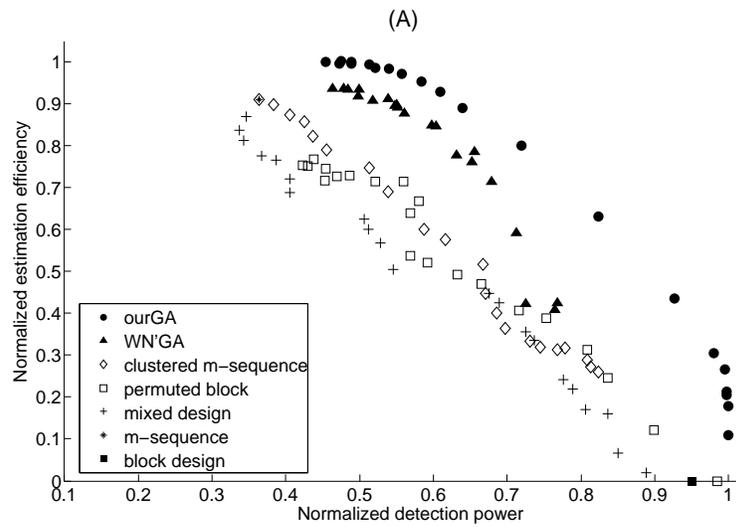
Figure 2.2: Normalized estimation efficiency vs. detection power for different designs: (A) individual stimulus effects; (B) pairwise contrasts.

effects and Figure 2.2 (B) for pairwise contrasts. Again, we compare our designs to WN's designs. Our designs significantly outperform theirs.

In addition, design efficiencies of mixed designs, clustered m-sequences and permuted block designs obtained from Liu (2004) are also presented. We also show in Figure 2.2 the initial m-sequence-based designs of our algorithm, presented as $*$, and the initial block designs, denoted by ∎.

As demonstrated by Liu (2004), mixed designs, clustered m-sequences and permuted block designs can offer advantageous trade-offs between estimation efficiency and detection power when individual stimulus effects and pairwise contrasts are simultaneously of interest. Results for this case are presented in Section 2.4.

SIMULATION 3

In this simulation, we follow Buračas and Boynton (2002) to work on white noise and set $\boldsymbol{S}$ in model (2.1) to a vector of ones, accounting for the overall mean of the fMRI time series. We focus on two separate cases, namely estimating $\boldsymbol{h}$ and estimating $\boldsymbol{h}_i - \boldsymbol{h}_j$ for $1 \leq i < j \leq Q$. Different combinations of Q and L used by Liu (2004) are considered. Our GA then finds designs optimizing the estimation efficiency; i.e., $w_e = 1$. For this comparison, we include only random designs as initial designs in our GA. Due to the computation time, here we let the algorithm run for only 2,000 generations at each combination.

We compare our designs to m-sequence-based designs, which are demonstrated by Buračas and Boynton (2002) to have high estimation efficiencies. The values of F_e achieved by our designs and by m-sequence-based designs are presented in Table 2.1 and Table 2.2. The CPU time spent by our GA is also provided. Even without the help of the m-sequence-based design, our GA consistently finds better designs. As shown in Tables 2.1 and 2.2, the stimulus proportions of our designs are again in good agreement with those optimal values approximated by Liu and Frank (2004).

Table 2.1: The F_e-values and the proportions of the stimuli: individual stimulus effects

number of types (Q)	2	3	4	6	7	8	10	12
length of design (L)	242	255	624	342	511	728	1330	2196
F_e-value								
our GA	41.17	33.34	68.39	26.76	36.20	46.68	72.73	105.18
m-sequence	40.43	31.80	63.08	24.33	31.94	40.72	61.38	85.39
Stimulus proportion								
our GA (min-max)	0.29	0.21-0.23	0.17	0.12-0.13	0.10-0.11	0.09-0.10	0.08	0.06-0.07
approximated optimum	0.29	0.21	0.17	0.12	0.10	0.09	0.08	0.06
CPU time (hours)	0.07	0.11	0.46	0.37	0.68	1.35	4.51	13.09

Table 2.2: The F_e-values and the proportions of the stimuli: pairwise contrasts

number of types (Q)	2	3	4	6	7	8	10	12
length of design (L)	242	255	624	342	511	728	1330	2196
F_e-value								
our GA	56.52	39.04	75.19	25.59	33.38	42.02	62.43	86.56
m-sequence	38.74	30.23	61.70	23.57	31.42	39.95	60.00	84.14
Stimulus proportion								
our GA (min-max)	0.49	0.32-0.33	0.24-0.25	0.16-0.17	0.14	0.12-0.13	0.10	0.08-0.09
approximated optimum	0.50	0.33	0.25	0.17	0.14	0.13	0.10	0.08
CPU time (hours)	0.06	0.11	0.47	0.43	0.87	1.70	5.70	16.49

2.4 CONCLUSIONS AND DISCUSSION

In this chapter, we propose an algorithm to search for optimal ER-fMRI designs. Our proposed algorithm works for any combination of the four popular objectives in ER-fMRI, but is flexible enough to accommodate other goals as well. Through simulations, we show that our algorithm outperforms others currently in use by researchers when either the individual stimulus effects or pairwise contrasts are of interest.

Conceptually, our algorithm follows Holland's (1975) notion of building blocks; see also Goldberg (1989). Rooted in the fundamental theorem of GAs, also known as the schema theorem, the building block hypothesis views these constructs as the driving engine for GAs (Goldberg, 1989). Ensuring a good supply of these building blocks is thus one of the key steps for developing good GAs (Goldberg, 2002; Ahn, 2006). The inclusion of good ER-fMRI

designs as both initial designs and immigrants follows this concept. Furthermore, using good ER-fMRI designs as initial designs also means that our algorithm starts from a good position.

The m-sequence-based design is not included as an initial design in Simulation 3 because our design is compared to this design. For Simulation 3 and quite a few other situations, we can find designs yielding higher estimation efficiencies than m-sequence-based designs without the benefit of an m-sequence-based design among the initial designs. However, this can be hard when both $\boldsymbol{h}$ and pairwise contrasts between $\boldsymbol{h}_i$s are of interest, and the model is with white noise but with neither drift nor trend. For that particular situation, the optimal stimulus proportion is $1/(Q+1)$ and the m-sequence-based design is known to be near-optimal (Liu and Frank, 2004; Liu, 2004). Note that the m-sequence-based designs are known to exist only when $Q+1$ is a prime or a prime power. In contrast, our GA is flexible enough to accommodate any number of stimulus types.

While good initial fMRI designs help to expedite the search, the well-defined design criterion ensures that our GA, when it evolves, finds a better design. As pointed out previously, WN's design criterion is a moving target during the search. Achieving a better design is thus not guaranteed. By contrast, our MO-criterion provides a stable, clear target for the search algorithm.

Our algorithm approximates $\max(F_e)$ and $\max(F_d)$ that are needed for our MO-criterion. A possible alternative is to follow Liu and Frank (2004) to find analytical approximations. For the special cases of the Simulation 2, we apply their approach to find the bound for F_e. It is 34.16 when focusing on individual stimulus effects and is 42.50 for pairwise contrasts. These analytical approximations are larger than our numerical ones, which are 31.96 and 38.21, respectively. However, it is unknown whether their approximated $\max(F_e)$ can actually be achieved by any design. Also, the analytical approximation to $\max(F_d)$ depends on a parameter θ_{min}; see Liu and Frank (2004) for details. Deciding the value of θ_{min} suitable for each situation can be hard. Furthermore, the requisite bounds should adapt to a wide range of conditions, such as different correlation structures and nuisance terms. While these

situations can easily be accommodated in our approach, it can be difficult to analytically derive bounds best suited to each circumstance.

Our algorithm can also be applied when individual stimulus effects and pairwise contrasts are simultaneously of interest. For illustration, consider the same conditions as in Simulation 2 of Section 2.3, where a second-order polynomial drift and AR(1) noise are assumed; such assumptions are closer to reality, compared to the model with white noise and without drift or trend. Figure 2.3 presents the F_e^*-value versus the F_d^*-value achieved by our designs, WN's designs, and the designs studied by Liu (2004). Note that, in the case of detection, the matrix C after (2.3) is the identity matrix for Figure 2.2 (A), and the rows of C for Figure 2.2 (B) represent the pairwise contrasts. Following Liu (2004), the matrix C for Figure 2.3 combines all of these rows into one matrix. Similar comments apply for the estimation problem. Again, our algorithm yields better designs.
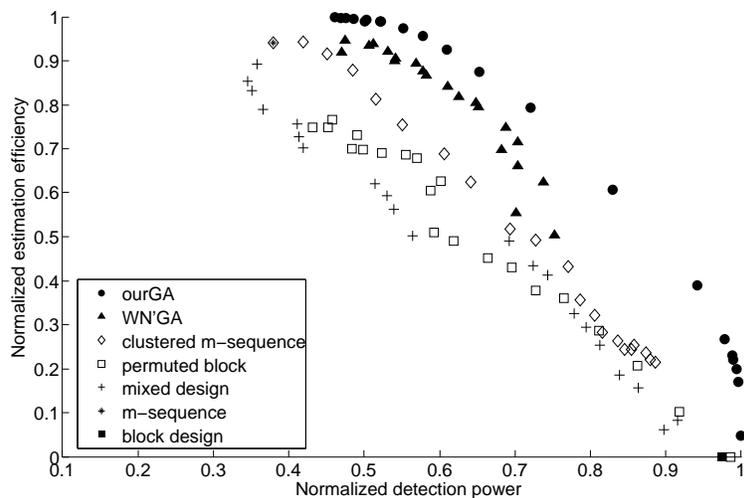
Figure 2.3: Normalized estimation efficiency vs. detection power for different designs when both individual stimulus effects and pairwise contrasts are of interest.

We note that it should be possible to find clustered m-sequences and permuted block designs to reach efficiencies as high as those of our designs. However, unlike our algorithm, the procedures to generate these designs do not attempt to maximize a design optimality

criterion, so that finding good designs of these types using existing algorithms depends on luck. One may be able to develop an effective algorithm that uses the concepts on which these designs are based for finding efficient designs, but that would also require some procedures to hone in on the optimal stimulus frequencies. Pursuing this is beyond the scope of the current work.

One additional advantage of our GA that is not elaborated here, but in Kao et al. (2007), is the formulation of the statistical model when $ISI \neq m\mathrm{TR}$ for any integer m. Our approach applies the discretization interval of Dale (1999) for the HRF parametrization. Denoting the length of this interval as ΔT, we set ΔT to the greatest value dividing both the ISI and TR. The resulting linear models agree with those of WN when $ISI = m\mathrm{TR}$ for some integer m, but our parameters remain interpretable when $ISI \neq m\mathrm{TR}$ for any integer m. Specifically, the ith HRF parameter in WN's model corresponds to the height of the HRF at the ith scan after the stimulus onset. Each parameter in their models may simultaneously represent more than one height of the HRF when $ISI \neq m\mathrm{TR}$. By contrast, our underlying model faithfully reflects the fluctuation in the HRF, and thus results in a more rigorous model formulation.
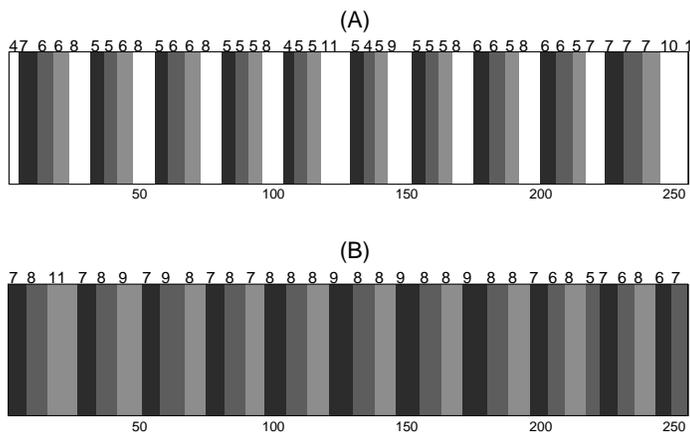
Figure 2.4: Best designs for detection found by our GA: (A) individual stimulus effects; (B) pairwise contrasts.

Appendix

We provide in Figure 2.4 the best designs for detecting activation found by our GA, assuming a second-order Legendre polynomial drift and a stationary AR(1) noise with a correlation coefficient of 0.3; for details see Section 2.3. Figure 2.4 (A) shows the design when the interest lies in individual stimulus effects and Figure 2.4 (B) is for pairwise contrasts. Different shades indicate different stimulus types with white representing rest. The number above each shaded bar presents the number of stimulus types included in that block. Both designs look like block designs. While rest is included in the first design, it is expelled by our GA when the interest lies only in pairwise contrasts. Note that the initial block designs for both searches contain rests.

2.5 References

Ahn, C. W. (2006), *Advances in Evolutionary Algorithms: Theory, Design and Practice*, Studies in computational intelligence, Berlin, New York: Springer.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007), *Optimum Experimental Designs, with SAS*, Great Britain: Oxford University Press.

Bandettini, P. A. and Cox, R. W. (2000), "Event-Related fMRI Contrast When Using Constant Interstimulus Interval: Theory and Experiment," *Magnetic Resonance in Medicine*, 43, 540–548.

Birn, R. M., Cox, R. W., and Bandettini, P. A. (2002), "Detection versus Estimation in Event-Related fMRI: Choosing the Optimal Stimulus Timing," *NeuroImage*, 15, 252–264.

Buračas, G. T. and Boynton, G. M. (2002), "Efficient Design of Event-Related fMRI Experiments Using M-Sequences," *NeuroImage*, 16, 801–813.

Buxton, R. B., Liu, T. T., Martinez, A., Frank, L. R., Luh, W. M., and Wong, E. C. (2000), "Sorting Out Event-Related Paradigms in fMRI: The Distinction Between Detecting an Activation and Estimating the Hemodynamic Response," *NeuroImage*, 11, S457.

Callan, A. M., Callan, D. E., Tajima, K., and Akahane-Yamada, R. (2006), "Neural Processes Involved with Perception of Non-Native Durational Contrasts," *Neuroreport*, 17, 1353–1357.

Dale, A. M. (1999), "Optimal Experimental Design for Event-Related fMRI," *Human Brain Mapping*, 8, 109–114.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. (1995), "Analysis of fMRI Time-Series Revisited," *NeuroImage*, 2, 45–53.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, Massachusetts: Addison-Wesley.

— (2002), *The Design of Innovation : Lessons from and for Competent Genetic Algorithms*, Boston: Kluwer Academic Publishers.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications To Biology, Control, and Artificial Intelligence*, Ann Arbor: University of Michigan Press.

— (1992), *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, Complex adaptive systems, Cambridge, Massachusetts: MIT Press, 1st ed.

Josephs, O., Turner, R., and Friston, K. (1997), "Event-Related fMRI," *Human Brain Mapping*, 5, 243–248.

Kao, M.-H., Mandal, A., Lazar, N., and Stufken, J. (2007), "Multi-Objective Optimal Experimental Designs for Event-Related fMRI Studies," Tech. rep., Department of Statistics, University of Georgia, http://www.stat.uga.edu/∼amandal/MO-fMRI.pdf.

Liu, T. T. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part II: Design of Experiments," *NeuroImage*, 21, 401–413.

Liu, T. T. and Frank, L. R. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part I: Theory," *NeuroImage*, 21, 387–400.

Ramautar, J. R., Slagter, H. A., Kok, A., and Ridderinkhof, K. R. (2006), "Probability Effects in the Stop-Signal Paradigm: The Insula and the Significance of Failed Inhibition," *Brain Research*, 1105, 143–154.

Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998), "Event-Related functional MRI: Past, Present, and Future," *PNAS*, 95, 773–780.

Summerfield, C., Egner, T., Mangels, J., and Hirsch, J. (2006), "Mistaking a House for a Face: Neural Correlates of Misperception in Healthy Humans," *Cerebral Cortex*, 16, 500–508.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of Experimental Design in fMRI: A General Framework Using A Genetic Algorithm," *NeuroImage*, 18, 293–309.

Wang, Y. P., Xue, G., Chen, C. S., Xue, F., and Dong, Q. (2007), "Neural Bases of Asymmetric Language Switching in Second-Language Learners: An ER-fMRI Study," *Neuroimage*, 35, 862–870.

Worsley, K. J. and Friston, K. J. (1995), "Analysis of fMRI Time-Series Revisited–Again," *NeuroImage*, 2, 173–181.

CHAPTER 3

MULTI-OBJECTIVE OPTIMAL EXPERIMENTAL DESIGNS FOR ER-FMRI USING

MATLAB[1]

3.1 INTRODUCTION

Event-related functional magnetic resonance imaging (ER-fMRI) is one of the leading technologies for studying human brain activity in response to mental tasks or stimuli. Before conducting an ER-fMRI experiment, a sequence of stimuli of one or more types interlaced with the control (rest or fixation) is prepared. This sequence of stimuli is presented to an experimental subject while the MR scanner scans his/her brain every few seconds. The blood oxygenation level dependent (BOLD) time series is collected from each brain voxel (a small region of the brain) for statistical analysis. See Josephs et al. (1997), Rosen et al. (1998), Dale (1999), and Bandettini and Cox (2000) for overviews of ER-fMRI.

One important design problem of ER-fMRI is to find an optimal sequence of the stimuli best suited to the researcher's needs. However, this problem is difficult due to the following reasons. First, the design space, consisting of all possible sequences of stimuli, is enormous. Searching over this space for a good design is hard. In addition, the flexibility of ER-fMRI allows researchers to consider two popular statistical objectives, namely estimation and detection. Estimation refers to the estimation of the hemodynamic response function (HRF), a function of time describing an effect of a single, brief stimulus. Detection is to identify brain regions that are activated by the stimuli. Considering both objectives in one study is not uncommon (see also Wager and Nichols, 2003), but this requires good multi-objective designs that efficiently achieve these two competing goals. Moreover, statistics is not the only concern for the design of ER-fMRI experiments. Psychology plays an important, even crucial, role. When a design sequence is patterned or easy to predict, psychological effects such as habituation or anticipation can occur to contaminate the data (Dale, 1999). A good design should help to avoid these confounds. Furthermore, customized requirements such as the required number of stimuli for each stimulus type can also arise to make the problem even more complicated. As a result, searching for a good multi-objective design is inevitable. We need well-defined multi-objective design criteria (MO-criteria) for evaluating the quality of

competing designs, an efficient search algorithm and a program that implements such an algorithm.

In this chapter, we develop a program using MATLAB (The MathWorks, Inc., 2006) for finding multi-objective optimal ER-fMRI designs. Our program implements the approach proposed by Kao et al. (2009a), which includes rigorously formulated models, well-defined MO-criteria and a genetic algorithm (GAs). They incorporate knowledge about the performance of well-known ER-fMRI designs to increase the effectiveness and efficiency of their approach. As demonstrated in their paper, this approach is more efficient than the previous methods and is flexible enough to accommodate different experimental conditions and assumptions. To make the best use of this approach, our program allows users to specify the experimental conditions based on their needs. The designs that we obtain can help researchers to achieve efficient statistical inference.

The rest of the chapter is organized as follows. Section 3.2 reviews the approach proposed by Kao et al. (2009a). Section 3.3 illustrates our computer codes. An example for using our program is in Section 3.4, followed by conclusion and discussion in Section 3.5.

3.2 Methodology Review

A typical ER-fMRI design can be viewed as an alignment of events, including the stimuli and the control. For convenience, the symbols $0, 1, ..., Q$ are used to represent the events with 0 indicating the control and i a type-i stimulus, $i = 1, ..., Q$; Q is the total number of stimulus types. A design, denoted by ξ, looks like $\xi = \{101201210...1\}$.

While being presented to an experimental subject, each stimulus lasts for a short period of time relative to the inter-stimulus interval (ISI), the fixed time interval between the onsets of consecutive events. We note that 0s in the sequence are "pseudo-events"; they help to calculate the onset times of stimuli. For example, with a 0 in between, the first, second and the third stimuli (1, 1, and 2) of ξ occur, respectively, 1ISI, 3ISI, and 4ISI seconds after the

outset of the experiment. The control fills up the time period between the end of a stimulus and the start of the next one.

Our goal is to find a best sequence of the events to efficiently achieve four popular objectives, which are 1) estimating the HRF, 2) detecting brain activation, 3) avoiding psychological confounds and 4) maintaining the desired frequency for each stimulus type. To define the design criteria for the first two statistical objectives, we need to specify the underlying models.

Models

Following previous approaches (e.g. Liu, 2004; Liu and Frank, 2004; Wager and Nichols, 2003), two popular linear models are considered for the two statistical objectives; see also Friston et al. (1995), Worsley and Friston (1995), and Dale (1999). These models are:

$$Y = Xh + S\gamma + e; \tag{3.1}$$

$$Y = Z\theta + S\gamma + \eta, \tag{3.2}$$

where Y is the voxel-wise BOLD time series, $h = (h_1', ..., h_Q')'$ is the parameter vector for the HRFs of the Q stimulus types, $X = [X_1 \cdots X_Q]$ is the design matrix, $\theta = (\theta_1, ..., \theta_Q)'$ represents the response amplitudes, $Z = Xh_0$ is the convolution of stimuli with an assumed basis, h_0, of the HRF, $S\gamma$ is a nuisance term describing the trend or drift of Y, and e and η are noise. We assume a known whitening matrix, V, such that Ve and $V\eta$ are white noise. The whitening matrix can be obtained empirically from previous experiments; see also Wager and Nichols (2003). Model (3.1) is typically used for estimating the HRF and model (3.2) is for detecting activation. Note that, for detection problems, a basis h_0 for the HRF needs to be assumed.

To enable the use of a finite set of interpretable parameters to capture the fluctuation of the continuous HRF over time, the discretization interval (Dale, 1999) is utilized for parameterizing the HRF in model (3.1). The length of the discretization interval, denoted

by ΔT, is set to the greatest value dividing both the ISI and TR; the TR is the time interval between consecutive MR scans. The HRF parameters, captured in the vector $\boldsymbol{h}$, then represent the heights of the HRF for each stimulus after every ΔT seconds following the stimulus onset. This parametrization is explained in detail in Kao et al. (2009b).

Design Criterion

For the two statistical objectives, two popular optimality design criteria, namely A- and D-optimality criteria (Atkinson et al., 2007), are considered. A-optimality aims at minimizing the average variance of estimators of parametric functions. A D-optimal design minimizes the generalized variance of estimators of linearly independent parametric functions, or, under normality, it minimizes the volume of simultaneous elliptical confidence regions for these parametric functions at any specified confidence level. The value of the design criterion for estimating the HRF, referred to as "estimation efficiency", is denoted by F_e. Likewise, the term "detection power" and the notation F_d are used to indicate the value of the design criterion for detecting activation. These criteria have the forms of $F_i = r_c/trace(\boldsymbol{M})$ for A-optimality, or of $F_i = det(\boldsymbol{M})^{-1/r_c}$ for D-optimality; i=d, e. Here, $\boldsymbol{M} = \boldsymbol{C}[\boldsymbol{W}'\boldsymbol{V}'(\boldsymbol{I} - P_{\boldsymbol{VS}})\boldsymbol{VW}]^-\boldsymbol{C}'$, $\boldsymbol{W} \equiv \boldsymbol{X}$ for F_e, $\boldsymbol{W} \equiv \boldsymbol{Z}$ for F_d, $\boldsymbol{I}$ is an identity matrix, $P_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A}'$ is the orthogonal projection on the vector space spanned by the column vectors of $\boldsymbol{A}$, $\boldsymbol{A}^-$ is a generalized inverse matrix of $\boldsymbol{A}$, $\boldsymbol{C}$ is a matrix of estimable linear combinations of the parameters, and r_c is the number of rows of $\boldsymbol{C}$. We note that F_e and F_d are "larger-the-better" criteria.

The third objective is to avoid psychological confounds. We would like a sequence that makes it difficult for a subject to anticipate future stimuli based on past stimuli. Designs minimizing the following criterion help to achieve this objective.

$$F_c = \sum_{r=1}^{R}\sum_{i=1}^{Q}\sum_{j=1}^{Q}\lfloor|n_{ij}^{(r)} - (n-r)P_iP_j|\rfloor.$$

Here, the sub-design excluding all 0s but retaining all the stimuli of the original design is considered. In this design criterion, n is the length of the sub-design, and $n_{ij}^{(r)}$ is the number

of times that i and j are r elements away in the sub-design; i.e., they are, respectively, the tth and the $(t+r)$th elements, $t = 1, ..., (n-r)$. P_i is the specified proportion for the type-i stimulus in the sub-design which may be taken as $1/Q$ if there is no preference, and $\lfloor |a| \rfloor$ is the integer part of the absolute value of a. R is a given integer; the unpredictability of a design increases when R increases. Therefore, F_c aims at having a design with each pair appearing a number of times that is proportional to the product of the specified proportions for the stimuli. Designs minimizing this criterion are said to be Rth order counterbalanced (cf. Wager and Nichols, 2003).

The fourth design criterion helps to maintain the desired stimulus frequency and is also defined on the sub-design. It is $F_f = \sum_{i=1}^{Q} \lfloor |n_i - nP_i| \rfloor$, where n_i is the number of the type-i stimulus in the sub-design. Designs achieving the desired stimulus frequency minimize F_f.

With these four individual design criteria, the family of MO-criteria is then defined as

$$\{F^* = w_c F_c^* + w_d F_d^* + w_e F_e^* + w_f F_f^* : \ w_i \geq 0, \ i = c, d, e, f; \ \sum_i w_i = 1\}, \qquad (3.3)$$

where w_is are weights selected based on the researcher's emphasis in a given study, $F_i^* = F_i/max(F_i)$ for $i = d, e$ and $F_i^* = 1 - [F_i/max(F_i)]$ for $i = c, f$. We note that, in Kao et al. (2009a), $min(F_i)$ is also considered when calculating F_i^*. Since these minimal values are zero, they are omitted here. Each member of this family can serve as an objective function of the search algorithm.

SEARCH ALGORITHM

The search algorithm of Kao et al. (2009a) is built upon the genetic algorithm technique (Holland, 1975; 1992). This technique is popular for solving optimization problems, in which good solutions (parents) are used to generate better ones (offsprings). To efficiently apply this technique to our problem, well-known results about good fMRI designs are incorporated so that the search over the huge design space can be carried out more efficiently. The outline of the algorithm is as follows:

Step 1. (Initial designs) Generate G initial designs consisting of random designs, an m-sequence-based design, a block design and their combinations. Use the objective function to evaluate the fitness of each initial design.

Step 2. (Crossover) With probability proportional to fitness, draw with replacement $G/2$ pairs of designs to crossover — select a random cut-point and exchange the corresponding fractions of the paired designs.

Step 3. (Mutation) Randomly select $q\%$ of the events from the G offspring designs. Replace these events by randomly generated ones. Here, an event is a stimulus or the control.

Step 4. (Immigration) Add to the population another I designs drawn from random designs, block designs and their combinations.

Step 5. (Fitness) Obtain the fitness scores of the offsprings and immigrants.

Step 6. (Natural selection) Keep the best G designs according to their fitness scores to form the parents of the next generation. Discard the others.

Step 7. (Stop) Repeat steps 2 through 6 until a stopping rule is met; e.g. after M_g generations. Keep track of the best design over generations.

In Step 1, m-sequences or m-sequence-based designs are generated following Liu (2004); see also Buračas and Boynton (2002). These designs are well-known for their high estimation efficiencies. Good designs for detection are block designs. A block design is a sequence where stimuli of the same type are clustered into blocks. For example, a two-stimulus-type block design with a block size of four can consist of repetitions of {111122220000}. Repetitions of {1111000022220000} and {11112222} are other possible patterns. In steps 1 and 4, block designs with various block sizes and patterns are considered. A fraction of an m-sequence-based design or a random design is combined with a fraction of a block design to form a mixed design. These mixed designs along with random designs are also included as part of the initial designs and immigrants.

Table 3.1: Input parameters: fields of the structure Inp; e.g., use $Inp.nSTYPE = 2$ to assign the number of stimulus types. $\boldsymbol{I}_a$ is the a-by-a identity matrix.

Parameter	Type	Description	Default
nSTYPE	integer	number of stimulus types (Q)	2
ISI	real	inter-stimulus interval; the time interval between event onsets (in seconds)	2.0
TR	real	time to repetition; the time interval between MR scans (in seconds)	2.0
dT	real	the discretization interval, $\Delta T =$ the greatest value dividing both the ISI and TR	2.0
nEvents	integer	number of events of the design	242
Smat	real matrix	the matrix $\boldsymbol{S}$ in models (3.1) and (3.2)	2nd-order polynomial
V2	real matrix	the square of the whitening matrix	see (3.4)
Opt	integer	0=A-optimality; 1=D-optimality	0
MOweight	real array	weights of the four objectives, $[w_c, w_d, w_e, w_f]$	$[\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}]$
basisHRF	real array	the basis function $\boldsymbol{h}_0$ for the HRF	canonical HRF of SPM2
durHRF	real	duration of the HRF (in seconds)	32.0
CX	real matrix	linear combinations of interest for the HRF parameters, h_{ij}; $i = 1, ..., Q, j = 1, ..., (1+\lfloor\frac{K}{\Delta T}\rfloor)$	$\boldsymbol{I}_{34}$
CZ	real matrix	linear combinations of interest for the amplitudes, θ_i; $i = 1, ..., Q$	$\boldsymbol{I}_2$
MaxFe	real	max(F_e)	39.2715
MaxFd	real	max(F_d)	132.0670
cbalR	integer	the order of counterbalancing, R, for F_c	3
stimFREQ	real array	the desired frequency of each stimulus type P_i; the array should sum to one	$[\frac{1}{2}\ \frac{1}{2}]$
sizeGen	integer	number of designs in each GA generation; an even number	20
qMutate	real	rate of mutation	0.01
nImmigrant	integer	number of immigrants	4
StopRule	integer	1=stop the search after numITR generations; 2=stop the search when there is no significant improvement	1
numITR	integer	total number of GA generations when StopRule = 1; check the accumulated improvement every numITR generations when StopRule = 2	10000
improve	real	when StopRule = 2, the value of δ in (3.5)	10^{-7}
SaveEvery	integer	save results every SaveEvery generations; 0=save results after completing the search	0
Nonlinear	integer	0=assume linearity; 1=incorporate nonlinear effects	0

3.3 CODE DESCRIPTION

In this section, we describe our MATLAB program, including the input parameters, the way to run the program and the output variables.

3.3.1 INPUT PARAMETERS

Table 3.1 presents all the input parameters of our program along with their default values. Due to the complexity of ER-fMRI, a few parameters are needed for specifying experimental conditions and assumptions, which might vary from experiment to experiment. An example code, Par_Assign.m, is therefore provided to help to assign these parameter values. The values are specified through the MATLAB structure *Inp*. Some fields of *Inp* are detailed below. Descriptions for other input parameters can be found in Table 3.1.

NUISANCE TERM

The nuisance term $\boldsymbol{S}$ in (3.1) and (3.2) is specified in *Inp.Smat*. By default, $\boldsymbol{S}$ corresponds to a second-order Legendre polynomial drift. The degree of the polynomial can be changed in Par_Assign.m through the *PolyOrder* variable. While polynomial drift is popular (e.g. Worsley et al., 2002; Liu, 2004), other nuisance terms can also be considered.

WHITENING MATRIX

The square of the whitening matrix, $\boldsymbol{V}^2$, described after (3.2) is specified in *Inp.V2*. By default, the following matrix, which corresponds to a stationary AR(1) process, is considered:

$$\boldsymbol{V}^2 = \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{pmatrix}. \tag{3.4}$$

The value of ρ is set to 0.3 and it can be changed in Par_Assign.m through the *rho* variable. Other whitening matrices can be considered as well.

Table 3.2: Output variables: fields of the structure *Out*; e.g., *Out.bestList* contains the best design achieved.

Variables	Description
bestList	the best design achieved
bestLists	the designs achieved over generations
bestOVF	the maximum of the objective function that we achieve
OVF	the values of the objective function achieved over GA generations
bestidvF	$[F_c^*, F_d^*, F_e^*, F_f^*]$ of the best design that we achieve
idvF	$[F_c^*, F_d^*, F_e^*, F_f^*]$ over GA generations
timespend	CPU time spent for the search (in seconds)

LINEAR COMBINATIONS OF PARAMETERS

Linear combinations of the parameters of interest are specified in *Inp.CX* for model (3.1) and in *Inp.CZ* for model (3.2). These fields are, by default, set to identity matrices, allowing the study of individual stimulus effects. The number of columns for *Inp.CX* equals to the length of $\boldsymbol{h}$. For a K-second HRF (by default, $K = 32$), the length of $\boldsymbol{h}$ is $Q(1 + \lfloor K/\Delta T \rfloor)$. The number of columns for *Inp.CZ* is Q, corresponding to the length of $\boldsymbol{\theta}$.

In addition to setting *Inp.CX* and *Inp.CZ* to identity matrices, researchers might also be interested in pairwise contrasts between stimulus types. Kao et al. (2009b) provides a systematic study of designs for convex combinations of these two interests. The example code, Exp_combinedInterest.m, is provided here for the case where equal weights are assigned to both interests.

MAXIMAL VALUES OF THE INDIVIDUAL CRITERIA

Values of $\max(F_e)$ and $\max(F_d)$ are assigned through *Inp.MaxFe* and *Inp.MaxFd*, respectively. These values are used to standardize F_e and F_d before combining them into an MO-criterion. Theoretical values of $\max(F_e)$ and $\max(F_d)$ are generally not available. We

therefore obtain numerical approximations by performing "pre-runs" of our program. For approximating $\max(F_e)$, we use *Inp.MaxFe = 1* and *Inp.MOweight = [0 0 1 0]*; i.e., $w_e = 1$ in (3.3). This is equivalent to using the non-standardized F_e as the objective function. The F_e-value achieved by the design that we obtain approximates $\max(F_e)$. Similarly, we can use *Inp.MaxFd = 1* and *Inp.MOweight = [0 1 0 0]* to find the optimal design for detection and to numerically approximate $\max(F_d)$. These approximations can then be specified in *Inp.MaxFe* and *Inp.MaxFd* for further searches for multi-objective optimal designs. PreRun_Fe.m is an example code for approximating $\max(F_e)$ and PreRun_Fd.m is for $\max(F_d)$. The maximal values of the other two criteria, F_c and F_f, are automatically calculated in our program.

Basis for the HRF

We need to assume a basis for the HRF when using model (3.2). A popular choice is the canonical HRF of SPM2 (The Wellcome Trust Centre for Neuroimaging, 2003), a popular software for fMRI. This basis is a combination of two Gamma distributions. In our program, we use this canonical HRF, scaled to have a maximum of one, as the default setting for $\boldsymbol{h}_0$. The parameters, such as the time-to-peak and time-to-onset, used to create the canonical HRF can also be altered in Par_Assign.m. By changing *Inp.basisHRF*, other basis functions can be considered as well.

Stopping Rules

We consider two stopping rules. The first stopping rule terminates the search after M_g generations. The second stopping rule is inspired by Liefvendahl and Stocki (2006). This second method considers the accumulated improvement of the design efficiency from the $\ell n + 1$st generation to the $(\ell + 1)n$th generation; $\ell = 0, 1, 2, ...$ Denote the accumulated improvement by $[F^*(\xi^*_{(\ell+1)n}) - F^*(\xi^*_{\ell n+1})]$, we stop the search at the $(\ell + 1)n$th generation when the following condition is met (for given n and δ):

$$F^*(\xi^*_{(\ell+1)n}) - F^*(\xi^*_{\ell n+1}) \leq \delta[F^*(\xi^*_n) - F^*(\xi^*_1)]. \tag{3.5}$$

To use the first stopping rule, we set *Inp.StopRule = 1*, and *Inp.numITR = M_g*. By setting *Inp.StopRule = 2*, the second stopping rule is considered. In this case, we set *Inp.numITR* to n (say, 100) and *Inp.improve* to δ (e.g., 10^{-7}).

3.3.2   RUNNING THE CODE

The m-file Par_Assign.m can directly be used to perform the search for optimal designs. With user-specified parameter values, this m-file calls the subroutine *fMRIMOD(Inp)* to start the search. Our program requires subroutines from SPM2 (The Wellcome Trust Centre for Neuroimaging, 2003) and *mttfmri* (Liu, 2004a, see also Liu, 2004). From SPM2, we need spm_Gpdf.m and spm_hrf.m to calculate the canonical HRF. From *mttfmri*, we apply gen_mseq.m, qadd.m, qmult.m, mseq2.m and return_mtaps.m to generate m-sequence-based designs, and stimpatch.m for plotting the final design. These m-files are freely downloadable from The Wellcome Trust Centre for Neuroimaging (2003) and Liu (2004a), respectively.

3.3.3   OUTPUT VARIABLES

The output variables of our program are listed in Table 3.2. The best design achieved by our program is contained in *Out.bestList* and its design efficiency is *Out.bestOVF*. Our program also tracks the best designs over generations (*Out.Lists*) and their design efficiencies (*Out.OVF*). The value for each individual criterion F^*_i is also provided. Time spent on the search is recorded, too.

3.4   AN EXAMPLE

An illustrative example is described in this section. We consider ISI $= TR = \Delta T = 2$s, so that *Inp.ISI = 2.0*, *Inp.TR = 2.0* and *Inp.dT = 2.0*. The number of stimulus types is set to $Q = 2$ (*Inp.nSTYPE = 2*). A total of 242 events (stimuli plus the control) are included
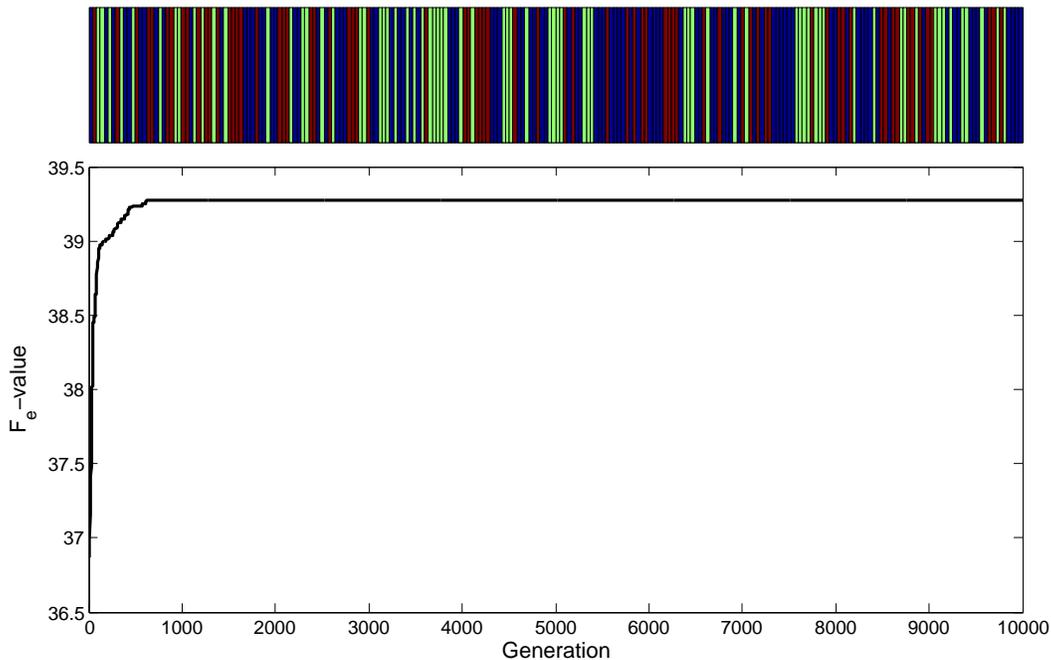
Figure 3.1: The design achieved by our program for estimation and F_e-values achieved over GA generations.

in the design sequence; i.e., $Inp.nEvents = 242$. A second-order polynomial drift and the AR(1) noise with the correlation coefficient of 0.3 are assumed for models (3.1) and (3.2). A-optimality criterion ($Inp.Opt = 0$) is used for both statistical objectives, including estimation and detection.

The canonical HRF, scaled to have a maximum of one, is used as the basis function $\boldsymbol{h}_0$ of model (3.2); see also Wager and Nichols (2003); Wager et al. (2005). After discretization using ΔT, this basis is assigned to $Inp.basisHRF$. Following the default setting of the canonical HRF, the duration of the HRF in model (3.1) is $K = 32$s, so that $Inp.durHRF = 32.0$. The number of parameters contained in each $\boldsymbol{h}_i$ of model (3.1) is therefore $17(= 1 + \lfloor K/\Delta T \rfloor)$,

and the length of $\boldsymbol{h}$ is 34. To investigate individual stimulus effects, we set $Inp.CX = \text{eye}(34)$ and $Inp.CZ = \text{eye}(2)$; they are identity matrices.

For algorithmic parameters, we use the first stopping rule to terminate the search after $M_g = 10,000$ generations. Each generation consists of $G = 20$ designs. The mutation rate is $q = 1\%$ and the number of immigrants is set to $I = 4$. Therefore, $Inp.StopRule = 1$, $Inp.numITR = 10000$, $Inp.sizeGen = 20$, $Inp.qMutate = 1$ and $Inp.nImmigrant = 4$.

We implement our simulations by using MATLAB (version 7.3) on a Linux cluster with 64-bit AMD Opteron, dual-processor, mix of single-core node and dual-core node; each core has 2GB RAM and the Linux operating system is 2.6.9-78.0.5.ELsmp.

We first find the (near-)optimal design for estimating the HRF by setting $Inp.MOweight = [0\ 0\ 1\ 0]$ and $Inp.MaxFe = 1$. The resulting design is presented in Figure 3.1 along with the curve of the achieved efficiencies over generations. In Figure 3.1, each bar in the design indicates an event. Different colors represent different event types with blue indicating the control. This design looks rather random and its F_e-value is 39.2715. The CPU time spent for this search is 0.23 hours.

We then search for the best design for detecting activation. We set ($Inp.MOweight = [0\ 1\ 0\ 0]$) and $Inp.MaxFd = 1$. As presented in Figure 3.2, the resulting design looks like a block design. This design starts with five 0s, followed by eight stimuli of the first type and nine stimuli of the second type. The F_d-value achieved by this design is 132.0670. We spend 0.13 hours of CPU time on this search.

We can also assign equal weights to the four objectives to search for a multi-objective optimal design; i.e., $Inp.MOweight = [1/4\ 1/4\ 1/4\ 1/4]$. The maximal values of F_e and F_d are approximated numerically; i.e., $Inp.MaxFe = 39.2715$ and $Inp.MaxFd = 132.0670$. For F_c, we require a third-order counterbalancing property (R = 3), so that $Inp.cbalR = 3$. For F_f, equal frequencies for the two stimulus types are required; i.e., $P_i = 1/2$, i = 1, 2, and $Inp.stimFREQ = [1/2\ 1/2]$. Note that, when assigning $Inp.stimFREQ$, we do not take into account the number of the control event. Therefore, the actual frequency of each stimulus
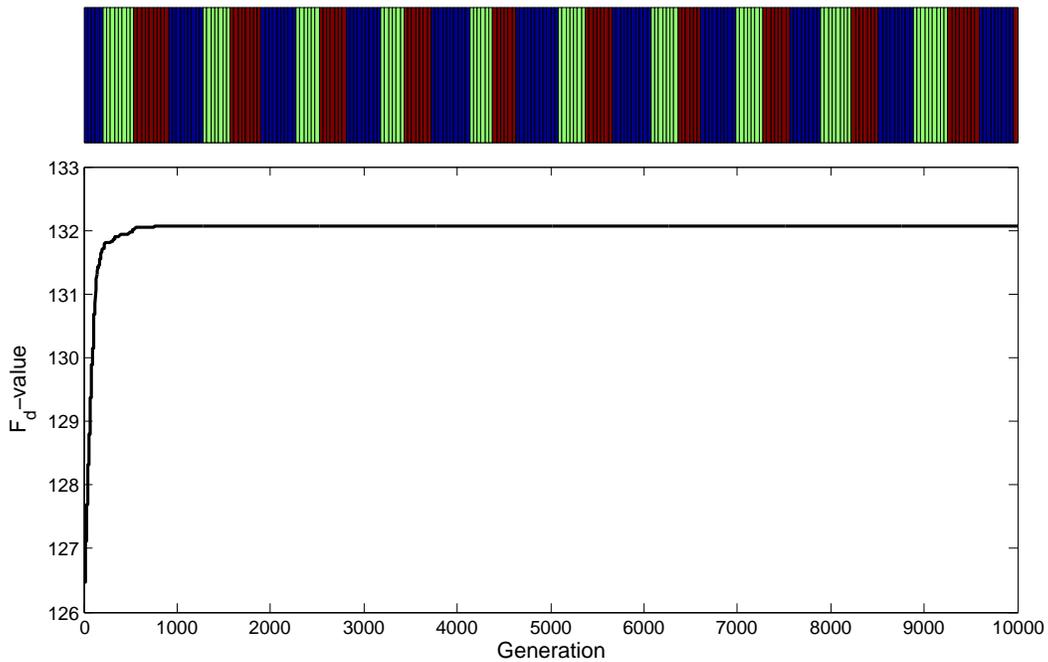
Figure 3.2: The design achieved by our program for detection and F_d-values achieved over GA generations.

type in this example is less than 1/2. The number of the control event is decided by the GA search based on other requirements (design criteria). In our experience, the number of the control event is greatly influenced by the linear combinations of parameters of interest; see also, Kao et al. (2009b). The actual stimulus frequency of our designs agrees with the approximated optimal stimulus frequency of Liu and Frank (2004).

The parameter values for this last search are the same as those listed in Table 3.1 and those in Par_Assign.m. This search requires 0.42 hours of CPU time.

3.5 CONCLUSION AND DISCUSSION

Optimal designs are crucial to the success of ER-fMRI experiments. Due to the nature of ER-fMRI, planning a good design is very complicated. Therefore, an efficient program that helps to find such good designs is called for. In this chapter, we develop a program using MATLAB to search for multi-objective optimal experimental designs for ER-fMRI. The algorithm utilized in our program is proposed by Kao et al. (2009a), which is shown to outperform the previous approaches. We describe the use of our program. An example is provided for illustration. In addition to default settings, we allow the users to assign the parameter values so that our program can achieve designs best suited to the researcher's needs.

The approach considered in this chapter is built upon popular linear models. However, the assumption of linearity may be invalid when the stimuli are very close, and a 'saturation' in the accumulated BOLD response is observed (Wager et al., 2005). To take into account such a nonlinear effect, Wager and Nichols (2003) propose to use 2 to replace the elements of the matrix $\boldsymbol{Z}$ in model (3.2) that are greater than 2. This is also allowed in our program by setting *Inp.Nonlinear* to 1. Developing a more sophisticated method for incorporating such nonlinear effects can be useful.

3.6 REFERENCES

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007), *Optimum Experimental Designs, with SAS*, Great Britain: Oxford University Press.

Bandettini, P. A. and Cox, R. W. (2000), "Event-Related fMRI Contrast When Using Constant Interstimulus Interval: Theory and Experiment," *Magnetic Resonance in Medicine*, 43, 540–548.

Buračas, G. T. and Boynton, G. M. (2002), "Efficient Design of Event-Related fMRI Experiments Using m-Sequences," *NeuroImage*, 16, 801–813, part 1.

Dale, A. M. (1999), "Optimal Experimental Design for Event-Related fMRI," *Humman Brain Mapping*, 8, 109–114.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. (1995), "Analysis of fMRI Time-Series Revisited," *NeuroImage*, 2, 45–53.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, Ann Arbor: University of Michigan Press.

— (1992), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, Complex adaptive systems, Cambridge, Massachusetts: MIT Press, 1st ed.

Josephs, O., Turner, R., and Friston, K. (1997), "Event-Related fMRI," *Human Brain Mapping*, 5, 243–248.

Kao, M.-H., Mandal, A., Lazar, N., and Stufken, J. (2009a), "Multi-Objective Optimal Experimental Designs for Event-Related fMRI Studies," *NeuroImage*, 44, 849–856.

Kao, M.-H., Mandal, A., and Stufken, J. (2009b), "Optimal Design for Event-Related Functional Magnetic Resonance Imaging Considering Both Individual Stimulus Effects and Pairwise Contrasts," *Statistics and Applications*, to appear (Special Volume in Honour of Professor Aloke Dey).

Liefvendahl, M. and Stocki, R. (2006), "A Study on Algorithms for Optimization of Latin Hypercubes," *Journal of Statistical Planning and Inference*, 136, 3231–3247.

Liu, T. T. (2004a), *mttfmri*: Multiple Trial Type fMRI MATLAB Toolbox, University of California at San Diego, Center for Functional Magnetic Resonance Imaging.

— (2004b), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part II: Design of Experiments," *NeuroImage*, 21, 401–413.

Liu, T. T. and Frank, L. R. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part I: Theory," *NeuroImage*, 21, 387–400.

Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998), "Event-Related Functional MRI: Past, Present, and Future," *PNAS*, 95, 773–780.

The MathWorks, Inc. (2006), *MATLAB – The Language of Technical Computing, Version 7.3*, The MathWorks, Inc., Natick, Massachusetts.

The Wellcome Trust Centre for Neuroimaging (2003), *SPM2 – Statistical Parametric Mapping*, The Wellcome Trust Centre for Neuroimaging at the University College London, London, United Kingdom.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of Experimental Design in fMRI: A General Framework Using a Genetic Algorithm," *NeuroImage*, 18, 293–309.

Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. (2005), "Accounting for Nonlinear BOLD Effects in fMRI: Parameter Estimates and a Model for Prediction in Rapid Event-Related Studies," *NeuroImage*, 25, 206–218.

Worsley, K. J. and Friston, K. J. (1995), "Analysis of fMRI Time-Series Revisited – Again," *NeuroImage*, 2, 173–181.

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., and Evans, A. C. (2002), "A General Statistical Analysis for fMRI Data," *NeuroImage*, 15, 1–15.

CHAPTER 4

OPTIMAL DESIGN FOR EVENT-RELATED FUNCTIONAL MAGNETIC RESONANCE IMAGING CONSIDERING BOTH INDIVIDUAL STIMULUS EFFECTS AND PAIRWISE CONTRASTS[1]

4.1 Introduction

ER-fMRI is one of the leading technologies for studying human brain activity in response to brief mental stimuli or tasks. Unlike the traditional fMRI, where long-period stimuli are used, ER-fMRI takes advantage of an ultra-fast MR scanner to allow the study of an effect due to a single, brief stimulus. ER-fMRI is a popular technique for brain mapping in both medical practice and scientific research and is arguably the most important advance in neuroscience (Rosen et al., 1998; Josephs and Henson, 1999; Culham, 2006).

A design for an ER-fMRI study consists of a sequence of stimuli of one or more types interlaced with a control condition (rest or fixation). Finding an optimal sequence of the stimuli best suited to the researcher's need can be arduous. There are multiple reasons for this. First, an ER-fMRI design is a long sequence of finite numbers, typically consisting of hundreds of elements. The design space containing all possible arrangements is thus enormous and irregular (Buračas and Boynton, 2002; Liu, 2004). Second, the nature of ER-fMRI requires consideration of multiple objectives in a study. These objectives involve not only statistical goals but also psychological constraints. In addition, customized requirements can arise to make the problem even more complicated. To overcome these difficulties, an efficient search algorithm along with a well-defined multi-objective design criterion is called for.

An efficient approach to search for multi-objective optimal designs for ER-fMRI is proposed by Kao et al. (2008). Their search algorithm is a genetic algorithm (GA) and their multi-objective design criteria are convex combinations of criteria for single objectives. Two popular statistical objectives in ER-fMRI are estimation and detection. Estimation refers to the estimation of the hemodynamic response function (HRF), a function of time describing the effect on the brain of a single, brief stimulus. Detection aims at investigating whether a region is activated by each stimulus type. This is accomplished by separately studying the amplitudes (or the peaks) of the HRFs evoked by different stimulus types. For both statistical objectives, a researcher may be interested in studying individual stimulus effects and pairwise contrasts of stimulus effects (Amaro and Barker, 2006; Liu and Frank, 2004; Don-

aldson and Buckner, 2002). Kao et al. (2008) consider optimal designs for each of these two interests. However, following Liu and Frank (2004) and Liu (2004), in this chapter we study designs that are efficient if both individual effects and pairwise contrasts are of interest. In contrast to earlier work, our approach allows user-specified weights for individual stimulus effects and pairwise contrasts.

The search algorithm proposed by Kao et al. (2008) is adopted to search for optimal designs, but we define a different family of MO-criteria to meet the goal of this study. Each MO-criterion is a weighted sum of the criteria for estimation and detection, and each of the latter criteria is defined based on a convex combination of the criteria for individual stimulus effects and pairwise contrasts. We study efficient designs that we obtain by using different weighting schemes, and compare them to designs currently in use by fMRI researchers.

In the following section, we briefly introduce background information regarding ER-fMRI designs. Section 4.3 describes our methodology including the statistical model, the design criteria and the search algorithm. Numerical results are presented in Section 4.4 and conclusions are provided in Section 4.5.

4.2   ER-FMRI DESIGNS

An ER-fMRI design is an alignment of events, including stimuli of different types and the control. For convenience, the symbols $0, 1, ..., Q$ are used to represent the events, where Q is the total number of stimulus types. The control is indicated by 0 and a type-i stimulus is denoted by i, $i = 1, ..., Q$. A design, denoted by ξ, looks like $\xi = \{101201210...1\}$. While being presented to the experimental subject, each event lasts for a short period of time relative to the inter-stimulus interval (ISI), the fixed time interval between event onsets. We note that 0s in a sequence are "pseudo-events". They may be thought of as periods of rest for the subject, even though the subject may still experience effects from previous events.

While the design sequence is presented, the MR scanner scans the subject's brain every few seconds; the time duration between two scans is referred to as TR or time to repetition.

The blood oxygenation level dependent (BOLD) signal at each brain voxel (a small region of the brain) is collected every TR seconds to form a voxel-wise fMRI time series. A design issue for ER-fMRI is to allocate the stimuli so that statistical inference (related to estimation or detection) on these time series is most efficient, in some sense.

Depending on the study objectives, several well-known ER-fMRI designs currently in use by researchers are block designs, m-sequence-based designs, mixed designs, permuted block designs, and clustered m-sequences (Liu, 2004). In ER-fMRI, a block design is a sequence where stimuli of the same type are clustered into blocks. For example, a two-stimulus-type block design with a block size of four can consist of repetitions of {111122220000}. Repetitions of {1111000022220000} and {11112222} are other possible patterns. Block designs are good for detection because, at a region that is activated by a particular stimulus type, the lingering effects evoked by stimuli of that type will accumulate to create strong signals. The difference in the signal intensity between activation and non-activation increases, and this helps in identifying activation. Agreeing with this intuition, block designs yield high design efficiencies when the detection problem is the only concern.

The m-sequence-based designs are m-sequences (Barker, 2004; Godfrey, 1993) and designs constructed from m-sequences. These sequences can be constructed from Galois fields or Reed-Muller codes (cf. MacWilliams and Sloane, 1977, Ch. 14), and look rather random with no clear pattern. They only exist if $Q + 1$ is a prime or prime power. The use of these designs for estimating the HRF is first proposed by Buračas and Boynton (2002). Liu and Frank (2004) and Liu (2004) also study these designs. The m-sequence-based designs have high efficiencies for estimation.

Mixed designs, permuted block designs, and clustered m-sequences are studied by Liu and Frank (2004) and Liu (2004) for the case when both estimation and detection are of interest. It is shown that there are designs in these classes that offer advantageous trade-offs between the two competing statistical objectives. A mixed design is formed by concatenating a fraction of a block design with a fraction of an m-sequence (or a random design). By changing

the length of the "blocky" part, and hence that of the "random" part, the resulting designs can move toward having high efficiencies for estimation or high efficiencies for detection. Permuted block designs can be generated by, repeatedly, exchanging positions of two randomly chosen events in a block design. The efficiency for estimation is gradually increased, at the expense of the ability for detection. Clustered m-sequences are created by permuting events in an m-sequence so that the resulting design becomes more "blocky". The design gradually moves toward having a higher efficiency for detection.

4.3   Methodology

4.3.1   Statistical Models

In this section, we specify the underlying model for the two primary statistical objectives, namely estimation and detection. As in Wager and Nichols (2003) and Liu and Frank (2004), two popular linear models are considered (see also Friston et al., 1995; Worsley and Friston, 1995; Dale, 1999):

$$Y = Xh + S\gamma + e, \text{ and} \tag{4.1}$$

$$Y = Z\theta + S\gamma + \eta, \tag{4.2}$$

where Y is the voxel-wise fMRI time series, $h = (h'_1, ..., h'_Q)'$ is the parameter vector for the HRFs of the Q stimulus types, $X = [X_1 \cdots X_Q]$ is the design matrix, $\theta = (\theta_1, ..., \theta_Q)'$ represents the response amplitudes, $Z = Xh_0$ is the convolution of stimuli with an assumed basis, h_0, for the HRF, $S\gamma$ is a nuisance term describing the trend or drift of Y, and e and η are noise. Following Wager and Nichols (2003), we assume a known whitening matrix, V, such that Ve and $V\eta$ are white noise. Model (4.1) is typically used for estimating the HRF and model (4.2) for detecting activation.

Our model formulation, explained in more detail in the remainder of this subsection, follows Kao et al. (2008). The major advantage of this model formulation lies in the use of the discretization interval (Dale, 1999) for parameterizing the HRF. The discretization

enables the use of a finite set of interpretable parameters to capture the fluctuation of the continuous HRF over time. The length of the discretization interval, denoted by ΔT, is set to the greatest value dividing both the ISI and TR. The HRF parameters, captured in the vector $\boldsymbol{h}$, then represent the heights of the HRF for each stimulus after every ΔT seconds following the stimulus onset. This parametrization is explained in the following example.

Example 4.3.1 *In Figure 4.1, we consider one stimulus type ($Q = 1$). The time interval between two consecutive events is 2s (ISI $= 2s$), and that between two successive scans is 3s (TR $= 3s$). An illustrative design is $\xi = \{110100...\}$ with three stimuli taking place at 2s, 4s and 8s, respectively. These three stimuli are presented by vertical bars in the figure, and are followed by curves that represent the evoked HRF. The heights of these three overlapping HRFs accumulate to form the dash-dot curve. This curve represents the noise-free and trend-free BOLD responses induced by the three stimuli; it corresponds to $\boldsymbol{Xh}$ in model (4.1).*

The four vertical lines correspond to the first four MR scans at which the BOLD signal is observed. The heights of the HRFs, or equivalently the effects of the stimuli, that contribute to the observed signal are indicated by the dots on the lines. These heights are different. Therefore, we need different parameters to represent them as well as any other heights that could possibly contribute to a response.

Under the combination of ISI and TR in this example, the heights that could contribute to an observation occur every second on the HRF curve. They are shown as dots and squares on the third curve in Figure 4.1. The reason for the 1 second intervals is that a scan can occur 1s, 2s, 3s, ... after the onset of an event. In general, this time difference must be a multiple of the greatest number dividing both ISI and TR. Setting ΔT to this greatest common divisor, our HRF parameters then describe the discretized HRF, $h((j-1)\Delta T)$, $j = 1, 2, ...$ Here, $h(t)$ is the HRF at time t following the stimulus onset; $t = 0$ corresponds to the stimulus onset. All heights that could possibly make a contribution to an observation are represented with this parametrization. In addition, irrelevant heights that will never contribute to a response are left out.
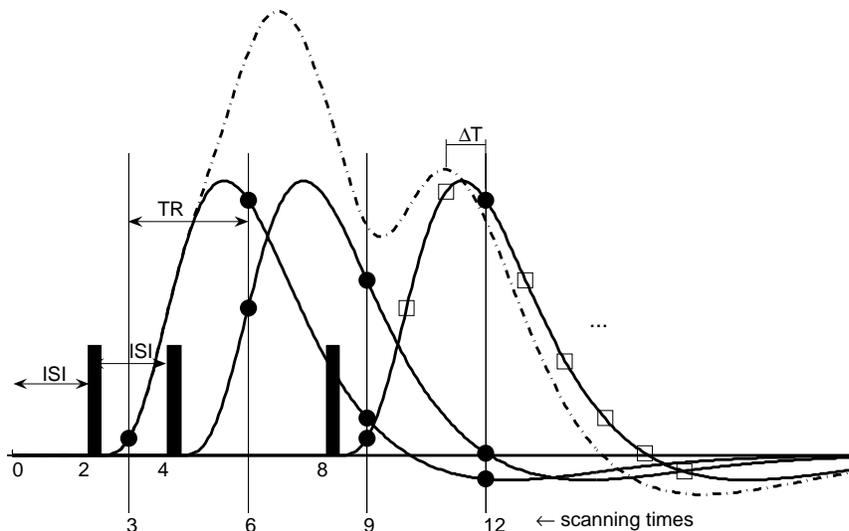
Figure 4.1: HRF parametrization with $Q = 1$, $ISI = 2$s, $TR = 3$s and $\Delta T = 1$s.

The parameter vector $\boldsymbol{h}_i = (h_{1i}, ..., h_{ki})'$ represents the HRF, $h_i(t)$, for the type-i stimulus. With ΔT as defined in Example 4.3.1, we use h_{ji} to denote the height $h_i((j-1)\Delta T)$; $j = 1, ..., k$. Here, the length of $\boldsymbol{h}_i$ is $k = 1 + \lfloor K/\Delta T \rfloor$, where $\lfloor a \rfloor$ is the greatest integer less than or equal to a and K is the duration of the HRF, counting from the stimulus onset to the complete return to baseline.

The matrix $\boldsymbol{X}$ in model (4.1) is determined by both the design sequence and the HRF parametrization. The matrix corresponding to Example 4.3.1 is provided below as an illustration. If the duration of the HRF is 32 seconds then (since $Q = 1$) there are 33 columns. Each column is linked to an h_{j1} and is labeled by $t_j = (j-1)\Delta T$ ($\Delta T = 1$s). Rows are labeled by scanning times, which are multiples of TR (= 3s).

$$
\boldsymbol{X} = \boldsymbol{X}_1 =
\begin{array}{r}
\\
3s \rightarrow \\
6s \rightarrow \\
9s \rightarrow \\
12s \rightarrow \\
\\
\end{array}
\begin{array}{cccccccccccc}
0s & 1s & 2s & 3s & 4s & 5s & 6s & 7s & 8s & 9s & 10s & 11s & \dots \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & \dots \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & \dots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots &
\end{array}
.
$$

While using model (4.2), the same basis function, $\boldsymbol{h}_0$, for the HRFs is assumed for detection for each stimulus type. Throughout this chapter, we consider $\boldsymbol{h}_0$ to be the canonical HRF of the software SPM2 (http://www.fil.ion.ucl.ac.uk/ spm/software/spm2/), which is a combination of two gamma distributions and is scaled to have a maximal value of one. In model (4.2), the matrix $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{h}_0$ represents the convolution of the stimuli with $\boldsymbol{h}_0$; see, e.g., Josephs et al. (1997) for details. The parameter $\boldsymbol{\theta}$, represents the maximal heights for the HRFs evoked by the stimuli. The larger θ_i the more the region is activated by the type-i stimulus.

As stated, the same basis function $\boldsymbol{h}_0$ is assumed for all the Q stimulus types. Only the amplitudes are allowed to vary. However, incorporating different basis functions for different stimuli is also possible in our approach. One can simply take $\boldsymbol{Z} = [\boldsymbol{X}_1 \boldsymbol{h}_0^{(1)} \, \dots \, \boldsymbol{X}_Q \boldsymbol{h}_0^{(Q)}]$, but this setting is beyond the scope of the current work.

4.3.2 Design Criteria

For models (4.1) and (4.2), we consider the parametric functions $\boldsymbol{C}_x \boldsymbol{h}$ and $\boldsymbol{C}_z \boldsymbol{\theta}$, respectively, where

$$
\boldsymbol{C}_x =
\begin{bmatrix}
(1 - \delta_x)\boldsymbol{I}_{Qk} \\
\delta_x \boldsymbol{D}_x
\end{bmatrix}
; \
\boldsymbol{C}_z =
\begin{bmatrix}
(1 - \delta_z)\boldsymbol{I}_Q \\
\delta_z \boldsymbol{D}_z
\end{bmatrix}.
$$

Here, $\boldsymbol{I}_a$ is the a-by-a identity matrix, $\boldsymbol{D}_z$ (with elements of 0 and ± 1) is the $(Q(Q-1)/2)$-by-Q matrix in which the rows correspond to the $Q(Q-1)/2$ pairwise contrasts, $\boldsymbol{D}_x = \boldsymbol{D}_z \otimes \boldsymbol{I}_k$

with $\otimes$ being the Kronecker product, and δ_x, $\delta_z \in [0,1]$. When $\delta_x = \delta_z = 0$, the parametric functions correspond to individual stimulus effects only. When δ_x and δ_z increase, more weight is assigned to pairwise contrasts. With $\delta_x = \delta_z = 1$ we have the case when only pairwise contrasts are of interest. We note that Kao et al. (2008) study the two extreme cases of 0 and 1 separately. Liu and Frank (2004) and Liu (2004) investigated the case when $\delta_x = \delta_z = 1/2$. In this study, we allow the values of δ_x and δ_z to vary between 0 and 1. For simplicity, we consider $\delta_x = \delta_z$. However, the equality between the two coefficients is not required in our approach.

With these parametric functions of interest, we define the design criteria for estimation and detection. For estimation, the design criterion is

$$r_{cx}\left\{trace\left\{\boldsymbol{C}_x\left[\boldsymbol{X}'\boldsymbol{V}'(\boldsymbol{I} - \boldsymbol{P_{VS}})\boldsymbol{V}\boldsymbol{X}\right]^{-1}\boldsymbol{C}'_x\right\}\right\}^{-1}$$

$$= r_{cx}\left\{(1 - \delta_x)^2 trace[\boldsymbol{M}_x^{-1}] + \delta_x^2 trace[\boldsymbol{D}_x\boldsymbol{M}_x^{-1}\boldsymbol{D}'_x]\right\}^{-1},$$

where ξ is the design, $\boldsymbol{M}_x = \boldsymbol{M}_x(\xi) = \boldsymbol{X}'\boldsymbol{V}'(\boldsymbol{I} - \boldsymbol{P_{VS}})\boldsymbol{V}\boldsymbol{X}$ is the information matrix for $\boldsymbol{h}$, $P_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ is the orthogonal projection onto the vector space spanned by the columns of $\boldsymbol{A}$, and r_{cx} is Q for $\delta_x = 0$, $Q(Q-1)/2 + Q$ for $\delta_x \in (0,1)$ and $Q(Q-1)/2$ for $\delta_x = 1$.

Similarly, we define the design criterion for detection as

$$r_{cz}\left\{trace\left\{\boldsymbol{C}_z\left[\boldsymbol{Z}'\boldsymbol{V}'(\boldsymbol{I} - \boldsymbol{P_{VS}})\boldsymbol{V}\boldsymbol{Z}\right]^{-1}\boldsymbol{C}'_z\right\}\right\}^{-1}$$

$$= r_{cz}\left\{(1 - \delta_z)^2 trace[\boldsymbol{M}_z^{-1}] + \delta_z^2 trace[\boldsymbol{D}_z\boldsymbol{M}_z^{-1}\boldsymbol{D}'_z]\right\}^{-1},$$

where $\boldsymbol{M}_z = \boldsymbol{M}_z(\xi) = \boldsymbol{Z}'\boldsymbol{V}'(\boldsymbol{I} - \boldsymbol{P_{VS}})\boldsymbol{V}\boldsymbol{Z}$, and r_{cz} is kQ for $\delta_z = 0$, $kQ(Q-1)/2 + kQ$ for $\delta_z \in (0,1)$ and $kQ(Q-1)/2$ for $\delta_z = 1$. The weights assigned to individual stimulus effects may be thought of as $\lambda_x = (1 - \delta_x)^2/[(1 - \delta_x)^2 + \delta_x^2]$ and $\lambda_z = (1 - \delta_z)^2/[(1 - \delta_z)^2 + \delta_z^2]$, with $1 - \lambda_x$ and $1 - \lambda_z$ being the weights for pairwise contrasts. To explicitly present the dependence of the design criteria on the design and the weights, we use $F_e(\xi; \lambda_x)$ to indicate the criterion for estimation and $F_d(\xi; \lambda_z)$ is for detection.

In general, we define the family of MO-criteria as

$$\left\{ F^* = wF_e^*(\xi; \lambda_x) + (1 - w)F_d^*(\xi; \lambda_z) : w \in [0, 1], \lambda_x, \lambda_z \in [0, 1] \right\}, \qquad (4.3)$$

where (suppressing ξ, λ_x and λ_z)

$$F_j^* = \frac{F_j - \min(F_j)}{\max(F_j) - \min(F_j)}, j = e, d.$$

We use the extreme values of F_d and F_e to normalize these two criteria before combining them. This is done to ensure comparability between the two criteria; see also, Imhof and Wong (2000). The values of w, λ_x and λ_z can be assigned based on the researcher's discretion. After specifying these values, an MO-criterion is well-defined and the search algorithm of Kao et al. (2008) can be applied to search for optimal designs. The search algorithm is based on a genetic algorithm and is briefly introduced in the next section.

4.3.3 SEARCH ALGORITHM

GAs (Holland, 1975; 1992) are popular for solving optimization problems, in which good solutions (parents) are used to generate better ones (offsprings). The GA proposed by Kao et al. (2008) takes advantage of well-known results about good fMRI designs so that the search over the huge design space can be carried out more efficiently. The outline of the algorithm is as follows:

Step 1. (Initial designs) Generate G initial designs consisting of random designs, an m-sequence-based design, a block design and their combinations. Use the objective function to evaluate the fitness of each initial design.

Step 2. (Crossover) With probability proportional to fitness, draw with replacement $G/2$ pairs of designs to crossover — select a random cut-point and exchange the corresponding fractions of the paired designs.

Step 3. (Mutation) Randomly select $q\%$ of the events from the G offspring designs. Replace these events by randomly generated ones.

Step 4. (Immigration) Add to the population another I designs drawn from random designs, block designs and their combinations.

Step 5. (Fitness) Obtain the fitness scores of the offsprings and immigrants.

Step 6. (Natural selection) Keep the best G designs according to their fitness scores to form the parents of the next generation. Discard the others.

Step 7. (Stop) Repeat steps 2 through 6 until a stopping rule is met (e.g., after M generations). Keep track of the best design over generations.

Details about this algorithm can be found in Kao et al. (2008). It is worth noticing that the objective function used in Step 1 and Step 5 can be taken as the design criterion for estimation, or that for detection, or an MO-criterion. To use our MO-criterion, the extreme values of F_e and F_d are required. Theoretical values of $\max(F_e)$ and $\max(F_d)$ are generally not available. They are thus approximated numerically by the GA using the non-standardized function F_e (or F_d) as the objective function. The values of $\min(F_e)$ and $\min(F_d)$ are set to zero, corresponding to designs for which the parametric functions of interest are non-estimable.

When searching for optimal designs, we follow Kao et al. (2008) to use $G = 20$, $q = 1$, $I = 4$ and $M = 10,000$. A larger M does not seem to lead to significantly better designs.

4.4 NUMERICAL RESULTS

We study optimal ER-fMRI designs through a series of numerical simulations. The focus is on investigating the impact of $\lambda = \lambda_x = \lambda_z$, for which we will consider the values of 1, 3/4, 2/3, 1/2, 1/3, 1/4, 1/8, 1/16 and 0. The case $\lambda = 1/2$ is also studied by Liu (2004). The number of stimulus types (Q) considered ranges from 2 to 4. The length of the design is $L = 242$ for $Q = 2$, $L = 255$ for $Q = 3$ and $L = 624$ for $Q = 4$. Under these combinations of Q and L, an m-sequence exists. We will consider two cases for the model. Case I assumes that errors are white noise, i.e., $\boldsymbol{V} = \boldsymbol{I}$, and that $\boldsymbol{S}$ is a vector of ones. For Case II, we

assume a stationary AR(1) error process with a correlation coefficient of 0.3, while $\boldsymbol{S\gamma}$ is taken to be a second-order Legendre polynomial drift. The resulting models for this case are closer to those that are used by fMRI researchers. The ISI and TR are both set to 2 seconds, so that ΔT is also 2 seconds. Under these conditions, we search for optimal designs using our genetic algorithm approach with specified optimality criteria.

4.4.1 DESIGNS FOR DETECTION

Figure 4.2: The designs obtained by our approach for detection with various λ when $Q = 2$. The rows correspond to $\lambda = 1$, 3/4, 1/2, 1/4, 1/16, and 0, respectively. The first column is for Case I and the second column for Case II.

Here, we study the case when only detection is of interest; i.e., $w = 0$ in (4.3). As described in Section 4.2, block designs are expected to be optimal and our approach yields designs with a block structure. Figures 4.2 and 4.3 present these designs for $Q = 2$ and $Q = 3$, respectively. Different shades of grey indicate different stimulus types and white represents the control.

Figure 4.3: The designs obtained by our approach for detection when $Q = 3$. The rows correspond to $\lambda = 1$, 3/4, 1/2, 1/4, 1/16, and 0, respectively. The first column is for Case I and the second column for Case II.

The number above each block is the number of events contained in that block. For example, the top-left design in Figure 4.2 starts with seven controls (pseudo-events), followed by eight stimuli of one type and ends with a stimulus of the other type. We do not present the designs for $Q = 4$ since they provide little additional insight.

From Figures 4.2 and 4.3, the occurrence of the control condition decreases with λ. The range of the frequencies of the stimulus types in our designs is presented in Table 4.1. For both cases, the results agree with the approximated optimal stimulus frequencies that Liu and Frank (2004) derived for Case I.

Table 4.1: Stimulus frequencies (range over different stimulus types) of our designs for detection

| | | | | | λ | | | | |
Q	1	3/4	2/3	1/2	1/3	1/4	1/8	1/16	0
					Case I				
2	0.29-0.30	0.31	0.32	0.33	0.35-0.36	0.36-0.37	0.39-0.40	0.42-0.43	0.50
	(0.30)	(0.31)	(0.32)	(0.33)	(0.36)	(0.37)	(0.40)	(0.42)	(0.50)
3	0.21-0.22	0.24	0.24	0.25	0.26-0.27	0.27	0.29	0.30-0.31	0.33
	(0.21)	(0.23)	(0.24)	(0.25)	(0.26)	(0.27)	(0.29)	(0.30)	(0.33)
4	0.17	0.18-0.19	0.19	0.20	0.21	0.21-0.22	0.22-0.23	0.23	0.25
	(0.17)	(0.18)	(0.19)	(0.20)	(0.21)	(0.22)	(0.23)	(0.23)	(0.25)
					Case II				
2	0.29-0.30	0.31-0.32	0.31-0.32	0.33-0.34	0.36	0.37	0.40	0.42	0.50
3	0.22	0.23-0.24	0.24-0.25	0.25-0.26	0.26-0.27	0.27-0.28	0.29-0.30	0.29-0.31	0.33-0.35
4	0.17	0.18-0.19	0.19	0.20-0.21	0.21	0.21-0.22	0.22-0.23	0.23-0.24	0.25

note. values in parentheses are the approximated optimal stimulus frequencies from Liu and Frank (2004)

In Table 4.2, we present relative efficiencies for the designs that we obtain. For example, for $Q = 3$ under Case I, the entry 93.9 for the row labeled $F_d^*(\xi; \lambda = 1)$ and the column labeled $\xi_{1/2}^d$ indicates that the optimal design for $\lambda = 1/2$, $Q = 3$ and Case I, say $\xi_{1/2}^d$, has an efficiency of 93.9% for detection if $\lambda = 1$. If we denote the optimal design for $\lambda = 1$ by ξ_1^d, this means that

$$\frac{F_d(\xi_{1/2}^d; 1)}{F_d(\xi_1^d; 1)} = 93.9\%$$

Of course, $\xi_{1/2}^d$ has an efficiency of 100% if $\lambda = 1/2$, but this value is not shown in Table 4.2.

From Table 4.2, the optimal designs for pairwise contrasts ($\lambda = 0$) have an efficiency of less than 7% if interest is only in individual stimulus effects ($\lambda = 1$). On the other hand, optimal designs for individual stimulus effects are more robust, with an efficiency of at least 58.8% for pairwise contrasts. The table can also be used to find designs that achieve similar relative efficiencies for the two competing interests. For $Q = 2$, $\xi_{1/8}^d$ achieves similar $F_d^*(\xi; 1)$- and $F_d^*(\xi; 0)$-values. This holds for Case I and Case II. For $Q = 3$ and $Q = 4$, the same is true for $\xi_{1/4}^d$ and $\xi_{1/3}^d$, respectively. Moving away from these λs, we find designs that achieve a higher relative efficiency for one interest than for the other. Note that, equal weight for

Table 4.2: Relative efficiency (percentage) for individual stimulus effects and pairwise contrasts of our designs for detection

Q	Criterion	Designs for Detection								
		ξ_1^d	$\xi_{3/4}^d$	$\xi_{2/3}^d$	$\xi_{1/2}^d$	$\xi_{1/3}^d$	$\xi_{1/4}^d$	$\xi_{1/8}^d$	$\xi_{1/16}^d$	ξ_0^d
						Case I				
2	$F_d^*(\xi; \lambda = 1)$	100.0	99.5	99.0	97.3	94.0	89.7	78.3	64.9	4.9
	$F_d^*(\xi; \lambda = 0)$	58.8	62.2	63.7	66.8	70.7	73.7	79.8	85.2	100.0
3	$F_d^*(\xi; \lambda = 1)$	100.0	98.7	97.4	93.9	87.3	81.6	67.6	56.9	6.1
	$F_d^*(\xi; \lambda = 0)$	63.6	69.8	71.8	75.6	79.7	82.9	88.6	90.1	100.0
4	$F_d^*(\xi; \lambda = 1)$	100.0	97.3	95.5	90.7	83.2	77.6	61.9	47.9	3.0
	$F_d^*(\xi; \lambda = 0)$	68.5	76.1	78.2	81.8	85.4	87.4	91.8	94.1	100.0
						Case II				
2	$F_d^*(\xi; \lambda = 1)$	100.0	99.6	99.3	97.4	93.4	90.4	79.3	65.6	5.5
	$F_d^*(\xi; \lambda = 0)$	59.6	63.1	64.3	67.5	71.6	73.7	79.5	84.8	100.0
3	$F_d^*(\xi; \lambda = 1)$	100.0	98.5	97.5	93.7	88.1	82.7	64.8	55.9	6.7
	$F_d^*(\xi; \lambda = 0)$	64.4	69.9	71.6	75.5	79.0	81.6	84.0	90.6	100.0
4	$F_d^*(\xi; \lambda = 1)$	100.0	98.8	96.8	91.7	84.1	75.5	62.4	48.1	3.6
	$F_d^*(\xi; \lambda = 0)$	68.1	74.2	75.9	80.2	83.7	84.4	88.7	91.6	100.0

individual effects and pairwise contrasts (i.e., $\lambda = 1/2$) does not necessarily yield balanced relative efficiencies for the two interests.

4.4.2 DESIGNS FOR ESTIMATION

In this simulation, we study the optimal designs when the focus is on estimation; i.e., $w = 1$ in (4.3). For this situation, m-sequence-based designs are often recommended. They are included as initial designs of our search algorithm, which is used again to search for optimal designs under the current setting. Table 4.3 presents the $F_e(\xi; \lambda)$-value of our designs relative to the initial m-sequence-based design (in percentage). From the table, the efficiency of the m-sequence-based designs can be improved markedly when λ moves away from $1/2$. When λ

Table 4.3: Efficiency (percentage) for estimation of our designs relative to the m-sequence-based designs

Q	Our Designs for Detection								
	ξ_1^d	$\xi_{3/4}^d$	$\xi_{2/3}^d$	$\xi_{1/2}^d$	$\xi_{1/3}^d$	$\xi_{1/4}^d$	$\xi_{1/8}^d$	$\xi_{1/16}^d$	ξ_0^d
	Case I								
2	101.8	100.3	100.2	100.2	101.9	103.9	109.6	116.8	146.6
3	104.9	101.0	100.3	100.5	102.0	103.8	109.3	114.2	131.8
4	108.9	101.3	100.3	100.0	101.0	102.3	106.4	110.3	122.3
	Case II								
2	106.5	105.3	105.1	105.5	107.6	109.9	116.9	125.1	157.3
3	109.7	106.2	105.8	106.2	108.6	110.8	116.7	122.2	141.6
4	112.6	105.1	104.2	103.7	105.5	107.2	111.7	115.8	128.8

is close to $1/2$, it is hard for Case I to find designs that are better than the m-sequence-based designs. In that particular situation, the stimulus frequencies of the m-sequence-based designs are close to the approximated optimal stimulus frequencies of Liu and Frank (2004). Our designs also yield similar frequencies in that situation. For other values of λ, our approach yields designs with higher efficiencies and the stimulus frequencies of the resulting designs are consistently in good agreement with the optimal stimulus frequencies.

It is known that the m-sequence-based designs can be suboptimal under correlated noise and the appearance of drift; see, e.g., Kao et al. (2008) and Buračas and Boynton (2002). In contrast, our approach can adapt to these situations and lead to better designs. This is reflected in Table 4.3 for Case II. Even for $\lambda = 1/2$, when the stimulus frequency of our designs is similar to that of the m-sequence-based designs, the pattern of our designs beats that of the m-sequence-based designs.

Table 4.4: Relative efficiency (percentage) for individual stimulus effects and pairwise contrasts of our designs for estimation

Q	Criterion	Designs for Estimation								
		ξ_1^e	$\xi_{3/4}^e$	$\xi_{2/3}^e$	$\xi_{1/2}^e$	$\xi_{1/3}^e$	$\xi_{1/4}^e$	$\xi_{1/8}^e$	$\xi_{1/16}^e$	ξ_0^e
						Case I				
2	$F_e^*(\xi; \lambda = 1)$	100.0	99.1	98.9	97.2	94.4	89.9	77.4	67.3	10.6
	$F_e^*(\xi; \lambda = 0)$	61.5	65.7	66.4	69.4	73.0	76.8	83.3	87.4	100.0
3	$F_e^*(\xi; \lambda = 1)$	100.0	98.2	96.2	94.6	87.9	83.7	71.0	56.6	11.6
	$F_e^*(\xi; \lambda = 0)$	66.8	72.5	75.2	77.2	81.5	83.6	88.6	92.1	100.0
4	$F_e^*(\xi; \lambda = 1)$	100.0	97.3	95.1	91.5	83.0	76.5	62.4	48.4	11.6
	$F_e^*(\xi; \lambda = 0)$	68.7	75.7	78.3	81.6	85.6	87.8	91.7	94.7	100.0
						Case II				
2	$F_e^*(\xi; \lambda = 1)$	100.0	99.1	99.1	97.4	92.8	88.3	79.8	65.7	13.9
	$F_e^*(\xi; \lambda = 0)$	61.4	65.5	65.5	69.2	74.5	78.1	82.6	88.2	100.0
3	$F_e^*(\xi; \lambda = 1)$	100.0	98.3	97.3	93.6	86.6	82.0	68.4	57.7	12.9
	$F_e^*(\xi; \lambda = 0)$	66.2	71.8	73.4	77.5	82.1	84.3	89.0	91.5	100.0
4	$F_e^*(\xi; \lambda = 1)$	100.0	97.6	95.7	90.2	81.9	76.5	61.5	48.7	8.0
	$F_e^*(\xi; \lambda = 0)$	68.5	75.2	77.4	81.4	85.8	87.8	91.9	94.4	100.0

Table 4.4, which is to be read in a similar way as Table 4.2, presents the relative efficiency achieved by our designs. Similar as for detection, ξ_1^e is more robust to a change in interests than ξ_0^e. Designs $\xi_{1/8}^e$ for $Q = 2$, $\xi_{1/4}^e$ for $Q = 3$ and $\xi_{1/16}^e$ for $Q = 4$ provide similar relative efficiencies for the two interests.

4.4.3 MULTI-OBJECTIVE DESIGNS

In this simulation, we allow the weight w to increase from 0 to 1 in steps of 0.05, thereby, gradually shifting emphasis from the detection problem to the estimation problem. For each Q and λ and for both cases, 21 optimal designs are obtained (one for each w-value). Figure 4.4 presents the $F_e^*(\xi; \lambda)$- versus $F_d^*(\xi; \lambda)$-values of the resulting designs for $Q = 2$ and Case

I. Similar figures for other settings provide no further insight and are therefore omitted. The designs introduced in Section 4.2 are also presented in the figure. The block design in Figure 4.4 is the initial block design of our search algorithm. Mixed designs are generated by combining a block design (our initial design) and an m-sequence-based design via crossover (Step 2 of the search algorithm in Subsection 4.3.3). This m-sequence-based design along with permuted block designs and clustered m-sequences are generated from the program provided by Liu (2004). For the permuted block designs, we choose a block design with a block size of eight since other block sizes do not seem to yield better results. As demonstrated in Liu (2004), for Case I and $\lambda = 1/2$, these designs provide an advantageous trade-off between the two statistical objectives, estimation and detection. Our designs provide a better trade-off, even for that special case.

We note that the designs considered by Liu (2004) have a fixed stimulus frequency of $1/(Q+1)$. They can be good when $\lambda = 1/2$ because this is the optimal stimulus frequency for that case. However, when moving away from $\lambda = 1/2$, these designs are sub-optimal. Our approach finds much better designs as shown in Figure 4.4.

Table 4.5 presents means and standard deviations of the stimulus frequencies of our designs, which are computed over the 21 designs generated for each setting and the different stimulus types. Small standard deviations show that the stimulus frequencies vary little over the designs and the stimulus types. Again, these frequencies agree with the approximated optimal stimulus frequencies in Table 4.1.

4.5 Discussion and Conclusions

In this chapter, we study optimal ER-fMRI designs when both individual stimulus effects and pairwise contrasts are of interest. These two interests are among the main concerns of fMRI researchers. Therefore, when planning ER-fMRI designs, it is crucial to find designs that are efficient for both interests.
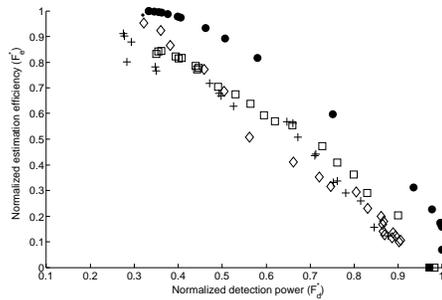
$\lambda = 1$
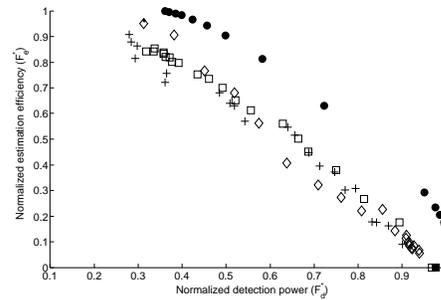
$\lambda = 1/4$
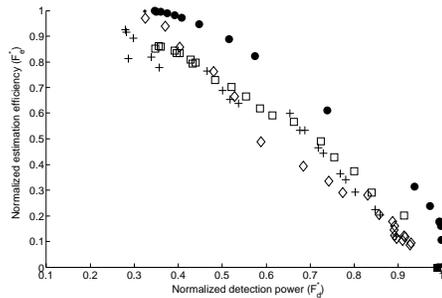
$\lambda = 3/4$
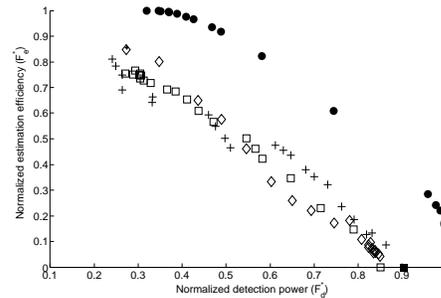
$\lambda = 1/16$
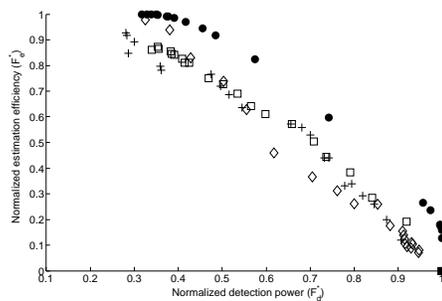
$\lambda = 1/2$
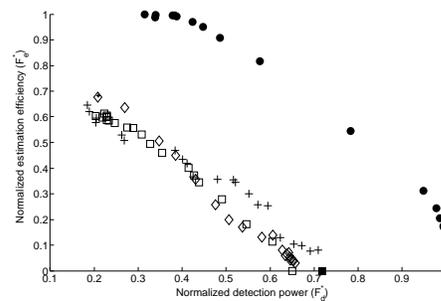
$\lambda = 0$

Figure 4.4: $F_e^*(\xi; \lambda)$-values versus $F_d^*(\xi; \lambda)$-values of designs obtained for Case I with $Q = 2$. •: designs found by our approach; *: m-sequence; ▪: block design; ◇: clustered m-sequences; □: permuted block designs; +: mixed designs;

Table 4.5: Stimulus frequencies (mean and standard deviation) of our designs

Q				λ				
	1	3/4	2/3	1/2	1/3	1/4	1/8	1/16	0
					Case I				
2	0.302	0.318	0.324	0.339	0.360	0.370	0.400	0.425	0.494
	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
3	0.218	0.234	0.242	0.253	0.267	0.276	0.292	0.302	0.330
	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)
4	0.169	0.187	0.192	0.201	0.210	0.216	0.226	0.232	0.249
	(0.002)	(0.002)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
					Case II				
2	0.298	0.316	0.323	0.338	0.360	0.373	0.400	0.426	0.495
	(0.005)	(0.007)	(0.006)	(0.006)	(0.007)	(0.007)	(0.005)	(0.007)	(0.007)
3	0.218	0.234	0.242	0.253	0.267	0.276	0.292	0.303	0.330
	(0.006)	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)	(0.006)	(0.007)
4	0.169	0.187	0.192	0.202	0.211	0.216	0.226	0.233	0.249
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)

note. the mean and standard deviation are taken over the Q stimulus types and the 21 designs

Previous work either considers these two interests separately (Kao et al., 2008) or assigns equal weights to them (Liu and Frank, 2004; Liu, 2004). In contrast, we propose an approach for finding optimal designs allowing user-specified weights.

In our numerical results, we find (near-)optimal designs for each λ. The stimulus frequencies in these designs increase when λ decreases. This also means that the frequency of the control decreases when λ decreases. This phenomenon is expected from the approximated optimal stimulus frequencies derived by Liu and Frank (2004). Their approximation is derived for white noise and with neither drift nor trend, that is for our Case I. Our designs have stimulus frequencies that are in good agreement with the approximated optimum, not only for Case I but also for Case II.

Our numerical results show that the choice of $\lambda = 1/2$ does not necessarily yield a design that achieves similar relative efficiencies for the two interests. The values of λ that achieve this objective are, approximately, $\lambda = 1/8$ for $Q = 2$, $\lambda = 1/4$ for $Q = 3$, and $\lambda = 1/3$ for $Q = 4$.

We also observe that designs for individual stimulus effects retain a reasonable efficiency for pairwise contrasts and that they are relatively robust with respect to a change in interests. On the other hand, designs that are optimal for pairwise contrasts can have low efficiencies for estimating individual stimulus effects. These designs should not be used unless pairwise contrasts are the only concern.

Working with our defined MO-criteria, the search algorithm of Kao et al. (2008) can be applied for finding multi-objective optimal designs when both individual stimulus effects and pairwise contrasts are of interest. For detecting activation, the algorithm yields designs with a block structure. For estimating the HRF, we can find designs that work much better than the m-sequence-based designs. When considering these statistical objectives simultaneously, we find designs that provide advantageous trade-offs between the two competing objectives. The algorithm can accommodate user-specified weights for the two objectives.

In addition to statistical objectives, Kao et al. (2008) also consider psychological constraints, and customized requirements when finding multi-objective optimal designs. It is straightforward to include these additional objectives in our family of MO-criteria and the search algorithm of Kao et al. (2008) can still be used for finding optimal designs. However, for the sake of clarity, we have only focused on statistical objectives in this study.

4.6 References

Amaro, E. J. and Barker, G. J. (2006), "Study design in MRI: Basic principles," *Brain and Cognition*, 60, 220–232.

Barker, H. A. (2004), "Primitive maximum-length sequences and pseudo-random signals," *Transactions of the Institute of Measurement & Control*, 26, 339–348.

Buračas, G. T. and Boynton, G. M. (2002), "Efficient Design of Event-Related fMRI Experiments Using M-Sequences," *NeuroImage*, 16, 801–813.

Culham, J. C. (2006), "Functional neuroimaging: Experimental design and analysis," in *Handbook of Functional Neuroimaging of Cognition*, eds. Cabeza, R. and Kingstone, A., Cambridge, Massachusetts: MIT Press, 2nd ed., pp. 53–82.

Dale, A. M. (1999), "Optimal experimental design for event-related fMRI," *Humman Brain Mapping*, 8, 109–114.

Donaldson, D. I. and Buckner, R. L. (2002), "Effective Paradigm Design," in *Functional MRI: An Introdcution to Methods*, eds. Jezzard, P., Matthews, P. M., and Smith, S. M., New York: Oxford University Press, pp. 177–195.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., and Turner, R. (1995), "Analysis of fMRI Time-Series Revisited," *NeuroImage*, 2, 45–53.

Godfrey, K. (1993), *Perturbation signals for system identification*, New York: Prentice Hall.

Holland, J. H. (1975), *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, Ann Arbor: University of Michigan Press.

— (1992), *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*, Complex adaptive systems, Cambridge, Massachusetts: MIT Press, 1st ed.

Imhof, L. and Wong, W. K. (2000), "A graphical method for finding maximin efficiency designs," *Biometrics*, 56, 113–117.

Josephs, O. and Henson, R. N. A. (1999), "Event-related functional magnetic resonance imaging: modelling, inference and optimization," *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 354, 1215–1228.

Josephs, O., Turner, R., and Friston, K. (1997), "Event-related fMRI," *Human Brain Mapping*, 5, 243–248.

Kao, M.-H., Mandal, A., Lazar, N., and Stufken, J. (2008), "Multi-objective Optimal Experimental Designs for Event-Related fMRI Studies," (Submitted).

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

Liu, T. T. and Frank, L. R. (2004), "Efficiency, power, and entropy in event-related FMRI with multiple trial types: Part I: theory," *NeuroImage*, 21, 387–400.

MacWilliams, F. J. and Sloane, N. J. A. (1977), *The theory of error correcting codes*, North-Holland Pub. Co.

Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998), "Event-related functional MRI: Past, present, and future," *PNAS*, 95, 773–780.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of experimental design in fMRI: a general framework using a genetic algorithm," *NeuroImage*, 18, 293–309.

Worsley, K. J. and Friston, K. J. (1995), "Analysis of fMRI Time-Series Revisited–Again," *NeuroImage*, 2, 173–181.

Efficient Designs for Event-Related Functional Magnetic Resonance Imaging with Multiple Scanning Sessions[1]

5.1 INTRODUCTION

ER-fMRI is one of the leading technologies that uses an ultra-fast MR scanner to study human brain activity in response to brief mental stimuli or tasks (e.g., looking at pictures or tapping fingers). This pioneering technology is popular in both medical practice and scientific research, and is arguably the most important advance in neuroscience; see Rosen et al. (1998), Josephs and Henson (1999) and Culham (2006) for overviews of ER-fMRI.

An ER-fMRI design is a sequence of stimuli of one or more types interlaced with a control condition (e.g., rest or fixation). Without a carefully selected sequence, data collected can easily fail to provide sufficient information for medical investigations or for scientific questions of interest. Designs helping to efficiently render valid and precise statistical inference are therefore very important to avoid wasting precious resources.

However, obtaining good designs for ER-fMRI is an arduous task; it requires careful consideration of statistical models, study objectives, experimental settings and assumptions, and certain practical issues. In addition, the design space, which contains all possible designs, is enormous (e.g. Liu, 2004). Searching over this huge space for good designs is difficult. Moreover, the complex nature of ER-fMRI impels the need for considering multiple, competing objectives in one study. Obtaining designs that efficiently achieve these objectives is necessary, but it increases the complexity of the design problem. Furthermore, answers will hinge on the researcher's interests and experimental conditions. Therefore, we need an efficient, versatile approach that can accommodate a variety of situations to obtain designs best suited to the researcher's needs.

Kao et al. (2009a) propose an approach to search for efficient ER-fMRI designs. Their approach includes rigorous formulations of statistical models, well-defined multi-objective design criteria, and a search algorithm that incorporates knowledge about the performance of well-known ER-fMRI designs. This approach is shown to outperform methods known hitherto. Designs obtained by their approach can attain much higher efficiencies than designs that had been widely used by practitioners.

While the approach of Kao et al. (2009a) can efficiently yield good designs, they only study experiments with one scanning session. The same applies to other methods in the literature (e.g. Liu and Frank, 2004; Wager and Nichols, 2003). In practice, it can be hard for a subject to maintain a satisfactory performance throughout a long scanning session, and using multiple short sessions is therefore not uncommon. For example, Wang et al. (2007) consider two scanning sessions, each lasting six minutes and 24 seconds; subjects can rest between sessions. See, e.g., Brown et al. (2008), Harms et al. (2005) and Harms and Melcher (2002) for other examples. Taking this practical issue into account at the design stage is crucial, but, to our knowledge, there is no previous work that systematically studies this.

In this chapter, we obtain efficient designs for experiments with multiple scanning sessions. We assume no "pre-scanning" periods so that the stimuli are presented to an experimental subject only within each scanning session; see Section 5.5 for a further discussion. Statistical models, which are natural extensions of widely used models, are formulated, and design criteria for evaluating competing designs are defined. The search algorithm of Kao et al. (2009a) is applied to find designs optimizing these criteria. This approach is compared to various alternative methods. In our simulations, we demonstrate that these alternative methods can perform well for some cases while they can be bad for others. We also propose an algorithm that is restricted to a subclass of designs. The good performance of this algorithm indicates that efficient designs can be found more efficiently over this constraint class.

In the following section, we briefly introduce ER-fMRI designs. Statistical models, design criteria and approaches for obtaining efficient designs are presented in Section 5.3. Simulation results are provided in Section 5.4 and conclusions and a discussion are in Section 5.5.

5.2 Background

Before conducting an ER-fMRI experiment, a design sequence consisting of stimuli and the control is prepared. This sequence is presented to an experimental subject while the MR

scanner scans his/her brain every few seconds. The time interval between successive MR scans is called time-to-repetition or TR. A blood oxygenation level dependent (BOLD) time series is collected by the MR scanner at each brain voxel (a small region of the brain). This time series indicates the fluctuations of the MR signal intensity, which are linked to the change in the ratio of oxygenated to deoxygenated hemoglobin; see, e.g., Lazar (2008) and Cabeza and Kingstone (2006) for details. The BOLD time series is then used to estimate the hemodynamic response function (HRF; a function describing the change of signal intensity over time that is evoked by a single, brief stimulus) and detect brain activation. Estimation of the HRF and detection of brain activity are two common statistical goals for ER-fMRI. We want an optimal sequence of stimuli so that statistical inference (related to estimation and detection) is most efficient, in some sense.

Following convention, designs are presented as a finite sequence of integers. They look like $\xi = \{101201210...1\}$, where 0 represents the control and q a type-q stimulus; $q = 1, 2, ..., Q$ (= total number of stimulus types). When being presented to an experimental subject, each stimulus (e.g. a picture) lasts for a short period of time (several milliseconds) relative to the inter-stimulus interval (ISI). The ISI is the fixed time interval between the onsets of consecutive events; an event can be a stimulus or control. We note that 0s in the sequence are "pseudo-events". They help to calculate the onset times of the stimuli. For example, with a 0 in the second position of ξ, the first three stimuli (1, 1, and 2) occur at 1ISI, 3ISI, and 4ISI seconds after the outset of the experiment, respectively. The control fills up the time period from the end of a stimulus to the start of the next one.

Well known ER-fMRI designs include block designs, m-sequence-based designs, and mixed designs. In ER-fMRI, a block design is a sequence where stimuli of the same type appear in clusters. For example, a block design can consist of repetitions of $\{111122220000\}$. This design has two stimulus types and the block size is four. Repetitions of $\{1111000022220000\}$ and of $\{11112222\}$ are other possible patterns. Block designs are known to be good for detecting activation.

The m-sequence-based designs are m-sequences (Barker, 2004; Godfrey, 1993) and designs constructed from m-sequences. The m-sequences can be generated from Galois fields or Reed-Muller codes (cf. Ch. 14 of MacWilliams and Sloane, 1977) and only exist if $Q+1$ is a prime or prime power. They look rather random with no clear pattern, but are not easy to achieve through a random mechanism (Buračas and Boynton, 2002). In terms of A-optimality (see Section 5.3.2), these designs yield high efficiencies when the BOLD time series is assumed to be uncorrelated, exhibits neither drift nor trend, and the objective of the experiment combines interest in estimating individual HRFs and comparing HRFs for different stimulus types. Except for this particular case, m-sequence-based designs can be significantly outperformed by designs obtained by the approach of Kao et al. (2009a), as demonstrated in that paper.

While block designs are good for detection problems, they can be very inefficient for estimating the HRFs. Conversely, designs having high estimation efficiency can perform poorly in detecting activation. There is a trade-off between these two statistical goals; see, e.g., Liu and Frank (2004), Liu et al. (2001) and Buxton et al. (2000). An intermediate efficiency can be achieved by a mixed design. A design of this kind can be formed by concatenating a fraction of a block design with a fraction of an m-sequence-based design (or a random design). By changing the length of the "blocky" part, and hence that of the "random" part, the resulting designs can move toward having high efficiencies for estimation or high efficiencies for detection.

These well known designs are incorporated in the search algorithm of Kao et al. (2009a) to efficiently and effectively obtain good ER-fMRI designs for a single scanning session. Their algorithm is briefly introduced in Section 5.3.3. This algorithm is flexible enough to search for efficient designs for multiple scanning sessions. In the next section, we introduce a variation on popular linear models to take session effects into account. We also describe the design criteria and approaches for finding efficient designs under theses criteria.

5.3 METHODOLOGY

5.3.1 MODELS

Linear models are popular for modeling the BOLD time series and play an important role in studying designs for ER-fMRI (e.g. Kao et al., 2009a; Liu and Frank, 2004; Liu, 2004; Wager and Nichols, 2003; Liu et al., 2001; Dale, 1999; Friston et al., 1999). However, these models do not take into account the reality that experiments commonly involve multiple scanning sessions. We therefore extend the widely used models for estimation and detection by including session effects. This results in models (5.1) and (5.2) below for estimation and detection, respectively.

$$
\mathbf{Y} = \begin{bmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \\ \vdots \\ \mathbf{X}^{(B)} \end{bmatrix} \mathbf{h} + [\mathbf{I}_B \otimes \mathbf{S}] \begin{bmatrix} \boldsymbol{\gamma}^{(1)} \\ \boldsymbol{\gamma}^{(2)} \\ \vdots \\ \boldsymbol{\gamma}^{(B)} \end{bmatrix} + \mathbf{e};
\tag{5.1}
$$

$$
\mathbf{Y} = \begin{bmatrix} \mathbf{Z}^{(1)} \\ \mathbf{Z}^{(2)} \\ \vdots \\ \mathbf{Z}^{(B)} \end{bmatrix} \boldsymbol{\theta} + [\mathbf{I}_B \otimes \mathbf{S}] \begin{bmatrix} \boldsymbol{\gamma}^{(1)} \\ \boldsymbol{\gamma}^{(2)} \\ \vdots \\ \boldsymbol{\gamma}^{(B)} \end{bmatrix} + \boldsymbol{\eta},
\tag{5.2}
$$

where $\mathbf{Y}$ is a $T \times 1$ vector of a BOLD time series from a brain voxel, $\mathbf{h} = (\mathbf{h}'_1, ..., \mathbf{h}'_Q)'$ is the parameter vector for the HRFs of the Q stimulus types, $\mathbf{X}^{(b)} = [\mathbf{X}_1^{(b)} \cdots \mathbf{X}_Q^{(b)}]$ is the design matrix for the bth session, $b = 1, ..., B$ (=the number of scanning sessions), $\mathbf{X}_q^{(b)}$ is the design matrix for the qth stimulus type in that session, $\boldsymbol{\theta} = (\theta_1, ..., \theta_Q)'$ represents the response amplitudes, $\mathbf{Z}^{(b)} = [\mathbf{X}_1^{(b)} \mathbf{h}_0 \cdots \mathbf{X}_Q^{(b)} \mathbf{h}_0]$ is the convolution of stimuli with an assumed basis, $\mathbf{h}_0$, of the HRF, $\mathbf{S}\boldsymbol{\gamma}^{(b)}$ is a nuisance term describing the trend or drift of $\mathbf{Y}$ within the bth session, $\mathbf{I}_B$ is the $B \times B$ identity matrix, and $\mathbf{e}$ and $\boldsymbol{\eta}$ are noise. A known whitening matrix, $(\mathbf{I}_B \otimes \mathbf{V})$, is assumed so that $(\mathbf{I}_B \otimes \mathbf{V})\mathbf{e}$ and $(\mathbf{I}_B \otimes \mathbf{V})\boldsymbol{\eta}$ are white noise;

i.e., $E\{(\boldsymbol{I}_B \otimes \boldsymbol{V})\boldsymbol{e}\} = E\{(\boldsymbol{I}_B \otimes \boldsymbol{V})\boldsymbol{\eta}\} = \boldsymbol{0}$, $Cov\{(\boldsymbol{I}_B \otimes \boldsymbol{V})\boldsymbol{e}\} = Cov\{(\boldsymbol{I}_B \otimes \boldsymbol{V})\boldsymbol{\eta}\} = \sigma^2 \boldsymbol{I}$, where $\sigma^2 > 0$ is unknown. We assume that each session is of the same length, so that the $\boldsymbol{X}^{(b)}$s have the same number of rows. Moreover, by using the same $\boldsymbol{S}$, the trend (or drift) in each session is assumed to be of the same functional shape, but the parameters, $\boldsymbol{\gamma}^{(b)}$s, are allowed to be different.

Note that, when $B = 1$ these models reduce to the ones considered in Kao et al. (2009a). For estimation problems, $\boldsymbol{h}$ is the parameter of interest, whereas detection problems focus on studying $\boldsymbol{\theta}$. In general, estimable parametric functions $\boldsymbol{C}_x \boldsymbol{h}$ and $\boldsymbol{C}_z \boldsymbol{\theta}$ are investigated. We omit details regarding the construction of each design matrix $\boldsymbol{X}^{(b)}$ and model parametrization since this is identical as in Kao et al. (2009a); see also Kao et al. (2009b).

5.3.2 DESIGN CRITERIA

To evaluate the quality of a design, we consider the A-optimality criterion (e.g. Atkinson et al., 2007); A-optimality aims at minimizing the average variance of estimators of parametric functions $\boldsymbol{C}_x \boldsymbol{h}$ and $\boldsymbol{C}_z \boldsymbol{\theta}$. We formulate these design criteria as "larger-the-better" criteria, which have the form of:

$$r_c \left\{ trace \left\{ \boldsymbol{C} \left[\boldsymbol{W}'(\boldsymbol{I}_B \otimes \boldsymbol{V}')(\boldsymbol{I}_T - \boldsymbol{P}_{(\boldsymbol{I}_B \otimes \boldsymbol{V})(\boldsymbol{I}_B \otimes \boldsymbol{S})})(\boldsymbol{I}_B \otimes \boldsymbol{V})\boldsymbol{W} \right]^{-} \boldsymbol{C}' \right\} \right\}^{-1}$$

$$= r_c \left\{ trace \left\{ \boldsymbol{C} \left[\sum_{b=1}^{B} \boldsymbol{W}^{(b)'} \boldsymbol{V}'(\boldsymbol{I}_{T_B} - \boldsymbol{P}_{\boldsymbol{V}\boldsymbol{S}})\boldsymbol{V}\boldsymbol{W}^{(b)} \right]^{-} \boldsymbol{C}' \right\} \right\}^{-1}.$$

Here, $\boldsymbol{W} = [\boldsymbol{W}^{(1)'} \cdots \boldsymbol{W}^{(B)'}]'$ with $\boldsymbol{W}^{(b)} \equiv \boldsymbol{X}^{(b)}$ for estimation problems, and $\boldsymbol{W}^{(b)} \equiv \boldsymbol{Z}^{(b)}$ for detecting activation; $T_B = T/B$ and $P_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'$ is the orthogonal projection matrix onto the vector space spanned by the column vectors of $\boldsymbol{A}$, $\boldsymbol{A}^{-}$ is a generalized inverse matrix of $\boldsymbol{A}$, $\boldsymbol{C}$ is a matrix of linear combinations of the parameters, and r_c is the number of rows of $\boldsymbol{C}$.

We denote this design criterion for estimation by F_e and for detection by F_d. The value of F_e is referred to as "estimation efficiency" and the F_d-value is called "detection power". It is not uncommon to consider both of these statistical objectives in one experiment. As

in previous studies, we consider the following family of multi-objective design criteria (MO-criteria):

$$\{F^* = wF_d^* + (1-w)F_e^*; w \in [0,1]\}, \tag{5.3}$$

where w is a weight selected based on the researcher's emphasis on the detection problem and F_d^* and F_e^* are standardized criteria:

$$F_i^* = \frac{F_i}{\max(F_i)}, i = d, e.$$

The standardization is used to ensure scale comparability between the two criteria. Note that $\min(F_i) = 0$, which corresponds to the case when some of the parametric functions are non-estimable; see also Kao et al. (2009a).

For a given w, we consider the approaches introduced in the following section to search for designs optimizing F^*.

5.3.3 SEARCHING FOR OPTIMAL DESIGNS

KNOWLEDGE-BASED GENETIC ALGORITHMS

The algorithm of Kao et al. (2009a) for searching for efficient ER-fMRI designs is based on the genetic algorithm (GA) technique. GAs (Holland, 1975; 1992) are popular for solving optimization problems, in which good solutions (parents) are used to generate better ones (offsprings). The GA proposed by Kao et al. (2009a) takes advantage of well known results about good ER-fMRI designs so that the search over the huge design space can be carried out more efficiently.

Their algorithm starts with G initial designs consisting of random designs, an m-sequence-based design, a block design and mixed designs. With probability proportional to fitness (in terms of the objective function), $G/2$ pairs of designs are then selected to produce G offsprings through crossover — randomly selecting a cut-point and exchanging the corresponding fractions of the paired designs. A portion, α_m, of the combined events from the G

offspring designs are randomly selected to mutate; these selected events are replaced by randomly generated ones. To help escape local optima and to provide building blocks (or good traits), Kao et al. (2009a) add to the GA population another $\alpha_I \times G$ (α_I is a small fraction) designs drawn from random designs, block designs and mixed designs. The enlarged population is then pruned to maintain a constant population size of G. Only the fittest survive to the next generation. The best solution is kept track of along generations and the process is repeated until a stopping rule is met, e.g., until a pre-specified number of generations has been reached.

The approach for finding efficient designs for our current problem is to use the GA search of Kao et al. (2009a) under the models in that paper (i.e., without session effects) and simply split the resulting design into B "short designs", each corresponding to one session. We will denote this approach by GA-L1. A second possibility is to use the same GA to search for models (5.1) and (5.2), so that session effects are included. We denote this approach by GA-L2.

As a third approach, we propose a new GA search over a smaller design space. The idea of this algorithm is to find a "short design" for the first session, and obtain designs for the other sessions by permuting the stimulus types. While other permutations can also be considered, we focus on cyclic permutations. The cyclic permutation is easy to implement and, in our experience, it yields efficient designs. The design for the bth session is constructed by replacing the symbol q in the short design for the first session by $q + b - 1 \ (mod \ Q)$ (we use Q to represent a residue of zero), $q = 1, 2, ..., Q$, $b = 1, 2, ..., B$. For example, with $Q = 3$ stimulus types of 1, 2 and 3 in the design for the first session are replaced, in order, by 2, 3, and 1 for the second session. By juxtaposing the short design and its permuted versions, we obtain a "juxtaposed design" for the entire experiment.

Our algorithm searches for a short design that yields a juxtaposed design maximizing the MO-criterion F^*. We follow Kao et al. (2009a) to incorporate well known ER-fMRI designs. To allow a more effective use of knowledge-based immigrants, we replace the worst parents

by immigrants before selecting good parents into the mating pool. This increases the chance to make good use of potential building blocks supplied by immigrants. Our algorithm, which is referred to as GA-L3, is detailed below.

Step 1. (Initial designs) For the first session, generate G initial short designs consisting of random designs, an m-sequence-based design, a block design and mixed designs.

Step 2. (Permutation) For each short design, cyclically permute the stimulus types to create $B - 1$ additional short designs. Form the juxtaposed design and calculate its fitness (e.g., F_e-, F_d- or F^*-values).

Step 3. (Immigration) Replace the worst $\alpha_I \times G$ designs in the current generation by immigrants drawn from random designs, block designs and mixed designs. Here, α_I is a small fraction. Calculate the fitness of these immigrants as in Step 2.

Step 4. (Crossover) With probability proportional to fitness, select with replacement G designs and form $G/2$ pairs of distinct parents. Use single-point crossover to generate G offspring designs. That is, randomly select a position in the design sequence and exchange the sub-sequences prior to the selected position of paired parents.

Step 5. (Mutation) Randomly select a portion α_m of the combined events from the G offspring designs. Replace these events by randomly generated ones. As in Step 2, obtain the fitness scores of the resulting G offsprings.

Step 6. (Natural selection) Out of the $2G$ designs from steps 3 and 5, keep the best G designs according to their fitness to form the next generation. Discard the others.

Step 7. (Stop) Repeat steps 3 through 6 until a stopping rule is met (e.g., after M_g generations). Keep track of the best design over generations.

As do Kao et al. (2009a), m-sequence-based designs or m-sequences are generated following Liu (2004). We include the m-sequence-based design that yields the highest estimation

efficiency as one of the initial designs. When these designs are not available, a random design will be used instead.

The initial block design has the highest detection power among designs of different numbers of blocks and of two different patterns. In this pool of candidate block designs, the number of blocks for each stimulus type ranges among one to five, ten, 15, 20, 25, 30, and 40, whenever possible. The two patterns include repetitions of NABC and NANBNC, where N is a block of rests and A, B and C represent blocks of stimuli of different types.

The combination of a block design with an m-sequence-based design or a random design is obtained through crossover. These mixed designs constitute a portion, e.g., one-third, of the initial designs. The remaining initial designs are formed by random designs.

In Step 3, each immigrant consists of a portion of a block design and a portion of a random design; the relative length of the two parts is randomly decided. When the length for the random part is too short (e.g., less than 10 events), a block design is used. Similarly, an immigrant can be a random design. The block-design portion is randomly selected to have one to ten blocks (as possible) and one of the two aforementioned patterns.

When searching for efficient designs, we follow Kao et al. (2009a) to use G (size of a generation) = 20, α_m (rate of mutation) = 0.01, α_I (proportion of immigrants) = 0.2 and stops after $M_g = 10,000$ generations. A larger M_g does not seem to lead to significantly better designs.

Note that, to use an MO-criterion as the objective function in the algorithm, the maxima of F_e and F_d are needed. Theoretical values of $\max(F_e)$ and $\max(F_d)$ are generally not available. They are approximated numerically by the GA using the non-standardized functions F_e and F_d as the objective functions, respectively.

A MATLAB program implementing this algorithm can be found in http://www.stat.uga.edu/~amandal/.

HEURISTIC APPROACHES

Some heuristic approaches can also be used. These approaches might provide "short-cuts" for finding efficient designs for the entire experiment. The idea is to find an efficient short design for a single session and, in some way, combine B such designs for B sessions. More specifically, we apply the GA of Kao et al. (2009a) to find a short design that maximizes the design criteria for a single session, namely (for estimation and detection):

$$r_c \left\{ trace \left\{ \boldsymbol{C} \big[\boldsymbol{W}^{(1)'} \boldsymbol{V}' (\boldsymbol{I}_{T_B} - \boldsymbol{P_V} \boldsymbol{S}) \boldsymbol{V} \boldsymbol{W}^{(1)} \big]^{-} \boldsymbol{C}' \right\} \right\}^{-1}.$$

The following approaches are then considered for a design for the B sessions:

GA-S1: Use the same short design for each session.

GA-S2: Use different designs for different sessions; these designs, with similar efficiencies, are generated by multiple runs of the GA.

GA-S3: Permute the stimulus types of the design for the first session to obtain designs for other sessions.

For GA-S3, we consider again cyclic permutations. Note that GA-S3 searches for an efficient short design for a single session, and the short design obtained is then used to generate a juxtaposed design for all sessions through cyclic permutations. This is different from GA-L3, which aims for a short design that results in a efficient juxtaposed design with respect to an MO-criterion F^*.

5.4 SIMULATIONS

Through simulations, we demonstrate the performance of approaches introduced in the previous section. These approaches are summarized in Table 5.1. We focus on $Q = 2$, 3, and 4, and designs of lengths $L = 242$, 255 and 624, respectively. The number of sessions B is set to the number of stimulus types Q. The ISI and TR are both set to two seconds.

Table 5.1: Algorithms compared in the simulations

Name	Algorithm description
GA-S1	Use the same efficient short design for all sessions
GA-S2	Use different efficient short designs for different sessions
GA-S3	Cyclically permute stimulus types in an efficient short design to create designs for other sessions
GA-L1	Search for efficient designs under models without session effects
GA-L2	Search for efficient designs under models with session effects
GA-L3	Search for the short design that, with its cyclically permuted designs, yields an efficient juxtaposed design for all sessions

Two cases are considered. For models (5.1) and (5.2), Case I assumes white noise, i.e., $\boldsymbol{V} = \boldsymbol{I}$, and $\boldsymbol{S}$ is a vector of ones. In Case II, we assume that, for each session, $\boldsymbol{e}$ and $\boldsymbol{\eta}$ are stationary AR(1) processes with a correlation coefficient of 0.3, and $\boldsymbol{S}$ corresponds to a second-order Legendre polynomial. While Case I is frequently studied in the literature, Case II is closer to the settings used by practitioners. We consider two parametric functions, which are individual stimulus effects and pairwise contrasts. Specifically for detection problems, where $\boldsymbol{C}_z\boldsymbol{\theta}$ is of interest, $\boldsymbol{C}_z = \boldsymbol{I}_Q$ for individual stimulus effects, and $\boldsymbol{C}_z = \boldsymbol{D}$ for pairwise contrasts. Here, each row of $\boldsymbol{D}$ corresponds to the difference between θ_{q_1} and θ_{q_2}; $1 \leq q_1 \neq q_2 \leq Q$. For estimation problems, we use $\boldsymbol{C}_x = \boldsymbol{C}_z \otimes \boldsymbol{I}_k$. The value of k, which is the length of $\boldsymbol{h}_q$ in model (5.1), is decided by the selected basis $\boldsymbol{h}_0$ of the HRF for model (5.2). A common choice of $\boldsymbol{h}_0$ is the canonical HRF of SPM2 (http://www.fil.ion.ucl.ac.uk/spm), a popular software for analyzing fMRI data. The duration of the canonical HRF is 32s, counting from the stimulus onset to the complete return of the HRF to baseline. With both ISI and TR equal to two, the corresponding k is 17 ($= 1 + \frac{32}{2}$); see Kao et al. (2009b) for details.

For MO-criteria F^*, we allow the weight w to increase from 0 to 1 in steps of 0.1, thereby gradually shifting emphasis from the estimation problem to the detection problem. A total of 11 deigns are generated from each approach for each case. In Figure 5.1, we present the

F_e^*-values against F_d^*-values for the designs obtained with $Q = 2$, 3 and 4 and $\boldsymbol{C} = \boldsymbol{I}$. The figure for $\boldsymbol{C} = \boldsymbol{D}$ is omitted since it provides similar information. For clarity, we only include in the figure the designs obtained by GA-L1, GA-L2, GA-L3 and GA-S3 (GA-S3 performs best among the three heuristic approaches). Note that, to calculate F_d^* and F_e^*, we approximate $\max(F_e)$ and $\max(F_d)$ via GA-L2.

In the figure, GA-S3 performs relatively well for $Q = 2$ and 4. However, the designs generated by this approach are less efficient for $Q = 3$, especially when more weight is assigned to F_e^*. On the other hand, GA-L2 and GA-L3 consistently obtain the most efficient designs. We note that GA-L3 searches over a constrained design space compared to GA-L2. The result indicates that efficient designs can be found over this smaller space. In addition, designs obtained by GA-L1, which ignores session effects, can be inefficient, especially for Case II.

For a single session, the short designs found by GA-L3 are less efficient than those of GA-S3. However, juxtaposed designs obtained from GA-L3 become more efficient than those of GA-S3. To demonstrate this, we present in Table 5.2 the ratios of F_e-values of the designs obtained from GA-L3 to those from GA-S3. For example, with Case II, $Q = 3$, and $\boldsymbol{C}_z = \boldsymbol{D}$, the short design found by GA-L3 attains only 69% of the estimation efficiency of that found by GA-S3. However, the juxtaposed design corresponding to GA-L3 reaches 127% of the F_e-value achieved by the GA-S3 design for all sessions.

Table 5.2: Ratio of F_e-values (GA-L3 to GA-S3, in percentage)

Q	Design	Case I		Case II	
		$\boldsymbol{C}_z = \boldsymbol{I}$	$\boldsymbol{C}_z = \boldsymbol{D}$	$\boldsymbol{C}_z = \boldsymbol{I}$	$\boldsymbol{C}_z = \boldsymbol{D}$
2	All sessions	103	104	102	107
	Single session	91	96	92	94
3	All sessions	109	124	118	127
	Single session	70	69	64	69
4	All sessions	108	112	109	112
	Single session	97	79	65	80

We implement our simulations by using MATLAB (version 7.3) on a Linux cluster with 64-bit AMD Opteron, dual-processor, mix of single-core node and dual-core node; each core has 2GB RAM and the Linux operating system is 2.6.9-78.0.5.ELsmp. The CPU time spent by the six methods for obtaining the 11 designs under different cases are presented in Table 5.3. Since GA-S1 to GA-S3 obtain designs via different manipulations of short designs, we only present the total time spent for achieving the 11 short designs. For GA-S2, different short designs are generated by assigning different random seeds to start the GA. These designs can be obtained in parallel. As indicated by Table 5.3 and Figure 5.1, GA-L3 uses less CPU time than GA-L2 and still yields efficient designs.

We also note that more CPU time is consumed for most cases when $\boldsymbol{C}_z = \boldsymbol{D}$. The additional time mainly comes from the check for estimability. For $\boldsymbol{C}_z = \boldsymbol{I}$, we need $\mathcal{M} = \sum_{b=1}^{B} \boldsymbol{W}^{(b)'} \boldsymbol{V}'(\boldsymbol{I}_{T_B} - \boldsymbol{P}_{\boldsymbol{VS}}) \boldsymbol{V} \boldsymbol{W}^{(b)}$ to be nonsingular. For $\boldsymbol{C} = \boldsymbol{D}$, the parametric function can still be estimable when $\mathcal{M}$ is not of full rank. To ensure estimability for the latter case, we examine the equality $\boldsymbol{D}\mathcal{M}^{-}\mathcal{M} = \boldsymbol{D}$. The parametric function is estimable if this equality holds (Seber, 1977). Performing this check requires more CPU time.

Table 5.3: Total CPU time spent for obtaining the 11 designs (in hours)

Q		Case I		Case II	
		$\boldsymbol{C}_z = \boldsymbol{I}$	$\boldsymbol{C}_z = \boldsymbol{D}$	$\boldsymbol{C}_z = \boldsymbol{I}$	$\boldsymbol{C}_z = \boldsymbol{D}$
2	GA-S1 to GA-S3*	1.4	1.6	1.4	1.6
	GA-L3	2.4	2.9	2.3	3.0
	GA-L2	3.1	5.9	4.5	5.9
	GA-L1	3.0	5.9	2.9	6.0
3	GA-S1 to GA-S3*	2.3	2.6	2.4	2.7
	GA-L3	3.2	4.2	3.1	4.3
	GA-L2	4.5	10.1	6.8	9.9
	GA-L1	4.1	9.8	4.1	9.8
4	GA-S1 to GA-S3*	3.7	4.7	3.8	4.5
	GA-L3	16.9	19.2	17.4	19.5
	GA-L2	20.3	39.1	31.5	39.5
	GA-L1	19.1	38.8	19.5	38.2

*: GA-S2 uses B independent GA runs and takes B times the presented CPU time if these runs are performed sequentially
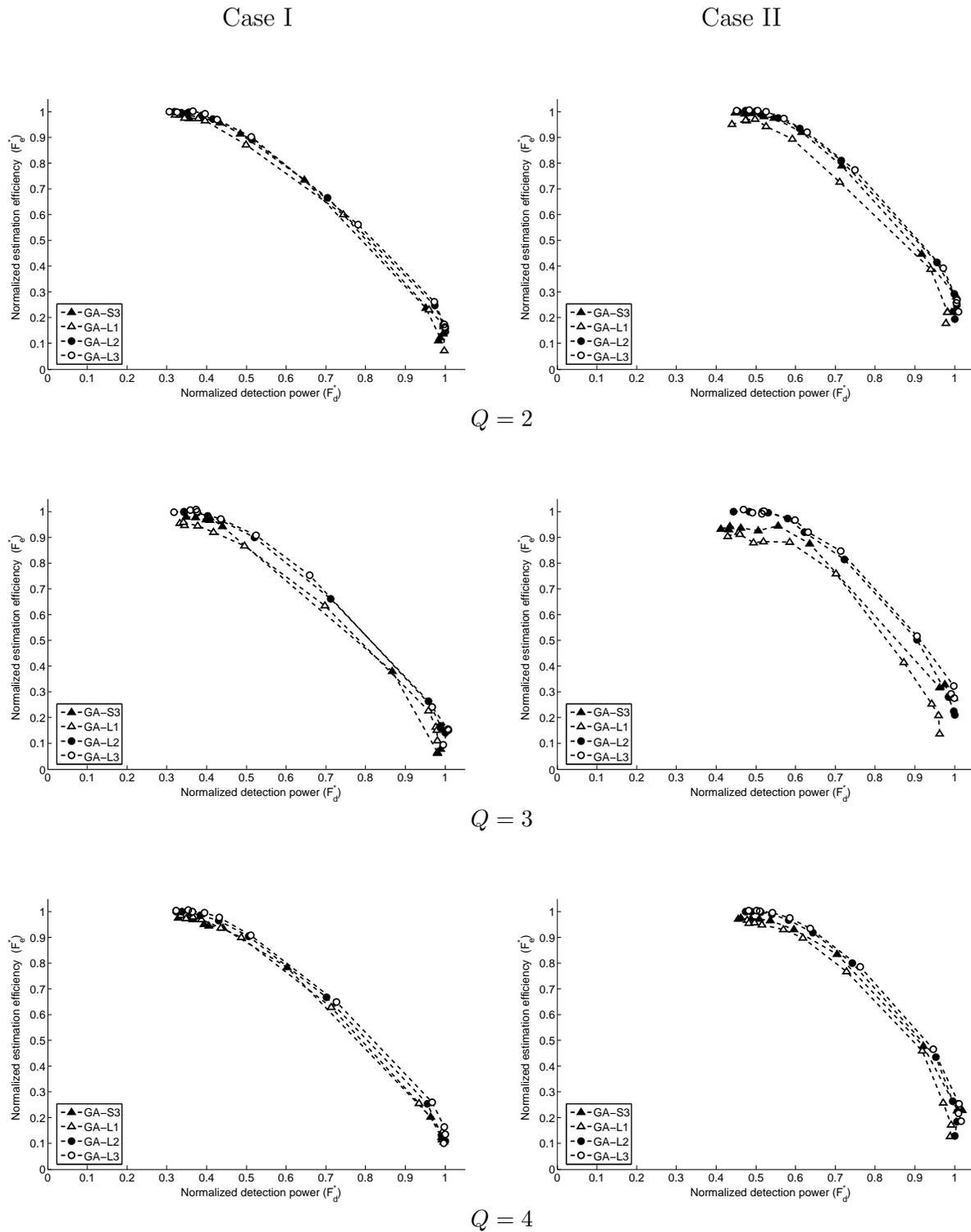
Figure 5.1: F_e^*-values versus F_d^*-values of designs obtained with $\boldsymbol{C} = \boldsymbol{I}$.

5.5 Conclusions and Discussion

In this chapter, we obtain efficient experimental designs for ER-fMRI when multiple sessions are included in one experiment. We compare the six approaches listed in Table 5.1 via simulations. In the simulation, we use $B = Q$ (the number of sessions = the number of stimulus types). Similar results are also observed for $B \neq Q$.

The first three approaches that we compared are GA-S1, GA-S2 and GA-S3. These approaches find (short) designs that are efficient for a single scanning session. These short designs are manipulated to form long designs for the entire experiment. GA-S1 uses the same short design for all sessions. For different sessions, GA-S2 utilizes different short designs with similar design efficiencies. GA-S3 cyclically permutes stimulus types in an efficient short design to form new short designs for subsequent scanning sessions.

The other three approaches search for efficient designs for the entire experiment directly. When evaluating competing designs, GA-L1 considers a model with no session effects, whereas the model utilized in GA-L2 is with session effects; both approaches search the entire design space for good designs. GA-L3 is similar to GA-L2 but searches over a restricted design space. The designs considered in GA-L3 are juxtaposed designs formed by a short design and permutations of that design. As in GA-S3, we consider cyclical permutations.

We evaluate the performance of all competing designs in experiments with multiple sessions. Among the approaches being compared, GA-L2 and GA-L3 consistently obtain the most efficient designs. We note that GA-L3 searches over a subclass of designs and it uses less CPU time than GA-L2 to obtain efficient designs. This indicates that good designs can be found more efficiently by searching over this restricted design space.

In addition, designs obtained under models without session effects (GA-L1) can be less efficient for experiments involving multiple scanning sessions. This is especially true when correlated noise and a drift in the response are assumed.

Among the heuristic approaches, GA-S3 performs best. This approach obtains a design for all sessions by cyclically permuting the stimulus types in an efficient short design for the

first session. For some cases, this approach obtains efficient designs, and it uses much less CPU time than GA-L2 and GA-L3. However, the approach does not perform well for other cases, especially when the length of a design for a session is short relative to the number of parameters of interest. Compared to these heuristic approaches, GA-L2 and GA-L3 are more reliable.

In this study, we consider designs with no pre-scanning periods; i.e., no stimuli are presented outside the scanning sessions. This assumption is also made by, e.g., Liu and Frank (2004) and Liu et al. (2001). Alternatively, one can consider pre-scanning periods and present extra stimuli to the subject before each scanning session as does Aguirre (2007). In that paper, a long design is divided into short designs for multiple sessions, and the last few stimuli of the previous session are presented again before each session. The last few stimuli of the last session are presented before the first session. Whereas Aguirre (2007) does not aim for designs yielding maximal efficiencies under multiple sessions, we might apply the GA approaches to search for such designs. However, pursuing this is beyond the scope of this chapter.

For the sake of clarity, we only focus on statistical objectives in this study. In addition to these objectives, Kao et al. (2009a) also consider psychological constraints, and customized requirements. Including these additional objectives in our family of MO-criteria is straightforward. In addition, we can also consider the case when the researcher's interest lies in both individual stimulus effect and pairwise contrasts. As in Kao et al. (2009b), the $\boldsymbol{C}_z$ matrix can then be taken as $\boldsymbol{C}_z = [\delta \boldsymbol{I}, (1-\delta)\boldsymbol{D}']'$; $\delta \in [0, 1]$.

5.6 References

Aguirre, G. K. (2007), "Continuous Carry-over Designs for fMRI," *NeuroImage*, 35, 1480–1494.

Atkinson, A. C., Donev, A., and Tobias, R. (2007), *Optimum Experimental Designs, with SAS*, Great Britain: Oxford U Pr.

Barker, H. A. (2004), "Primitive Maximum-Length Sequences and Pseudo-Random Signals," *Transactions of the Institute of Measurement & Control*, 26, 339–348.

Brown, M. R. G., Vilis, T., and Everling, S. (2008), "Isolation of Saccade Inhibition Processes: Rapid Event-Related fMRI of Saccades and Nogo Trials," *NeuroImage*, 39, 793–804.

Buračas, G. T. and Boynton, G. M. (2002), "Efficient Design of Event-Related fMRI Experiments Using M-Sequences," *NeuroImage*, 16, 801–813.

Buxton, R. B., Liu, T. T., Martinez, A., Frank, L. R., Luh, W. M., and Wong, E. C. (2000), "Sorting Out Event-Related Paradigms in fMRI: The Distinction Between Detecting an Activation and Estimating the Hemodynamic Response," *NeuroImage*, 11, S457.

Cabeza, R. and Kingstone, A. (2006), *Handbook of Functional Neuroimaging of Cognition*, Cognitive Neuroscience, Cambridge, Mass.: MIT Press, 2nd ed.

Culham, J. C. (2006), "Functional Neuroimaging: Experimental Design and Analysis," in *Handbook of Functional Neuroimaging of Cognition*, eds. Cabeza, R. and Kingstone, A., Cambridge, Mass.: MIT Press, 2nd ed.

Dale, A. M. (1999), "Optimal Experimental Design for Event-Related fMRI," *Human Brain Mapping*, 8, 109–114.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., and Dale, A. M. (1999), "Stochastic Designs in Event-Related fMRI," *NeuroImage*, 10, 607–619.

Godfrey, K. (1993), *Perturbation Signals for System Identification*, New York: Prentice Hall.

Harms, M. P., Guinan, J. J., Sigalovsky, I. S., and Melcher, J. R. (2005), "Short-Term Sound Temporal Envelope Characteristics Determine Multisecond Time Patterns of Activity in Human Auditory Cortex as Shown by fMRI," *Journal of Neurophysiology*, 93, 210–222.

Harms, M. P. and Melcher, J. R. (2002), "Sound Repetition Rate in the Human Auditory Pathway: Representations in the Waveshape and Amplitude of fMRI Activation," *Journal of Neurophysiology*, 88, 1433–1450.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications To Biology, Control, and Artificial Intelligence*, Ann Arbor: University of Michigan Press.

— (1992), *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, Complex adaptive systems, Cambridge, Mass.: MIT Press, 1st ed.

Josephs, O. and Henson, R. N. A. (1999), "Event-Related Functional Magnetic Resonance Imaging: Modelling, Inference and Optimization," *Philosophical Transactions: Biological Sciences*, 354, 1215–1228.

Kao, M.-H., Mandal, A., Lazar, N., and Stufken, J. (2009a), "Multi-Objective Optimal Experimental Designs for Event-Related fMRI Studies," *NeuroImage*, 44, 849–856.

Kao, M.-H., Mandal, A., and Stufken, J. (2009b), "Optimal Design for Event-Related Functional Magnetic Resonance Imaging Considering Both Individual Stimulus Effects and Pairwise Contrasts," *submitted*.

Lazar, N. A. (2008), *The Statistical Analysis of Functional MRI Data*, Statistics for Biology and Health, New York: Springer.

Liu, T. T. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part II: Design of Experiments," *NeuroImage*, 21, 401–413.

Liu, T. T. and Frank, L. R. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part I: Theory," *NeuroImage*, 21, 387–400.

Liu, T. T., Frank, L. R., Wong, E. C., and Buxton, R. B. (2001), "Detection Power, Estimation Efficiency, and Predictability in Event-Related fMRI," *NeuroImage*, 13, 759–773.

MacWilliams, F. J. and Sloane, N. J. A. (1977), *The Theory of Error Correcting Codes*, North-Holland Pub. Co.

Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998), "Event-Related functional MRI: Past, Present, and Future," *PNAS*, 95, 773–780.

Seber, G. A. F. (1977), *Linear Regression Analysis*, Wiley series in probability and mathematical statistics, New York: Wiley.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of Experimental Design in fMRI: A General Framework Using A Genetic Algorithm," *NeuroImage*, 18, 293–309.

Wang, Y. P., Xue, G., Chen, C. S., Xue, F., and Dong, Q. (2007), "Neural Bases of Asymmetric Language Switching in Second-Language Learners: An ER-fMRI Study," *NeuroImage*, 35, 862–870.

Hybrid Algorithms in Searching for Optimal Experimental Designs for Event-Related fMRI[1]

6.1 Introduction

Event-related functional magnetic resonance imaging (ER-fMRI) is one of the leading brain mapping techniques for studying brain activity in response to mental tasks or stimuli (Rosen et al., 1998; Josephs and Henson, 1999; Culham, 2006). Efficient designs that help to collect informative data for rendering valid and precise statistical inference are in great demand for this pioneering technology. However, due to the complex nature of ER-fMRI experiments, obtaining good designs is difficult; it requires consideration of a variety of statistical and practical issues, and involves major computational efforts. An efficient, versatile approach for obtaining good designs best suited for the researcher's needs is therefore crucial.

Kao et al. (2009a) propose an approach for finding optimal ER-fMRI designs. Their approach includes a genetic algorithm (GA) incorporating knowledge about the performance of well known ER-fMRI designs and is demonstrated to outperform previous methods. A MATLAB program implementing this approach is available in Kao (2009). Kao et al. (2009c) adopt this knowledge-based approach in searching for efficient designs when the researcher's interest lies in both individual stimulus effects and pairwise contrasts between stimulus types. An application regarding dividing a long scanning session into multiple short ones can be found in Kao et al. (2009b). This application reflects a common practice for protecting against a decrease in the experimental subject's attention level over a long session.

Following this line of research, we study here a family of algorithms for finding efficient ER-fMRI designs. This family consists of the knowledge-based GA of Kao et al. (2009a) and its variants. These variants are formed by considering two modifications regarding design selections. The first modification is applied at the stage where designs are selected into a mating pool. The second modification involves two methods for selecting designs to survive to the next GA generation.

The first modification considers selecting designs into a mating pool with probability proportional to Boltzmann transformed fitness (value of the design criterion). The Boltzmann transformation is used by simulated annealing (SA; Kirkpatrick et al., 1983), and is

considered here to allow a variety of selective pressure of the GA, which is the ratio of the probability of selecting the best design to the average selection probability of all designs in the population. Strong selective pressure may lead to a premature convergence since a good design, which is not necessarily optimal, can take over the entire GA population quickly. On the other hand, algorithms with weak selective pressure have wide population diversity, but they may converge slowly; see Michalewicz (1996).

The Boltzmann transformation is indexed by the "temperature", T_0. With a high temperature, the transformed fitness function is rather flat. Selecting designs based on this transformed fitness is similar to a random selection, and selection pressure in this case is weak. For low temperatures, selection pressure becomes strong, and good designs have larger chances to be selected into mating pool. Following the concept of SA, we gradually decrease T_0 during a GA search; i.e., a cooling schedule. At the beginning of the search, a cooling schedule allows exploration over a wide range in the design space containing all possible designs. It then helps to "exploit" more promising areas by selecting good designs to reproduce offsprings.

In addition, we also follow the stochastic evolutionary algorithm (Saab and Rao, 1991) to consider a warming schedule. By starting with a low temperature, this schedule helps to find good designs first. The temperature is then gradually increased, and the algorithm searches over a wider area of the designs space for better designs.

The second modification to the knowledge-based GA considers different ways for deciding the survival of designs. The GA of Kao et al. (2009a) consider an elitist-based method where the offspring designs compete with their parents and only the better designs can survive. Here, we also consider another approach which allows only offsprings to survive to the next generation. This is age-based.

We compare the performance of the GA and its variants in finding optimal ER-fMRI designs. We consider cases where stimuli can be present in the "warm-up" period of the MR scanner. In a typical ER-fMRI experiment, the first few MR scans are discarded due to this

warm-up period; the MR scanner is unstable during this period. However, stimuli can still be present in this period. For example, Aguirre (2007) presents stimuli of 15 seconds in this period; these stimuli are the same as the last few stimuli presented. Note that, although no valid observations are available during this period, the stimuli presented in this period still have impacts on the subsequent responses.

Our simulation results show that the original GA performs better than its variants, especially when the design space becomes large. We therefore advocate the use of this knowledge-based GA. We also observe that, while presenting stimuli in the warm-up period can increase the design efficiency in estimating the hemodynamic response function (HRF) as well as in detecting brain activation, the gain in design efficiency is small for a 16-second warm-up period.

In the next section, we briefly introduce ER-fMRI designs, the model considered in this study, design criterion and algorithms utilized. Simulation results are presented in Section 6.3 and a conclusion in Section 6.4.

6.2 METHODOLOGY

6.2.1 ER-fMRI DESIGNS

ER-fMRI designs are sequences of brief stimuli of one or more types with the control. When being presented to the experimental subject, each stimulus lasts for a very short period of time and is followed by another stimulus after a multiple of ISI (inter-stimulus interval) seconds. The gap in between stimuli are filled by the control (e.g., blank or fixation). For convenience, the design is usually represented by a sequence of finite numbers 0, 1, ... ,Q(=number stimulus types). The symbol i stands for a type-i stimulus, $i = 1, ..., Q$, and the inclusion of 0, which represents the control, helps to calculate the timing of each stimulus. For example, the onset times of the first and the second type-1 stimuli of the design $\xi = \{1012\cdots0\}$ are $2ISI$ apart, whereas the onset time of the third stimulus (a type-2 stimulus) is at ISI following the onset of the second type-1 stimulus.

For an active brain voxel, each stimulus evokes an effect that changes the strength of the MR signal over time. This change is described by a function of time, called HRF. A typical HRF arises in one or two seconds after a stimulus onset, peaks at about five to seven seconds, and falls back to the baseline followed by a undershooting that has a nadir occurring at about 15 seconds. The entire process, called the duration of the HRF, takes about 25 to 30 seconds; see also, Lazar (2008). It is noteworthy that the HRF can change across brain voxels, subjects, scanning sessions and stimulus types. Even so, assuming the same HRF in an experiment is not uncommon in practice, especially for the problem of detecting activated voxels (e.g. Makni et al., 2008).

In many experiments, the ISI is set to be (much) shorter than the duration of the HRF. Stimuli can thus be close in time, so that the resulting HRFs overlap. The signal values from overlapping HRFs are assumed to be cumulative and are collected by the MR scanner every TR (time-to-repetition) seconds along with noise. While a saturation can occur in the accumulated signals, it is very popular to assume that signals are additive especially when ISI is greater than four seconds (see e.g. Wager et al., 2005; Dale and Buckner, 1997). Under this additive assumption, the linear model framework is applied to analyze fMRI data, and the two common statistical objectives are estimation of the HRF and detection of activated voxels.

In the next subsection, we consider linear models reflecting the fact that the MR machine requires several seconds to warm up. Since data collected during this warm-up period are not stable, they are often deleted from the analysis. However, the effects of the stimulus presented in this period, if any, can still influence later responses. It is therefore important to take this influence into account if stimuli are present during the warm-up period. Our models accomplish this.

6.2.2 Models

The general linear model framework is popular for ER-fMRI (e.g., Kao et al., 2009a; Liu and Frank, 2004; Liu, 2004; Wager and Nichols, 2003; Liu et al., 2001; Dale, 1999; Friston et al., 1999). Following this popular framework, we consider the models presented in (6.1) and (6.2) for estimation and detection, respectively. Here, we modify the common model formulation to incorporate the warm-up period of the MR scanner, which is a pre-specified time period prior to the first valid data acquisition. The models are:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{e}, \text{ and} \tag{6.1}$$

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\eta}. \tag{6.2}$$

Except for the matrices $\boldsymbol{X}$ and $\boldsymbol{Z}$, the terms in these two models are the same as those in Kao et al. (2009a) where no warm-up period is assumed. Specifically, $\boldsymbol{Y}$ is a T-by-1 vector of the fMRI time series from a voxel, $\boldsymbol{h} = (\boldsymbol{h}_1', ..., \boldsymbol{h}_Q')'$ is the Qk-by-1 HRF parameter vector with $\boldsymbol{h}_q$ representing the HRF evoked by the qth stimulus types, $q = 1, ..., Q$, k is the length of $\boldsymbol{h}_q$, $\boldsymbol{S}\boldsymbol{\gamma}$ describes the trend or drift of $\boldsymbol{Y}$ with $\boldsymbol{\gamma}$ representing the corresponding unknown parameter vector, $\boldsymbol{\theta} = (\theta_1, ..., \theta_Q)'$ contains the Q response amplitudes, and $\boldsymbol{e}$ and $\boldsymbol{\eta}$ are noise. For estimation problems, $\boldsymbol{h}$ is of main interest. On the other hand, studying $\boldsymbol{\theta}$ helps us to identify whether a voxel is activated or not.

The matrix $\boldsymbol{X} = [\boldsymbol{X}_1 \cdots \boldsymbol{X}_Q]$ is a T-by-Qk design matrix. Originally in Kao et al. (2009a), the upper-triangle of $\boldsymbol{X}$ is zero; see also, Kao et al. (2009c) and Liu et al. (2001). When stimuli are present in the warm-up period, some elements in the upper-triangle become one. These elements reflect carry-over effects of the stimuli in the warm-up period to latter periods. To obtain such a design matrix $\boldsymbol{X}$, we can first construct an enlarged design matrix as if the data are also acquired during the warm-up period (so that the upper-triangle of this enlarged matrix is zero). The first few rows of the enlarged matrix, which correspond to the warm-up period, is then trimmed to form $\boldsymbol{X}$. The matrix $\boldsymbol{Z}$ is $\boldsymbol{X}\boldsymbol{h}_0$, which stands for the convolution of stimuli with an assumed basis, $\boldsymbol{h}_0$, of the HRF.

When stimuli are presented in the warm-up period, we follow Aguirre (2007) to place the last few stimuli of a design sequence in this period.

6.2.3 DESIGN CRITERIA

Following convention, we evaluate the quality of a design sequence based on functions of the corresponding information matrix. The A-optimality criterion, which attempts to minimize the averaged variance of parameter estimators, is considered. Our criteria have the form of:

$$\Phi_A(\xi) = \frac{r_c}{trace(\boldsymbol{C}[\boldsymbol{W}^T\boldsymbol{V}^T(\boldsymbol{I} - \boldsymbol{P_{vs}})\boldsymbol{V}\boldsymbol{W}]^-\boldsymbol{C}^T)}, \tag{6.3}$$

where $\boldsymbol{W} = \boldsymbol{X}$ for model (6.1), and $\boldsymbol{W} = \boldsymbol{Z}$ for model (6.2), $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{P_A} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^-\boldsymbol{A}^T$ is the orthogonal projection on the vector space spanned by the column vectors of $\boldsymbol{A}$, $\boldsymbol{A}^-$ is a generalized inverse matrix of $\boldsymbol{A}$, and ξ is the design being evaluated. We assume a known $\boldsymbol{V}$ so that $\boldsymbol{V}\boldsymbol{e}$ and $\boldsymbol{V}\boldsymbol{\eta}$ are white noise. The matrix $\boldsymbol{C}$ indicates the linear combinations of parameters of interest, and r_c is the number of rows of $\boldsymbol{C}$. When individual stimulus effects are of interest, $\boldsymbol{C}$ is set to an identity matrix. When pairwise contrasts between stimulus types are of interest, each row of $\boldsymbol{C}$ corresponds to a pairwise difference. Other parametric functions, $\boldsymbol{C}_x\boldsymbol{h}$ and $\boldsymbol{C}_z\boldsymbol{\theta}$, can also be considered.

To accommodate the two competing statistical objectives, we consider weighted sums of the two statistical criteria. Before combining them in a multi-objective criterion (MO-criterion), the individual criteria are standardized by their maximal values; see, e.g., Dette (1997) for a justification of standardized criteria. Our MO-criteria are defined as:

$$F(\xi) = w_d F_d^*(\xi) + w_e F_e^*(\xi),$$

where $F_d^*(.)$ and $F_e^*(.)$ are the standardized criteria for estimation and detection, respectively, and w_is are user-specified weights; $i = d, e$. Specifically,

$$F_i^*(\xi) = \frac{F_i(\xi)}{\max\limits_{\xi \in \Xi} F_i(\xi)}, i = d, e,$$

where $F_d(\xi)$ and $F_e(\xi)$ are $\Phi_A(\xi)$ of (6.3) with $\boldsymbol{W} = \boldsymbol{Z}$ and $\boldsymbol{X} = \boldsymbol{X}$, respectively, and Ξ is the design space containing all possible designs. Our goal is to find designs maximizing $F(.)$.

6.2.4 Search Algorithm

We compare algorithms in this study for finding optimal ER-fMRI designs that optimize the MO criterion, F. These algorithms are the knowledge-based GA of Kao et al. (2009a), and its variants. For the GA of Kao et al. (2009a), knowledge about the performance of well known ER-fMRI designs is utilized to construct the initial designs and immigrants. With a probability proportional to the value of the objective function, parents are selected from the initial designs to generate offsprings through the crossover operator, which exchanges corresponding subsequences of the paired parent designs. The mutation operator then perturbs a small fraction of elements of the offspring designs. Immigrants are introduced to the population to increase population diversity; they help to jump out of local optima and introduce good traits to the population. Elites or the better designs among parents, offspring and immigrants survive to form the next generation. The process is repeated and terminated after a stopping rule is met.

We modify this GA by considering the following Boltzmann transformation of the objective function (formulated for maximizing the objective function F):

$$B_{T_0}(F) = \exp\{\frac{F}{T_0}\},$$

where T_0 is a given temperature. Instead of directly using F, $B_{T_0}(F)$ is used for selecting parents. Specifically, the probability for each design to be selected into a mating pool is proportional to the $B_{T_0}(F)$-value. When T_0 is large, $B_{T_0}(F)$ is rather insensitive to the F-value. On the other hand, $B_{T_0}(F)$ changes drastically with the F-value under a small T_0. The selective pressure, which is the ratio of the probability of the best design being selected into a mating pool to the average selection probability of all designs in the population, is lower for the former than for the latter. It is known that an algorithm with a high selective pressure tends to achieve good designs quickly, but it might converge prematurely. The convergence rate under a low selective pressure is slow. However, it might lead to a better design.

Following simulated annealing, we gradually decrease the temperature T_0 during the search. For implementation, we apply double-loop approaches which consist of an inner loop and an outer loop (e.g. Jin et al., 2005). For each given temperature T_0, the inner loop is the GA with a Boltzmann transformed fitness for selecting parents. After M_I generations in the inner loop, the outer loop changes T_0. This process is then repeated M_O times, where M_O is the number of iterations in the outer loop. We also consider a warming schedule that gradually increases T_0 and the double-loop approach is again applied.

In addition to considering the Boltzmann transformation, we also consider two different methods for selecting designs to survive to the next generation. The GA of Kao et al. (2009a) uses the better designs among the parents, offsprings and immigrants to form the next generation. Another possible way is to use only offspring designs as the next generation of GA. Hereinafter, the former method is referred to as the elitist-based method and the latter one to as the age-based method. These two methods are applied to each of the three algorithms mentioned previously. We compare the following six algorithms in the next section:

- **elitist-based Algorithm I**: the GA of Kao et al. (2009a);

- **age-based Algorithm I**: the GA of Kao et al. (2009a) with the age-based selection method;

- **elitist-based Algorithm II**: the GA of Kao et al. (2009a) with a Boltzmann transformation, a cooling schedule, and elitist-based selection method;

- **age-based Algorithm II**: the GA of Kao et al. (2009a) with a Boltzmann transformation, a cooling schedule, and age-based selection method;

- **elitist-based Algorithm III**: the GA of Kao et al. (2009a) with a Boltzmann transformation, a warming schedule, and elitist-based selection method;

- **age-based Algorithm III**: the GA of Kao et al. (2009a) with a Boltzmann transformation, a warming schedule, and age-based selection method.

Note that MO criteria F involve normalized design criteria. When normalizing the design criteria, the maximal F_e- and F_d-values are approximated using the elitist-based Algorithm I.

6.3 SIMULATIONS

We study the performance of the six algorithms through simulations. Designs with stimuli of one to three types ($Q = 1$, 2, and 3) are considered, and the corresponding lengths of designs are $L = 255$, 242 and 255, respectively. Both the ISI and TR are set to two seconds and we consider a total of eight scans (16 seconds) in the warm-up period of the MR scanner. Therefore, when stimuli are present in the warm-up period, the design lengths become 263, 250 and 263 for $Q = 1$, 2 and 3, respectively. Similar settings are also considered by Liu (2004).

The $\boldsymbol{h}_0$ described under model (6.2) is set to the canonical HRF of the SPM2, which is a combination of two gamma distributions; see, (7.1) in Chapter 7. Corresponding to this $\boldsymbol{h}_0$, which has a duration of 32 seconds, the length of $\boldsymbol{h}_q$ for model (6.1) is $k = 17 = 1 + (32/\Delta T)$, where $\Delta T = 2$ is the greatest value dividing both the ISI and TR; $q = 1, ..., Q$. The noise for models (6.1) and (6.2) is set to follow a stationary AR(1) process with a correlation coefficient of $\rho = 0.3$, and the response is assumed to have a second-order polynomial drift. The matrix $\boldsymbol{C}$ of (6.3) is set to an identity matrix, which corresponds to the study of individual effects.

The number of designs in each generation or population size of elitist-based and age-based Algorithm I is set to 30, and the algorithm stops after 50,000 generations. The number of immigrants introduced to each generation is six and the mutation rate is 1%, meaning that for each generation one percent of elements from the 30 offspring designs are replaced by randomly generated ones.

As for Algorithms II and III, we stop the search after 500 runs, while the inner loops have 100 generations for each run. The temperature is adjusted in the outer loop based on a geometric schedule (e.g. Fouskakis and Draper, 2002). For the ith run, the temperature T_0
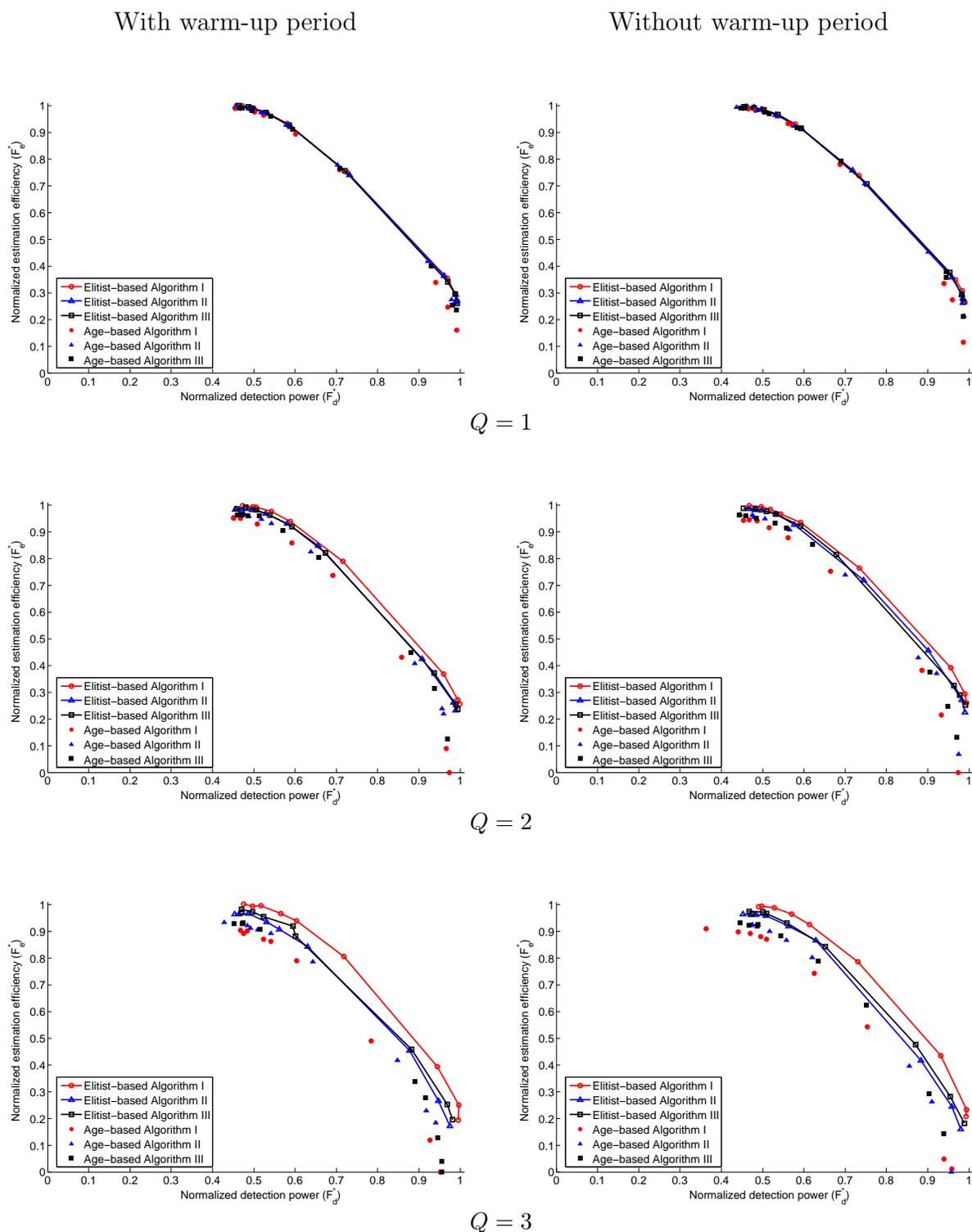
Figure 6.1: F_e^*-values versus F_d^*-values; left column: stimuli are present in the warm-up period; right column: stimuli are absent in the warm-up period

is:

$$
\begin{cases}
T_U\left(\frac{T_L}{T_U}\right)^{\frac{i-1}{M_O-1}}, & \text{for Algorithms II;} \\
T_L\left(\frac{T_U}{T_L}\right)^{\frac{i-1}{M_O-1}}, & \text{for Algorithms III.}
\end{cases}
$$

Here, T_L and T_U are the lower and upper bounds of the temperature, respectively, and they are set to 0.1 and 1 for this simulation.

We search for designs optimizing convex combinations of the two statistical objectives. By increasing the weight assigned to F_e from 0.1 to 0.9 in steps of 0.1, and decreasing the weight for F_d accordingly, we achieve nine designs from each algorithm. The normalized F_e^*-against F_d^*-values, for these designs are presented in Figure 6.1. From Figure 6.1, the elitist-based algorithms are slightly better than the age-based ones. In addition, the elitist-based Algorithm I can achieve better designs than those obtained by the other five methods when $Q = 2$ and 3. The six algorithms are comparable when $Q = 1$. While further investigations are needed, this figure indicates that, with larger design space, the elitist Algorithm I is better than the other algorithms being compared.

Table 6.1: F_e- and F_d-values of designs obtained from the elitist-based Algorithm I when stimuli are present/absent in the warm-up period

	$Q = 1$		$Q = 2$		$Q = 3$	
	F_e	F_d	F_e	F_d	F_e	F_d
Present	61.02	211.62	39.44	133.86	32.04	107.57
Absent	60.85	210.49	39.27	133.72	31.99	107.53

Table 6.1 presents the F_e- and F_d-values of designs obtained from the elitist Algorithm I for both cases where stimuli are present and absent in the warm-up period. While presenting stimuli in the warm-up period can increase the design efficiency, the difference between these two cases is small with a warm-up period of 16 seconds.

6.4 CONCLUSION

In this chapter, we compare algorithms for finding optimal ER-fMRI designs. These algorithms are the knowledge-based GA of Kao et al. (2009a) and its variants. We modify the

knowledge-based GA by considering the Boltzmann transformation on the design criteria. A cooling and a warming schedule are applied to change the "temperature" of the Boltzmann transformation, and hence alter the selective pressure of GAs. The cooling schedule gradually lowers the temperature, and thus increases the selective pressure. On the other hand, the warming schedule elevates the temperature during the GA search; the selective pressure is decreased accordingly.

In addition, we also consider two methods for selecting designs to form the next GA generation. One method is elitist-based and the other is age-based. While parent designs in a GA with the former selection method are allowed to pair with offsprings to reproduce new designs, the latter method allows only offspring designs to survive to the next generation before selecting designs into a mating pool. Combining these selection methods with the Boltzmann transformation, which is applied to select designs into a mating pool, we form five variants of the GA of Kao et al. (2009a). Through simulations, we compare the performance of these algorithms in finding ER-fMRI designs when the warm-up period of an MR scanner is taken into account. We follow Aguirre (2007) to place the last few stimuli of a design sequence to the warm-up period. The case where the stimuli are absent in the warm-up period is also considered.

Our simulations show that the original GA performs the best among the six algorithms being compared, regardless of the presence or absence of the stimuli in the warm-up period. This is especially true when the design space gets large. We also observe that the elitist-based algorithms are better than the age-based algorithms. From these observations, we strongly advocate the use of the simpler algorithm, the original GA of Kao et al. (2009a).

Other variants of the knowledge-based GA can also be found in Kao et al. (2007). When selecting designs to survive to the next GA generation, Kao et al. (2007) consider the use of a function formed by a transformed design criterion and a random noise; designs with larger values of this function can survive. Kao et al. (2007) form variants of the GA by considering different ways for changing the importance of the random noise (and hence the

transformed design criterion). When the random noise has a larger magnitude than the transformed design criterion, designs with low design efficiency can still be chosen to form the next generation. When the magnitude of the random noise is very small, only elites can survive. In Kao et al. (2007), the performance of the variants is shown to be similar to that of the original knowledge-based GA. The latter algorithm is therefore suggested since it is simpler.

6.5 REFERENCES

Aguirre, G. K. (2007), "Continuous Carry-over Designs for fMRI," *Neuroimage*, 35, 1480–1494.

Culham, J. C. (2006), "Functional Neuroimaging: Experimental Design and Analysis," in *Handbook of Functional Neuroimaging of Cognition*, eds. Cabeza, R. and Kingstone, A., Cambridge, Mass.: MIT Press, 2nd ed.

Dale, A. M. (1999), "Optimal Experimental Design for Event-Related fMRI," *Human Brain Mapping*, 8, 109–114.

Dale, A. M. and Buckner, R. L. (1997), "Selective Averaging of Rapidly Presented Individual Trials Using fMRI," *Human Brain Mapping*, 5, 329–340.

Dette, H. (1997), "Designing Experiments with Respect to 'Standardized' Optimality Criteria," *Journal of the Royal Statistical Society Series B-Methodological*, 59, 97–110.

Fouskakis, D. and Draper, D. (2002), "Stochastic Optimization: A Review," *International Statistical Review / Revue Internationale de Statistique*, 70, 315–349.

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., and Dale, A. M. (1999), "Stochastic Designs in Event-Related fMRI," *NeuroImage*, 10, 607–619.

Jin, R., Chen, W., and Sudjianto, A. (2005), "An Efficient Algorithm for Constructing Optimal Design of Computer Experiments," *Journal of Statistical Planning and Inference*, 134, 268–287.

Josephs, O. and Henson, R. N. A. (1999), "Event-Related Functional Magnetic Resonance Imaging: Modelling, Inference and Optimization," *Philosophical Transactions: Biological Sciences*, 354, 1215–1228.

Kao, M.-H. (2009), "Multi-Objective Optimal Experimental Designs for ER-fMRI Using MATLAB," *Journal of Statistical Software*, to appear.

Kao, M.-H., Mandal, A., Lazar, N., and Stufken, J. (2007), "Multi-Objective Optimal Experimental Designs for Event-Related fMRI Studies," Tech. rep., Department of Statistics, University of Georgia, http://www.stat.uga.edu/~amandal/MO-fMRI.pdf.

Kao, M.-H., Mandal, A., Lazar, N., and Stufken, J. (2009a), "Multi-Objective Optimal Experimental Designs for Event-Related fMRI Studies," *NeuroImage*, 44, 849–856.

Kao, M.-H., Mandal, A., and Stufken, J. (2009b), "Efficient Designs for Event-Related Functional Magnetic Resonance Imaging with Multiple Scanning Sessions," *Communications in Statistics - Theory and Methods*, to appear.

— (2009c), "Optimal Design for Event-Related Functional Magnetic Resonance Imaging Considering Both Individual Stimulus Effects and Pairwise Contrasts," *Statistics and Applications (Special Volume in Honour of Professor Aloke Dey)*, to appear.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983), "Optimization by Simulated Annealing," *Science*, 220, 671–680.

Lazar, N. A. (2008), *The Statistical Analysis of Functional MRI Data*, Statistics for Biology and Health, New York: Springer.

Liu, T. T. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part II: Design of Experiments," *NeuroImage*, 21, 401–413.

Liu, T. T. and Frank, L. R. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part I: Theory," *NeuroImage*, 21, 387–400.

Liu, T. T., Frank, L. R., Wong, E. C., and Buxton, R. B. (2001), "Detection Power, Estimation Efficiency, and Predictability in Event-Related fMRI," *NeuroImage*, 13, 759–773.

Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., and Ciuciu, P. (2008), "A Fully Bayesian Approach to the Parcel-Based Detection-Estimation of Brain Activity in fMRI," *NeuroImage*, 41, 941–969.

Michalewicz, Z. (1996), *Genetic Algorithms + Data Structures = Evolution Programs*, Berlin; New York: Springer-Verlag, 3rd ed.

Rosen, B. R., Buckner, R. L., and Dale, A. M. (1998), "Event-Related functional MRI: Past, Present, and Future," *PNAS*, 95, 773–780.

Saab, Y. G. and Rao, V. B. (1991), "Combinatorial Optimization by Stochastic-Evolution," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 10, 525–535.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of Experimental Design in fMRI: A General Framework Using A Genetic Algorithm," *NeuroImage*, 18, 293–309.

Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. (2005), "Accounting for Nonlinear BOLD Effects in fMRI: Parameter Estimates and a Model for Prediction in Rapid Event-Related Studies," *NeuroImage*, 25, 206–218.

CHAPTER 7

EFFICIENT EXPERIMENTAL DESIGNS UNDER A NONLINEAR MODEL FOR

EVENT-RELATED FMRI[1].

_____

[1]Kao, M.-H., Majumdar, D., Mandal, A., Stufken, J. To be submitted to *Journal of the American Statistical Association*

7.1 INTRODUCTION

Event-related functional magnetic resonance imaging (ER-fMRI) is a rapidly growing area. While many statistical methodologies have been developed for analyzing ER-fMRI data (for an overview, see, e.g. Lazar, 2008), the research on experimental designs is largely underdeveloped for this pioneering technology. As indicated by Lindquist (2008), "While the area of experimental design is a natural domain for statisticians to conduct research, it has so far been largely unexplored by members of the field. $\cdots$ the need for more advanced experimental designs will only increase further and this is clearly an area where statisticians can play an important role." In particular, we find no previous study on efficient designs for detecting activated brain voxels (equal sized volume elements of the brain) under nonlinear models. Our current study focuses on this important design problem.

Previous studies on ER-fMRI designs mainly focus on the linear model framework (e.g. Kao et al., 2009; Liu, 2004; Liu and Frank, 2004; Wager and Nichols, 2003; Liu et al., 2001). Under this framework, which is arguably the most popular for analyzing ER-fMRI data, a linear model involving an assumed hemodynamic response function (HRF; a function of time describing the change in the MR signal evoked by a single, brief stimulus) is considered for detecting voxels that are activated by the mental stimuli; see model (7.2) in Section 7.2. A valid assumption about the HRF is crucial since mis-specification of the HRF may lead to inaccurate statistical inference (e.g. Makni et al., 2008; Lindquist and Wager, 2007). However, researchers usually do not have full knowledge about the shape of the HRF, which can vary across brain voxels, experimental subjects and MR scanning sessions (Menz et al., 2006; Handwerker et al., 2004). Making a good assumption about the HRF can be difficult. One possible way to address this issue at the analysis stage is to estimate the HRF first and incorporate the estimated HRF in the model for detection. An experiment can therefore involve two statistical objectives, namely estimation and detection, even when the latter objective is the main concern (e.g. Lu et al., 2006).

Considering both statistical objectives, existing approaches for finding efficient ER-fMRI designs work with two linear models, one for each objective. Two objective functions, or design criteria, for evaluating the quality of designs are formed based on these models. Multi-objective (MO) designs reaching advantageous compromises between these two competing objectives are then pursued (e.g. Liu et al., 2001). However, such approaches still have the major disadvantage of having to assume a fixed HRF in the model for detection. Designs optimal for detecting activation under an assumed HRF may be inefficient when the assumed HRF is incorrect. In addition, only one design can be used for any subject in any scanning session. We would like a design capable of accommodating a variety of HRFs which are likely to occur in voxels of interest. To achieve this, we propose a more natural approach that involves one, single nonlinear model. Moving away from "dual model" approaches, the nonlinear model framework that we propose accommodates both estimation and detection in a unified setting. The uncertainly of the HRF can therefore be taken into account when planning designs for detecting activation.

However, obtaining good designs for nonlinear models is notoriously arduous since the information matrix depends on unknown parameter values. One popular approach is to find locally optimal designs, which requires a good guess of true parameter values (Chernoff, 1953). Such designs are optimal when the guessed parameter values are correct, but can be inefficient otherwise. Another approach, which is more computationally expensive, is to find pseudo-Bayesian designs under a pre-specified distribution of the parameters (Chaloner and Larntz, 1989); we adopt the term "pseudo-Bayesian" from Atkinson et al. (2007) to distinguish from the full Bayesian approaches involving utility functions (e.g. Chaloner and Verdinelli, 1995). With a design criterion that depends on the information matrix, pseudo-Bayesian designs optimize the expectation of the design criterion over the specified prior distribution of the parameters.

Both locally optimal designs and pseudo-Bayesian designs are pursued in this study. We follow previous studies to formulate the design problem as a discrete optimization problem.

The design space, which contains all possible designs, is enormous. There are $2^{255} \cong 5.8 \times 10^{76}$ possible designs for a design sequence that has of 255 elements and contains stimuli of one type and the control. Experiments with much larger scale are also not uncommon (e.g. Zhang and Yu, 2008). Searching over this huge design space for good designs requires an efficient search algorithm. Here, we adopt the knowledge-based genetic algorithm (GA) of Kao et al. (2009). This algorithm incorporates knowledge about the performance of well known ER-fMRI designs and is shown to be more efficient than approaches known hitherto.

The characteristics of locally optimal designs obtained by the GA are investigated. This provides important information about good features of designs required for different parameter values. In addition, by calculating the design efficiency relative to locally optimal designs, the performance of a design over a variety of parameter values can be evaluated. Pseudo-Bayesian designs that take into account the uncertainty of parameter values are also obtained. We note that, although pseudo-Bayesian designs maximize the expected design criterion, these designs may have undesirable characteristics. For example, with a uniform distribution on the parameter space, we can obtain a design that performs particularly well over one region but is inefficient over another, while a more balanced design efficiency is desired.

To help researchers obtain desirable designs suited for their needs and goals, we propose an idea borrowed from MO optimization problems. We would like to offer researchers a collection of efficient designs to choose from based on their requirements. To achieve this, we propose a simple greedy algorithm to transform the design problem to an MO problem, and then adopt the popular non-dominance sorting GA II (NSGA-II) of Deb et al. (2002) to search for such a collection of efficient designs.

In the following section, we briefly introduce the background information of ER-fMRI. Section 7.3 describes our methodologies including the statistical model, the derived design criteria and the algorithms for finding good ER-fMRI designs. A case study on efficient

designs for one stimulus type is provided in Section 7.4, followed by conclusion in Section 7.5.

7.2 BACKGROUND

The ER-fMRI design is a finite sequence of brief stimuli of one or more types interlaced with the control (rest or fixation). During an experiment, the design sequence is presented to the subject while an MR scanner scans the subject's brain every TR (time-to-repetition) seconds. While being presented, each stimulus in the design lasts for a very short period of time; e.g., several milliseconds to a few seconds. The time interval between two consecutive onsets of stimuli is a multiple of ISI (inter-stimulus interval) seconds. The "off times" are filled with the control.

To facilitate mathematical representation, finite numbers, 0, 1, ..., Q, are used to denote the elements of a design sequence, which might look like $\xi = \{1011\cdots0\}$. Here, Q is the total number of stimulus types, and i represents a stimulus of the ith type; $i = 1,...,Q$. The symbol 0, indicating the control, helps to calculate the timing of each stimulus. For example, the above ξ represents a design with a type-1 stimulus occurring at the onset of the experiment (Time 0). A 0 before the second one means that the second stimulus is at $2ISI$ seconds, while the third stimulus is provided at time $3ISI$.

For a brain voxel activated by the stimuli, each stimulus evokes a change in the MR signals over time. Current knowledge attributes this change to alterations in the ratio of oxy- to deoxy-hemoglobin in the cerebral blood vessels (Cabeza and Kingstone, 2006), and the HRF is used to describe such a change. A popular choice of the HRF is the "canonical HRF" of the software SPM2 (http://www.fil.ion.ucl.ac.uk/spm), which is formed by gamma density functions (see also, Wager et al., 2005):

$$
g(\tau; \boldsymbol{p}_c = (p_1, p_2, ..., p_7)) = \begin{cases} \frac{h(\tau; \boldsymbol{p}_c)}{\max\limits_{s} h(s; \boldsymbol{p}_c)}, & \tau \in [\max(0, p_6), p_7] \\ 0, & \text{otherwise} \end{cases}, \tag{7.1}
$$

where

$$h(\tau; \boldsymbol{p}_c) = Gamma(x = (\tau - p_6), \alpha_1 = \frac{p_1}{p_3}, \gamma_1 = p_3)$$

$$- \frac{1}{p_5} Gamma(x = (\tau - p_6), \alpha_2 = \frac{p_2}{p_4}, \gamma_2 = p_4),$$

$$\boldsymbol{p}_c = (6, 16, 1, 1, 6, 0, 32),$$

$$Gamma(x, \alpha, \gamma) = \frac{x^{\alpha-1} e^{-x/\gamma}}{\Gamma(\alpha)\gamma^{\alpha}},$$

and $\Gamma(.)$ is the gamma function. Here, τ represents time elapsed from a stimulus onset. The given $\boldsymbol{p}_c$-value decides the shape of the canonical HRF. A natural generalization is to set (part of) $\boldsymbol{p}_c$ as free parameters so that the shape of the HRF is determined by data. This generalization will be discussed in Section 7.3.

The HRF, which has a maximal value of one, can be viewed as describing the "standardized effects" of a single, brief stimulus. When the HRFs overlap (i.e., the onsets of stimuli are no more than $p_7 = 32$ seconds apart), their values accumulate. Linear accumulation is commonly assumed (e.g. Dale and Buckner, 1997). Under this assumption, the popular general linear model framework uses the following model to describe the fMRI time series (observations are acquired at each scan):

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\theta} + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\eta}, \tag{7.2}$$

where $\boldsymbol{Y}$ (T-by-1) is the fMRI time series from a voxel; $\boldsymbol{\theta} = (\theta_1, ..., \theta_Q)$ represents response amplitudes; $\boldsymbol{Z} = \boldsymbol{X}(\boldsymbol{I}_Q \otimes \boldsymbol{h}_0)$ corresponds to the convolution of the stimuli with an assumed HRF $\boldsymbol{h}_0$; $\boldsymbol{I}_Q$ is the Q-by-Q identity matrix; $\otimes$ is the kronecker product; $\boldsymbol{X} = [\boldsymbol{X}_1 \cdots \boldsymbol{X}_Q]$ is the zero-one design matrix decided by the design sequence, ISI and TR (for details, see, Kao et al., 2009); $\boldsymbol{S}\boldsymbol{\gamma}$ is a nuisance term describing the trend or drift of $\boldsymbol{Y}$ with the parameter vector $\boldsymbol{\gamma}$; and $\boldsymbol{\eta}$ is noise which is commonly assumed to follow a stationary AR(1) process (e.g. Zhang and Yu, 2008).

As mentioned previously, the canonical HRF of (7.1) is widely used as the assumed HRF $\boldsymbol{h}_0$. Specifically, the jth element of $\boldsymbol{h}_0$ is $g(\tau = (j-1)\Delta T; \boldsymbol{p}_c)$; $j = 1, ..., k$. Here, ΔT is the

smallest value dividing both the ISI and TR, and $k = 1 + \lfloor 32/\Delta T \rfloor$ is the length of $\boldsymbol{h}_0$; $\lfloor a \rfloor$ is the greatest integer less than or equal to a. The same $\boldsymbol{h}_0$ is assumed for all stimulus types, which is not uncommon in practice (see, e.g., Makni et al., 2008).

With model (7.2), the detection problems focus on studying θ_is; a large θ_i signals a voxel highly activated by the stimulus of the ith type; $i = 1, ..., Q$. We want designs that help render the most precise statistical inference about θ_i. Block designs, where stimuli of the same type appear in clusters, are well known to be efficient for this purpose (e.g. Henson, 2007). For $Q = 1$, a block design may consist of repetitions of $\{00001111\}$; this design has a block size of four. Crudely put, by using these designs, the MR signals evoked by the stimuli of the same type accumulate to become strong signals, which are easier to be detected by the MR scanner. The block size of efficient block designs can vary with the assumed HRF. For example, Henson (2007) shows a high efficiency of a block size of about $(16/ISI)$ under the canonical HRF, while Liu et al. (2001) find it efficient to use larger block sizes under an HRF formed by a single gamma density function. Therefore, it is necessary to take into account the uncertainty of the HRF when planning an ER-fMRI design, and the nonlinear model introduced in the next section accomplishes this.

7.3 Methodology

7.3.1   A Nonlinear Model

When finding efficient designs for detection problems, model (7.2) is widely considered. However, such a model has a major limitation; it does not take into account the uncertainty of the HRF. Therefore, we consider the following nonlinear model which is free from this limitation:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{H}(\boldsymbol{p})\boldsymbol{\theta} + \boldsymbol{S}\boldsymbol{\gamma} + \boldsymbol{\eta}. \tag{7.3}$$

This model is a natural generalization of model (7.2). The matrix $\boldsymbol{Z} = \boldsymbol{X}(\boldsymbol{I}_Q \otimes \boldsymbol{h}_0)$ is replaced by $\boldsymbol{X}\boldsymbol{H}(\boldsymbol{p}) = \boldsymbol{X}(\boldsymbol{I}_Q \otimes \boldsymbol{h}_0(\boldsymbol{p}))$, where $\boldsymbol{p}$ is a free HRF parameter vector that needs to be estimated by data, and the jth element of the vector $\boldsymbol{h}_0(\boldsymbol{p})$ is $g((j-1)\Delta T; \boldsymbol{p})$; $j = 1, ..., k$.

Instead of assuming a pre-specified HRF, the shape of the HRF in this nonlinear model is determined by data.

Model (7.3) is linked to the model considered by Wager et al. (2005). Following their suggestions, we treat two influential HRF parameters, namely p_1, time-to-peak, and p_6, time-to-onset, as free parameters, while keeping others at the design stage at their default values in $\boldsymbol{p}_c$; so $(p_2, p_3, p_4, p_5, p_7) = (16, 1, 1, 6, 32)$. Writing from hereon $\boldsymbol{p} = (p_1, p_6)$, the parameter space for $\boldsymbol{p}$ is also inspired by Wager et al. (2005):

$$\mathcal{P} = \{(p_1, p_6) : p_1 \in [3.5, 11], p_6 \in [-2.5, 1.5], p_1 + p_6 > 1\}. \tag{7.4}$$

Other parameter spaces can also be considered.

7.3.2   DESIGN CRITERIA

Our design criteria are functions of the information matrix of $\boldsymbol{\theta}$. We follow the technique to first linearize model (7.3) and then use the linearized model to obtain the approximated information matrix; see, e.g., Kempton et al. (2001) and Fedorov and Hackl (1997). This requires the differentiability of $\boldsymbol{H}(\boldsymbol{p})$ with respect to p_1 and p_6, which is verified in Appendix I.

The approximated information matrix of $\boldsymbol{\theta}$ is:

$$\mathcal{M}(\xi; \boldsymbol{\theta}, \boldsymbol{p}) = \boldsymbol{E}'\boldsymbol{E} - \boldsymbol{E}'\boldsymbol{L}(\boldsymbol{L}'\boldsymbol{L})^{-}\boldsymbol{L}'\boldsymbol{E}, \tag{7.5}$$

where

$$\boldsymbol{E} = (\boldsymbol{I}_T - P_{\boldsymbol{Vs}})\boldsymbol{V}\boldsymbol{X}\boldsymbol{H}(\boldsymbol{p})$$

$$\boldsymbol{L} = (\boldsymbol{I}_T - P_{\boldsymbol{Vs}})\boldsymbol{V}\boldsymbol{X}\left\{\left[\left(\frac{\partial \boldsymbol{H}(\boldsymbol{p})}{\partial p_1}\right), \left(\frac{\partial \boldsymbol{H}(\boldsymbol{p})}{\partial p_6}\right)\right](\boldsymbol{I}_2 \otimes \boldsymbol{\theta})\right\},$$

$P_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-}\boldsymbol{A}'$ is the orthogonal projection matrix onto the vector space spanned by the column vectors of $\boldsymbol{A}$, $\boldsymbol{A}^{-}$ is a generalized inverse matrix of $\boldsymbol{A}$, and $\boldsymbol{V}$ is a known matrix such that $\boldsymbol{V}\boldsymbol{\eta}$ is white noise. To evaluate the quality of a design, we consider the following

A-optimality criterion:

$$\Phi(\xi; \boldsymbol{\theta}, \boldsymbol{p}) = \frac{trace(\mathcal{M}^{-1}(\xi; \boldsymbol{\theta}, \boldsymbol{p}))}{Q}.$$

Designs minimizing this criterion minimize the average variance of estimators of the θ_is. In our simulation studies, we first find locally optimal designs minimizing $\Phi(\xi; \boldsymbol{\theta}_0, \boldsymbol{p}_0)$ under given $(\boldsymbol{\theta}_0, \boldsymbol{p}_0)$. These designs are desirable if the prior guess of the parameter values $(\boldsymbol{\theta}_0, \boldsymbol{p}_0)$ is accurate. However, they can be inefficient when the parameter values are mis-specified; see Section 7.4. With locally optimal designs, the relative efficiency of any other design, ξ, can be obtained:

$$RE(\xi; \boldsymbol{\theta}_0, \boldsymbol{p}_0) = \frac{\Phi(\xi^*(\boldsymbol{\theta}_0, \boldsymbol{p}_0); \boldsymbol{\theta}_0, \boldsymbol{p}_0)}{\Phi(\xi; \boldsymbol{\theta}_0, \boldsymbol{p}_0)},$$

where $\xi^*(\boldsymbol{\theta}_0, \boldsymbol{p}_0)$ is the locally optimal design obtained with given $(\boldsymbol{\theta}_0, \boldsymbol{p}_0)$. The relative efficiency reflects the performance of designs at $(\boldsymbol{\theta}_0, \boldsymbol{p}_0)$. Evaluated at different parameter values, RE is useful for measuring the quality of designs across the parameter space; see also Gotwalt et al. (2009) and Dette (1997).

To obtain designs that work relatively well across the parameter space, we also consider the pseudo-Bayesian criterion:

$$\Phi_I(\xi) = \int_{\mathcal{P}} \int_{\Theta} RE(\xi; \boldsymbol{\theta}, \boldsymbol{p}) \pi_1(\boldsymbol{\theta}) \pi_2(\boldsymbol{p}) d\boldsymbol{\theta} d\boldsymbol{p},$$

where Θ is the parameter space for $\boldsymbol{\theta}$, $\mathcal{P}$ is defined in (7.4), and $\pi_1(\cdot)$ and $\pi_2(\cdot)$ are given density functions for $\boldsymbol{\theta}$ and $\boldsymbol{p}$, respectively. Here, we follow de Pasquale et al. (2008) to assume independence of activation amplitudes ($\boldsymbol{\theta}$) and the HRF (characterized by $\boldsymbol{p}$). Finding designs maximizing this criterion is computationally expensive, and hence a surrogate criterion is commonly considered (e.g. Woods et al., 2006):

$$\bar{\Phi}(\xi) = \sum_{\boldsymbol{p} \in \mathcal{P}_d} \sum_{\boldsymbol{\theta} \in \Theta_d} RE(\xi; \boldsymbol{\theta}, \boldsymbol{p}) \pi_1^d(\boldsymbol{\theta}) \pi_2^d(\boldsymbol{p}).$$

Here, $\mathcal{P}_d$ and Θ_d are a finite subset of $\mathcal{P}$ and Θ, respectively, and $\pi_i^d(\cdot)$s are the corresponding probability mass functions. To further decrease the computational burden, we utilize the following theorem to reduce the parameter space Θ. The proof of the theorem is straightforward and is therefore omitted; see also Bose and Stufken (2007).

Theorem 7.3.1 $\mathcal{M}(\xi; a\boldsymbol{\theta}, \boldsymbol{p}) = \mathcal{M}(\xi; \boldsymbol{\theta}, \boldsymbol{p})$ *for all* $a \neq 0$.

7.3.3 SEARCH ALGORITHMS

Kao et al. (2009) propose a knowledge-based GA for finding optimal ER-fMRI designs. This algorithm starts with a set of initial designs containing well-known ER-fMRI designs and random designs. These designs are selected into a mating pool with probability proportional to their fitness (value of the design criterion). The offspring designs are produced by the selected designs via one-point crossover, which exchanges corresponding portions (starting from the first element to a randomly selected element) of paired designs. Mutation operator then perturbs a small number of elements in offspring designs. Immigrants including random designs and block designs of different block sizes are also introduced into the population. These immigrants not only help to jump out of local optima, but also bring in good traits that are not easily achievable by a random mechanism; they are designed to facilitate the search. Elites among parents, offsprings, and immigrants survive to form the next generation. The process is then repeated until a stopping rule is met.

One possible stopping rule is to terminate the search after a pre-specified number of generations. Another method stops the process when there is no significant improvement. Following Kao (2009), and Liefvendahl and Stocki (2006) we calculate the cumulative improvement, in terms of fitness, from the $\ell n + 1$st generation of the GA to the $(\ell + 1)n$th generation for a given integer n; $\ell = 1, 2, \dots$. When this amount is less than or equal to the cumulative improvement achieved by the first n generations, the search is terminated.

To increase the chance of utilizing good traits provided by immigrants, we use immigrants to replace the worst designs in the current generation before selecting designs into a mating pool. When implementing this GA in Section 7.4, the population size of the GA is set to 50, mutation rate is 1%, and number of immigrants is ten. The magnitude of the improvement in the design efficiency is obtained every $n = 200$ generations. If, at any stage, the cumulative

improvement is less than or equal to 10^{-7} of that over the first 200 generations, we stop the search.

We use this GA to find locally optimal designs minimizing $\Phi(\cdot)$, and pseudo-Bayesian designs maximizing $\bar{\Phi}(\cdot)$. While pseudo-Bayesian designs take into account the uncertainty of parameters, these designs might still fail to fit the researcher's prior knowledge about the values of the parameters. For example, when all parameter values are equally important (i.e., assigning a uniform prior distribution to the parameters), the pseudo-Bayesian designs obtained do not guarantee relative design efficiencies that are balanced over the entire parameter space. Instead, these designs might perform particularly well over certain parameter values, but are very inefficient over other regions of the parameter space (see, e.g., Figure 7.7 in the next section).

This phenomenon is linked to a well known limitation of the weighted sum method for solving MO optimization problems (Deb, 2001). The limitation lies in the unknown, complicated relationship between weights assigned to objectives and achieved values of objective functions relative to the best values. Assigning equal weights does not necessarily lead to a solution with balanced relative efficiencies, even when such a solution exists. Similar limitations apply to other weighting schemes.

In dealing with MO problems, one popular alternative is to create a set of solutions for researchers to choose from. The common goal of such approaches is to obtain solutions that are (1) Pareto-optimal, i.e. no other solution is better than, or dominates, the proposed solutions in all objectives, and (2) the proposed solutions are well spread over the frontier formed by Pareto-optimal solutions. Huang and Wong (1998) describe an approach for obtaining such solutions by the weighted sum method with uniformly spaced weights. Although this approach finds Pareto-optimal solutions, the solutions may fail to meet condition (2). A more promising approach to achieve both goals is to use multi-objective evolutionary algorithms (MOEAs; Shukla and Deb, 2007) for finding well-spread, Pareto-optimal solutions. Here, we adopt a popular algorithm of this sort, namely NSGA-II (Deb et al., 2002).

The driving engines of the NSGA-II are two sorting mechanisms — non-dominated sorting and crowding distance sorting. For a given class of solutions, denoted by O, the non-dominated sorting assigns the "non-domination levels" to the solutions. Starting with $i = 1$, it has the following two steps.

1. At the ith stage, identify the non-dominated set and the dominated set for O; no solutions in O dominate any solutions in the non-dominated set, and solutions in the dominated set are dominated by at least one other solution of O. Solutions in the non-dominated set for the current stage are said to be of the ith level.

2. Exclude solutions of level i from O, increase i by 1, and repeat the previous step.

The procedure stops until all solutions are assigned with a level.

The crowding distance sorting works on solutions of the same non-domination level. A "crowding distance" is assigned to each solution to indicate the distance of a solution from other solutions in the objective space, the space formed by values of objective functions. Appendix II provides details about these two sorting mechanisms.

The NSGA-II works similar to GAs, but, instead of achieving one solution only, it simultaneously obtains a group of Pareto-optimal solutions that are well spread in the frontier formed by all Pareto-optimal solutions. During a search, better solutions are those with small non-dominance levels. Within the same non-dominance level, solutions with larger crowding distances are desirable. Solutions with these features have greater chances of survival in the NSGA-II. This algorithm is described in Appendix II.

To use the NSGA-II, we first transform the optimal design problem to a MO optimization problem using the following proposed algorithm. This greedy algorithm helps to reduce the number of objective functions involved.

- Step 0. Set $\Lambda = \emptyset$, $\Omega_0 = \Theta_d \times \mathcal{P}_d$, $\Xi_0 = \Xi$ (a set of locally optimal designs, one for each element in Ω_0), and $\boldsymbol{\beta} = (\boldsymbol{\theta}, \boldsymbol{p})$.

- Step 1. Set a tolerance level, τ_p (e.g. 0.95). Find

$$\boldsymbol{\beta}_0 = arg \max_{\boldsymbol{\beta} \in \Omega_0} |\Delta_{\beta,\tau_p}|, \text{ where}$$

$$\Delta_{\beta,\tau_p} \equiv \{\boldsymbol{\beta}_1 \in \Omega_0 : RE(\xi^*(\boldsymbol{\beta}); \boldsymbol{\beta}_1) \geq \tau_p\},$$

and $|A|$ is the cardinality of the set A. If $\boldsymbol{\beta}_0$ is not unique, choose one so that the average of RE-values over Δ_{β,τ_p} is maximized.

- Step 2. Set $\Lambda = \Lambda \cup \{\boldsymbol{\beta}_0\}$, $\Omega_0 = \Omega_0 - \Delta_{\beta_0,\tau_p}$ and $\Xi_0 = \Xi_0 - \{\xi^*(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \Delta_{\beta_0,\tau_p}\}$.

- Step 3. Repeat Steps 1 and 2 until $\Omega_0 = \emptyset$.

To implement this algorithm, we form Ξ by obtaining a locally optimal design for each $\boldsymbol{\beta} \in \Omega_0$. After obtaining the final set Λ, we then use the elements of the set $\{RE(\cdot; \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Lambda\}$ as objective functions and form an MO problem with $M = |\Lambda|$ objectives. Well-spread, Pareto-optimal designs for this MO problem are then searched for by using the NSGA-II. Researchers can then choose suitable designs based on the needs and goals of an experiment. Note that, when $M = 1$, we achieve a design that is relatively efficient over the entire parameter space. The same algorithm can be applied to cases where $\Theta_d = \{0\}$, $\Theta_d = \{1\}$, or $\Theta_d = \{0, 1\}$.

7.4 Efficient Designs for One Stimulus Type

In this section, we find efficient ER-fMRI designs for model (7.3) when $\boldsymbol{\theta}$ is of main interest. We focus on $Q = 1$, where the design sequences consist of one type of stimulus and the control. Therefore, θ_1 is the only parameter of interest. Following Kao et al. (2009), the drift of the time series $\boldsymbol{Y}$ is a second-order Legendre polynomial, and $\boldsymbol{\eta}$ follows a stationary AR(1) process with a correlation coefficient of $\rho = 0.3$. The ISI is set to four seconds and TR to two seconds. For this ISI, the assumption of linear accumulations of the overlapping HRFs is likely to be valid. The length of the design sequence is $L = 255$, corresponding to

an experiment of 17 minutes. We note that, even for this simple case, there is no systematic study on efficient experimental designs for detection when the HRF is uncertain.

We use the GA described in Subsection 7.3.3 to search for locally optimal designs minimizing $\Phi(\xi; \theta_1, \boldsymbol{p})$ for $(\theta_1, \boldsymbol{p}) \in \Theta_d \times \mathcal{P}_d$. For $\mathcal{P}_d$, we consider a 10-by-10 regular grid on the smallest rectangle covering $\mathcal{P}$. Since the first point of the grid, $\boldsymbol{p} = (3.5, -2.5)'$, falls outside of $\mathcal{P}$, it is excluded and thus $|\mathcal{P}_d| = 99$. These 99 values are aligned in order, starting from $(3.5, -2.1)$ to $(11, 1)$, by first increasing the p_6-values. Due to Theorem 7.3.1, we can focus on $\Theta_d = \{0, 1\}$. A design that is optimal for $\theta_1 = 1$ is also optimal for any $\theta_1 \neq 0$, which corresponds to voxels that are activated or deactivated; see, e.g., Friston et al. (1998) for an example of deactivation. The case $\theta_1 = 0$ is linked to non-active voxels.

For each $(\theta_1, \boldsymbol{p}) \in \Theta_d \times \mathcal{P}_d$, we perform the GA ten times with different random seeds to obtain ten designs. Figure 7.1 presents the Φ-values of the designs obtained from these GA runs. For each $(\theta_1, \boldsymbol{p})$, the Φ-values of the ten designs are virtually identical, indicating the stability of the GA. Among the ten designs, we select the best design as the locally optimal design for further study. Consequently, a total of 198 locally optimal designs are obtained. These designs are investigated.

When finding pseudo-Bayesian designs as described, which requires locally optimal designs, we only use the locally optimal designs obtained from the first runs of the GA. This reflects a practical situation where only one GA run is performed for each parameter value.

All computations in this section are conducted by using MATLAB (version 7.3) on a Linux cluster with 64-bit AMD Opteron, dual-processor, mix of single-core node and dual-core node; each core has 2GB RAM and the Linux operating system is 2.6.9-78.0.5.ELsmp.

7.4.1 ROBUSTNESS WITH RESPECT TO MIS-SPECIFYING θ_1

We calculate, for each locally optimal design, the relative efficiency with respect to a mis-specification of θ_1. That is, for designs with $\theta_1 = 1$, denoted by $\xi^*(\theta_1 = 1, \boldsymbol{p})$, we find $RE(\xi^*(\theta_1 = 1, \boldsymbol{p}); \theta_1 = 0, \boldsymbol{p})$. The value of $RE(\xi^*(\theta_1 = 0, \boldsymbol{p}); \theta_1 = 1, \boldsymbol{p})$ is also obtained.
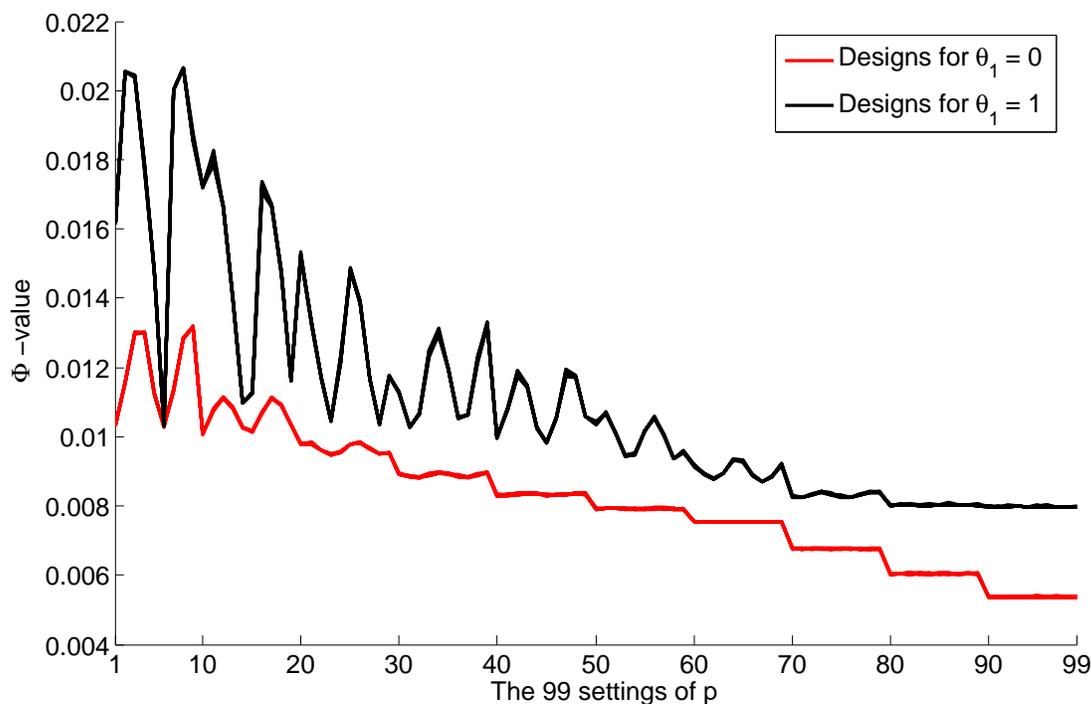
Figure 7.1: The Φ-values of the ten designs obtained by the ten GA runs for each given $(\theta_1, \boldsymbol{p})$

Figure 7.2 presents these relative efficiencies against the 99 values of $\boldsymbol{p}$. Designs for $\theta_1 = 1$ are generally more robust to a mis-specification of θ_1 than designs for $\theta_1 = 0$. A relative efficiency of at least 67% is attained by $\xi^*(\theta_1 = 1, \boldsymbol{p})$ under $(\theta_1 = 0, \boldsymbol{p})$. Whereas $RE(\xi^*(\theta_1 = 0, \boldsymbol{p}); \theta_1 = 1, \boldsymbol{p})$ can be as small as 3%.

Figure 7.2 indicates the importance of the designs with $\theta_1 = 1$. These designs are especially useful when researchers would like to have precise inference on voxels that are activated or deactivated. The case where $\theta_1 = 0$ is also of interest. For this particular case, the approximated information matrix of (7.5) is the same as the information matrix of θ_1 under
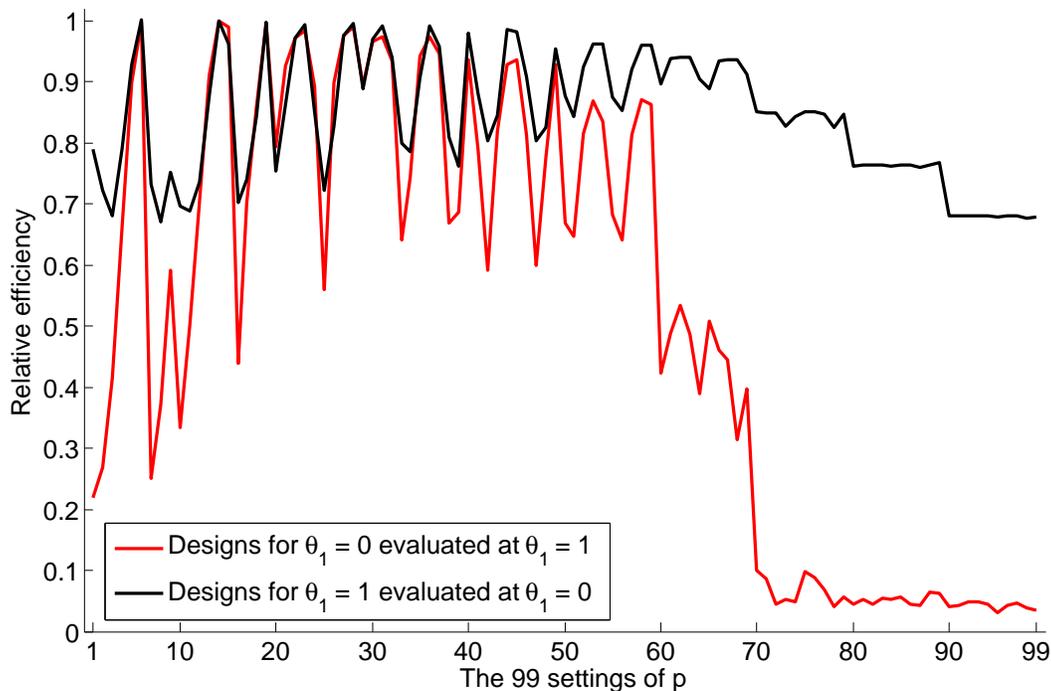
Figure 7.2: Relative efficiencies of the locally optimal designs under mis-specification of θ_1

the linear model (7.2). Studying this case also provides information about the change in the design efficiency for detection with respect to a mis-specification of $\boldsymbol{p}$ under the linear model. As mentioned previously, these designs are the best for making precise inference on non-active voxels.

In the following subsections, we study these two cases separately. As discussed in Section 7.5, our approach can also be applied to the situation where both $\theta_1 = 0$ and 1 are of interest.
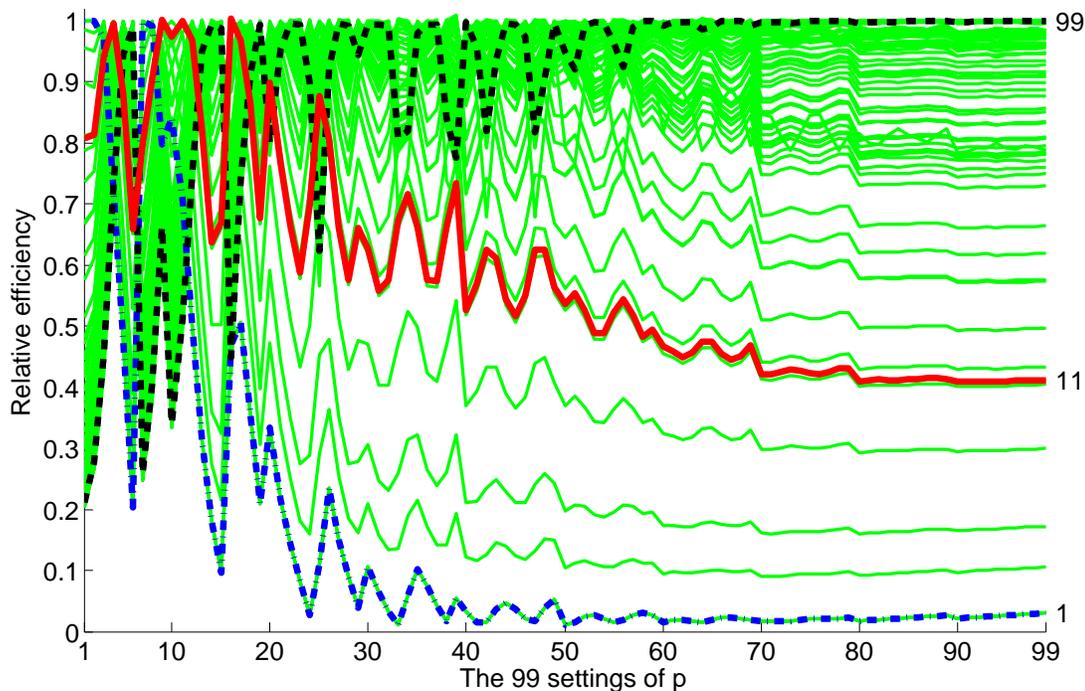
Figure 7.3: Relative efficiencies of the 99 locally optimal designs for $\theta_1 = 1$ evaluated at all of the 99 $\boldsymbol{p}$s of $\mathcal{P}_d$

7.4.2 Designs for $\theta_1 = 1$

Figure 7.3 presents the relative efficiencies, $RE(\xi^*(\theta_1 = 1, \boldsymbol{p}_0); \theta_1 = 1, \boldsymbol{p})$, of each of the 99 locally optimal designs evaluated at all of the 99 values of $\boldsymbol{p} \in \mathcal{P}_d$. For illustration, three designs are selected, which are the first ($\boldsymbol{p}_0 = (3.5, -2.1)$), 11th ($\boldsymbol{p}_0 = (4.3, -2.1)$) and last ($\boldsymbol{p}_0 = (11, 1)$) designs. These designs are presented in Figure 7.4 with black bars indicating 1s and white bars representing 0s. The first locally optimal design (the top design in Figure 7.4) consists of repetitions of $\{1\ 0\}$. As shown in Figure 7.3, this design does not perform

well when p_1 is large. On the other hand, the last design (the bottom design in Figure 7.4) is inefficient at places where the first design works well, but is efficient after the 60th setting of $\boldsymbol{p}$. This last design looks similar to a block design with a block size of three. The 11th design, which looks like a combination of the first and the last design, works well at some places where neither the first nor last design have a high relative efficiency.
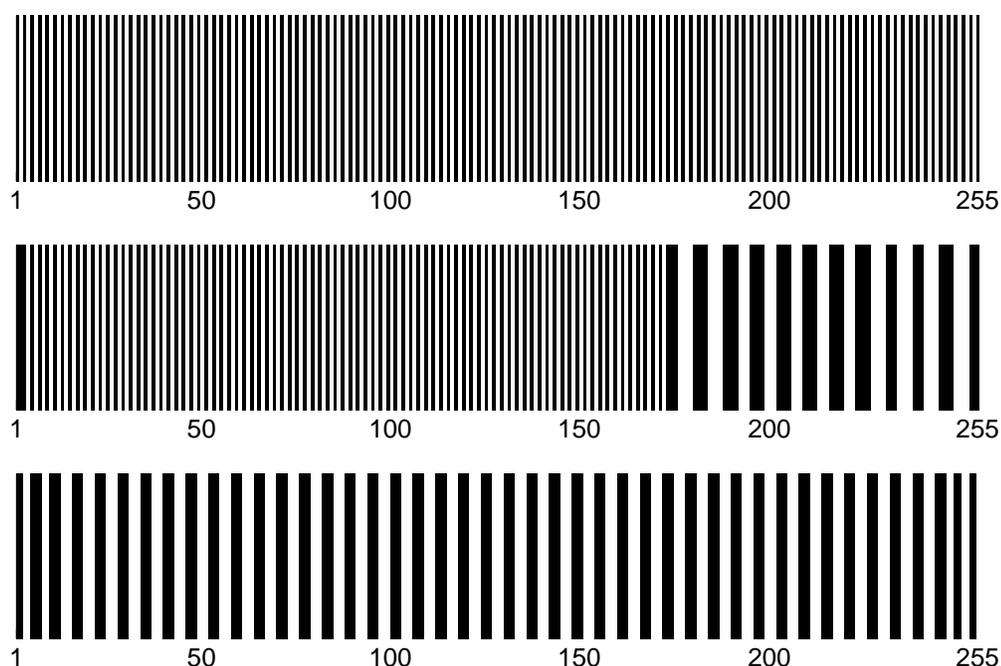
Figure 7.4: The first, eleventh and last locally optimal designs for $\theta_1 = 1$

These three designs are illustrative of the form of locally optimal designs. Thus characteristics of good designs are observed to be different across subregions of the parameter space. Our proposed approach helps to find efficient designs achieving advantageous compromises among these "competing" subregions. The greedy algorithm (Subsection 7.3.3) identifies $\boldsymbol{p}$s, at which designs performing relatively well over certain subregions can be generated; the

union of these subregions covers the entire $\mathcal{P}$. Using these values of $\boldsymbol{p}$ to form multiple objective functions, an MO problem is formulated and NSGA-II is used to search for well-spread, Pareto-optimal solutions (designs).

The greedy algorithm takes only a few seconds to identify four parameter values, which form an MO problem with four objectives. The NSGA-II with the settings described in Appendix II spends a CPU time of about six minutes to find 100 Pareto-optimal designs for these four objective functions. The relative efficiencies, with respect to the 99 values of $\boldsymbol{p}$, of the designs found by NSGA-II are presented in Figure 7.5. To facilitate the selection of designs, we pick four reference designs in Figure 7.5. These designs are selected based on the relative efficiency at the last $\boldsymbol{p} = (11, 1)$. Design 1 is relatively efficient for large p_1-value. This design works similar to a pseudo-Bayesian design maximizing $\bar{\bar{\Phi}}$ with a uniform prior over $\mathcal{P}_d$ and $\Theta_d = \{1\}$, which corresponds to the curve denoted by 'B' in Figure 7.5. Design 1 and the pseudo-Bayesian designs work well over a large subregion of the parameter space. When other parameter values are more important, we would certainly consider other designs such as the reference designs 2 or 3 that are easily obtained by our approach.

7.4.3 DESIGNS FOR $\theta_1 = 0$

The same approach as in the previous subsection is applied to find efficient designs for $\theta_1 = 0$. Similar to Figure 7.3, Figure 7.6 presents the relative efficiencies $RE(\xi^*(\theta_1 = 0, \boldsymbol{p}_0); \theta_1 = 0, \boldsymbol{p})$ for $\boldsymbol{p} \in \mathcal{P}_d$. This figure also provides information about the sensitivity in the design efficiency with respect to mis-specification of p_1 and p_6-values under the linear model (7.2).

For convenience, $\mathcal{P}_d$ is divided into three subsets: $\mathcal{P}_d^1$ contains the first 59 values of $\boldsymbol{p}$, $\mathcal{P}_d^2$ includes the 60th to the 69th values, and $\mathcal{P}_d^3$ the remaining elements. From Figure 7.6, the first 59 designs work relatively well over $\mathcal{P}_d^1$. However, their relative efficiencies can be smaller than 80% over $\mathcal{P}_d^2$ and $\mathcal{P}_d^3$. On the other hand, designs that are suitable for large p_1-values ($\mathcal{P}_d^3$) can be inefficient for small p_1. A closer look at these locally optimal designs reveals that the first 59 designs consist of blocks of sizes of about four ($=16/ISI$). The locally
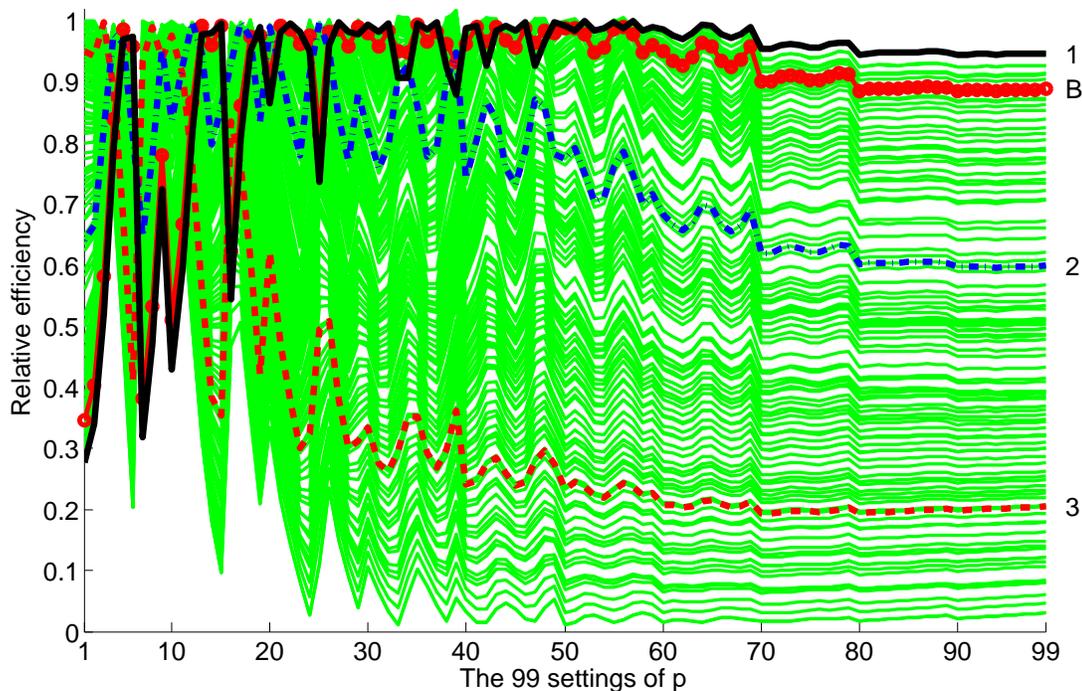
Figure 7.5: Relative efficiencies of the 100 designs obtained by NSGA-II for $\theta_1 = 1$ evaluated at all of the 99 $\boldsymbol{p}$s of $\mathcal{P}_d$

optimal designs from $\mathcal{P}_d^2$ have a slightly larger block size of about five. The pattern of the locally optimal designs for large p_1-values is more complicated; these designs include blocks of different sizes. The design that achieves a relative efficiency above 99% over $\mathcal{P}_d^3$ is formed by 43 ones followed by 85 zeros, 85 ones and 42 zeros. A numerical study (not shown) indicates that this latter design and all locally optimal designs obtained here are nearly trend-resistant for a quadratic trend (Afsarinejad, 2001; Cox, 1951). This is evidenced by the fact that the information matrices of these designs are nearly identical between the model with a second-order drift and the model with no drift or trend. It is noteworthy that the locally optimal
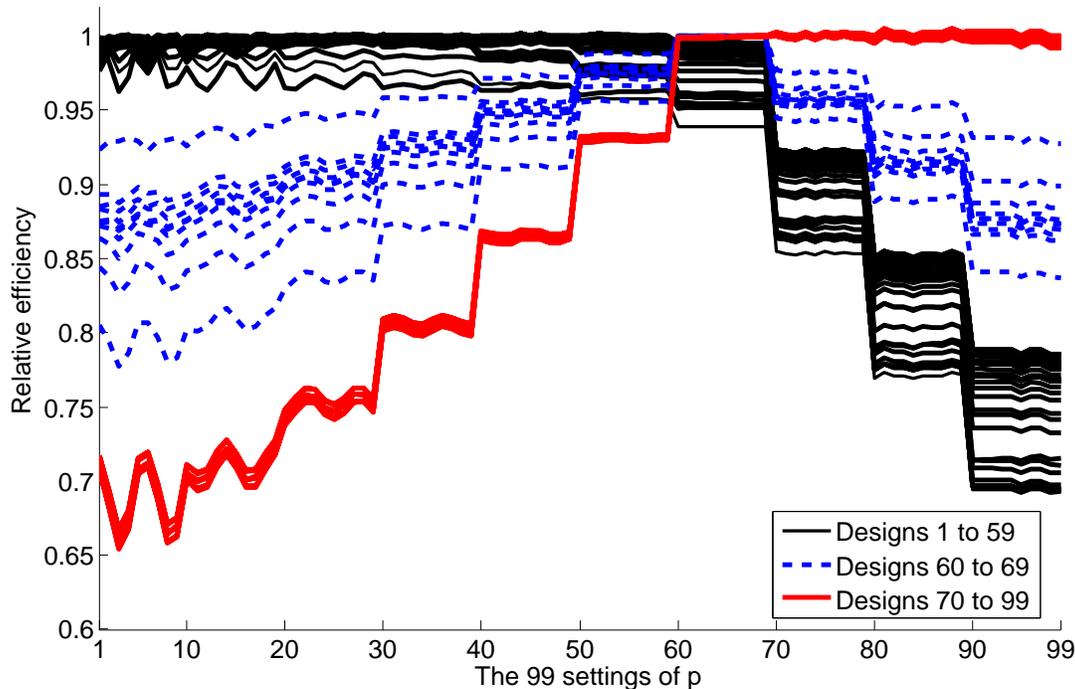
Figure 7.6: Relative efficiencies of the 99 locally optimal designs for $\theta_1 = 0$ evaluated at all of the 99 $\boldsymbol{p}$s of $\mathcal{P}_d$

designs obtained from the model with no drift or trend are not trend-resistant for a quadratic trend, especially when $\boldsymbol{p} \in \mathcal{P}_d^3$.

The greedy algorithm identifies two objective functions, namely $RE(\xi; \theta_1 = 0, \boldsymbol{p} = (3.5, -1.72))$ and $RE(\xi; \theta_1 = 0, \boldsymbol{p} = (8.5, 1))$. The NSGA-II is then used to find 100 Pareto-optimal designs for these objective functions. From Figure 7.7, these diverse designs achieve intermediate efficiencies between the two 'competing' subregions, $\mathcal{P}_d^1$ and $\mathcal{P}_d^3$ and are highly efficient over $\mathcal{P}_d^2$. Here, the NSGA-II spends a CPU time of about five minutes.
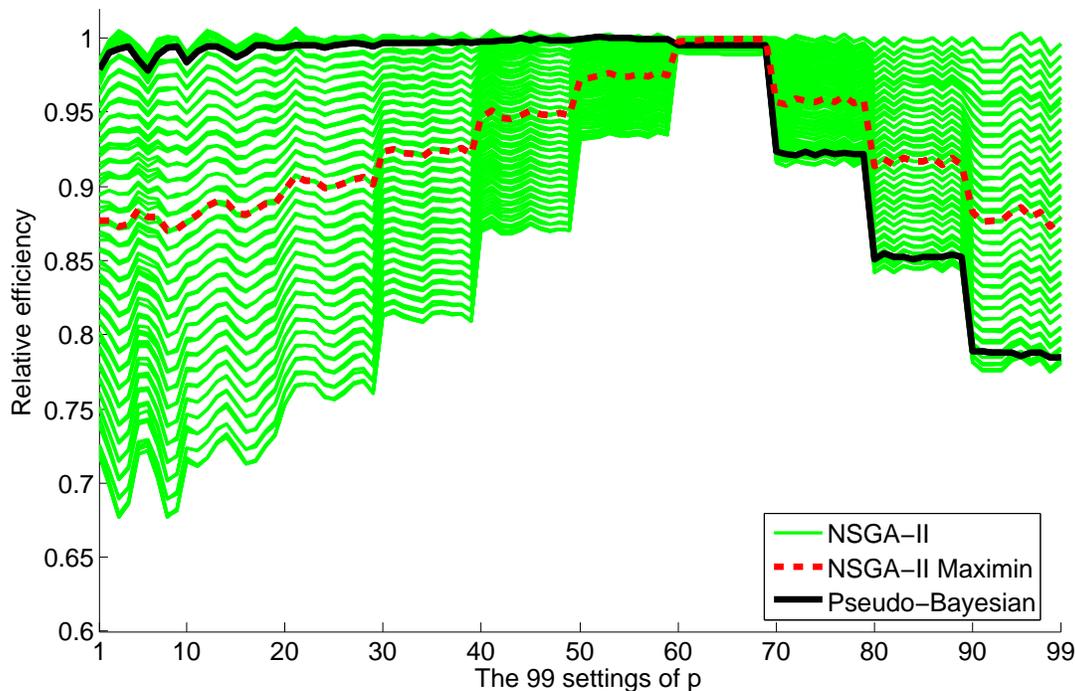
Figure 7.7: Relative efficiencies of the 100 designs obtained by NSGA-II for $\theta_1 = 0$ evaluated at all of the 99 $\boldsymbol{p}$s of $\mathcal{P}_d$

A pseudo-Bayesian design is also found by using the GA to maximize the $\bar{\Phi}$-value under a uniform prior over $\mathcal{P}_d$ ($\Theta_d = \{0\}$). The relative efficiency of this design over $\mathcal{P}_d$ is also presented in Figure 7.7. Although equal weights are assigned to the 99 values of $\boldsymbol{p}$, the resulting design is much more efficient in $\mathcal{P}_d^1$ than in $\mathcal{P}_d^3$. We would, instead, suggest the reference design in Figure 7.7 when all the parameter values are equally important. This reference design maximizes the minimal relative efficiency and it achieves at least 87% relative efficiencies over $\mathcal{P}_d$. Using our approach, other designs can also be chosen based on the goals and needs of the experiment.

7.5 Conclusion and Discussion

Our study focuses on finding efficient ER-fMRI designs with a nonlinear model. With this nonlinear model, we take into account the uncertainty in the HRF for selecting good designs for detection problems. This consideration is crucial since, typically, the HRF is not fully known to researchers. In addition, the HRF can change across voxels. Although an approach for finding designs that help render efficient statistical inference across different voxels is valuable, there is no previous study addressing this issue.

We derive the design criteria based on the nonlinear model. A knowledge-based GA is adopted to search for locally optimal designs and pseudo-Bayesian designs. We observe that the pseudo-Bayesian design may not always match the researcher's prior knowledge about the importance of the parameter values. This is linked to a limitation of the popular weighted sum method for MO optimization problems. To remedy this, we propose an approach to search for a set of well-spread, Pareto-optimal designs for researchers to choose from based on their goals and needs.

In our case study, we consider the design problem with one stimulus type. The parameter space for the parameter of interest (θ_1) is reduced to two points, 0 and 1. While zero only stands for one parameter value, the design efficiency calculated with $\theta_1 = 1$ is invariant over the entire real line, except for the origin. Both cases are studied and the results learned from $\theta_1 = 0$ can also be applied to the linear model framework. When there is a need to consider both cases, our approach can still work by considering $\Theta \times \mathcal{P}$ altogether. For this latter case, six objectives are identified when forming the MO problem, and a set of designs is obtained. The larger parameter space does bring in complexity and we find it useful to use reference designs (as those presented in Figure 7.5). A more sophisticated method for choosing designs is a future research of our interest. A better idea about the range of plausible parameter values is definitely helpful.

Here, we show the existence of $\partial \boldsymbol{H}(\boldsymbol{p})/\partial p_1$ and $\partial \boldsymbol{H}(\boldsymbol{p})/\partial p_6$, where $\boldsymbol{H}(\boldsymbol{p})$ of model (7.3) can be taken as:

$$\boldsymbol{H}(\boldsymbol{p}) = \boldsymbol{I}_Q \otimes \begin{pmatrix} g(\tau_1; \boldsymbol{p}) \\ g(\tau_2; \boldsymbol{p}) \\ \vdots \\ g(\tau_k; \boldsymbol{p}) \end{pmatrix},$$

where $g(\cdot; \boldsymbol{p})$ is defined in (7.1), $\boldsymbol{p} = (p_1, p_6)$ is a parameter vector, and $\tau_j = (j-1)\Delta T$; $j = 1, ..., k$. The following lemma ensures that $g(\tau; \boldsymbol{p})$ is well defined.

Lemma 1 *The function $g(\tau; \boldsymbol{p})$ is well defined; i.e., $\max\limits_s h(s; \boldsymbol{p}) > 0$, where $h(\cdot; \boldsymbol{p})$ is defined under (7.1).*

The lemma follows from observing that,

$$h(\tau = p_6 + \epsilon; \boldsymbol{p}) = \frac{(\epsilon)^{p_1-1} e^{-\epsilon}}{\Gamma(p_1)} - \frac{(\epsilon)^{15} e^{-\epsilon}}{6\Gamma(16)} > 0 \text{ when}$$

$$\epsilon \in (0, \min\{(32 - p_6), \big(\frac{\Gamma(p_1)}{6\Gamma(16)}\big)^{1/(p_1-16)}\}).$$

The maximal value of $h(\tau; \boldsymbol{p})$ does not depend on p_6; see also Wager et al. (2005). The existence of $\partial \boldsymbol{H}(\boldsymbol{p})/\partial p_6$ then follows from the partially differentiability of $h(\tau; \boldsymbol{p})$ with respect to p_6. Without loss of generality, we assume $p_6 = 0$ and omit it from $h(.)$ hereinafter.

For $\partial \boldsymbol{H}(\boldsymbol{p})/\partial p_1$, we adopt the Danskin's theorem addressed in Dem'yanov and Malozemov (1974, Chapter VI, Theorem 2.1):

Theorem 1 *(Danskin's theorem) Let $\mathcal{T}$ be compact and $\mathcal{P}_1$ be an open set. If $h(\tau; p_1)$ is continuous and continuously differentiable in p_1 on $\mathcal{T} \times \mathcal{P}_1$, then $\phi(p_1) \equiv \max\limits_{\tau \in \mathcal{T}} h(\tau; p_1)$ is differentiable at each point of $\mathcal{P}_1$ in any direction $\eta \in \mathbb{R}$ ($||\eta|| = 1$). Specifically,*

$$\frac{\partial \phi(p_1)}{\partial \eta} = \max_{\tau \in R_0} \big\langle \frac{\partial h(\tau, p_1)}{\partial p_1}, \eta \big\rangle$$

$$R_0 = \{\tau \in \mathcal{T} : h(\tau, p_1) = \phi(p_1) \equiv \max_{s \in \mathcal{T}} h(s, p_1)\},$$

where $< \cdot, \cdot >$ is the inner product.

We take $\mathcal{T}$ to be $[0, T_{\max}]$, where $T_{\max}$ is the experimental duration, and $\mathcal{P}_1$ is the smallest open set covering the valid range of p_1 in $\mathcal{P}$. The continuity, and continuously differentiability of $h(.)$ with respect to p_1 hold and the theorem ensures the directional differentiability of $h(.)$, and hence $g(.)$. When R_0 contains only one point, the usual differentiability of $h(.)$ with respect to p_1 holds; see also, Danskin (1967).

Lemma 2 $|R_0| = 1$; i.e., $\arg\max\limits_{s \in \mathcal{T}} h(s; p_1)$ is unique.

The first derivative of $h(\tau; p_1)$ with respect to τ is:

$$\frac{\partial h(\tau; p_1)}{\partial \tau} = e^{-\tau} \left[\frac{\tau^{p_1 - 2}(p_1 - 1 - \tau)}{\Gamma(p_1)} - \frac{\tau^{14}(15 - \tau)}{6\Gamma(16)} \right].$$

Finding $|R_0|$ is equivalent to finding positive roots of this function of τ. This is equivalent to finding $\tau > 0$ such that

$$\frac{6\Gamma(16)}{\Gamma(p_1)}(p_1 - 1 - \tau) = \tau^{16 - p_1}(15 - \tau).$$

The function on the left hand side is a linear function with a negative slope. The function on the right hand side increases from zero, peaks at $\tau = 15(16 - p_1)/(17 - p_1)$ and is concave decreasing afterward. These two functions can intersect at one or two points. Therefore, $|R_0| = 1$.

APPENDIX II

We describe the procedure of NSGA-II. We describe first the way to form a non-dominated set and assign crowding distances to designs.

FORMING A NON-DOMINATED SET

The following steps are used to form the non-dominated set Q_i from the set O containing designs being sorted. We label the designs in O from one to $|O|$. The M design criteria $\{RE(\xi; \boldsymbol{\beta}) : \boldsymbol{\beta} \in \Lambda\}$ are obtained from the greedy algorithm.

1. For each design ξ, obtain the value of each individual criterion: $(RE(\xi; \boldsymbol{\beta}_1), ..., RE((\xi; \boldsymbol{\beta}_M))$; $\boldsymbol{\beta}_j \in \Lambda$, $j = 1, ... M$.

2. Set $p = 1$ and let $P^* = \{\xi_1\}$; ξ_i is the first design in O.

3. Increase p to $p + 1$ and use $P^* \cup \xi_p$ to replace P^*, where ξ_p is the pth design in O.

4. For each $\xi_q \in P^* - \{\xi_p\}$:

   if $\xi_q \succ \xi_p$ (ξ_q dominates ξ_p), $P^* = P^* - \{\xi_p\}$, go to 3;

   else if $\xi_p \succ \xi_q$, $P^* = P^* - \{\xi_q\}$.

   We said that ξ_q dominates ξ_p if $RE(\xi_q; \boldsymbol{\beta}_j) > RE(\xi_p; \boldsymbol{\beta}_j)$ for all $j = 1, ..., M$.

5. If $p = |O|$, set $Q_i = P^*$ and stop. Otherwise, go to 3.

ASSIGNING CROWDING DISTANCES

The crowding distance assignment procedure for a set Q_i is:

1. Assign d_j (distance) $= 0$ for each $\xi_j \in Q_i$; $j = 1, ..., |Q_i|$.

2. Based on each $RE_{(\xi; \boldsymbol{\beta}_m)}$, sort the set Q_i and denote the sorted set by $Q_i^{(m)}$; $m = 1, ..., M$

3. Assign a large distance, say ∞ to the first and the last design in $Q_i^{(m)}$. For the jth design $\xi_{(j)}$ in $Q_i^{(m)}$, assign:

$$d_{I_j^m} = d_{I_j^m} + RE(\xi_{(j+1)}; \boldsymbol{\beta}_m) - RE(\xi_{(j-1)}; \boldsymbol{\beta}_m),$$

where I_j^m is the original index in Q_i of the design $\xi_{(j)}$; $j = 2, ..., |Q_i| - 1$ and $m = 1, ..., M$.

NSGA-II

The procedure of the NSGA-II is presented below. As mentioned previously, designs are better if they have lower non-domination level. If designs are of the same non-domination level, better designs are those with larger crowding distances.

- Step 0. Form a set of $2G$ initial designs, denoted by R_0, and set $t = 1$.

- Step 1. Form the set P_t of G better designs:

  - Step 1A. Set $i = 1$, $O = R_{t-1}$, and $P_t = \emptyset$.

  - Step 1B. Form the non-dominated set Q_i from O.

  - Step 1C. If $|P_t| + |Q_i| > G$, sort the designs of Q_i based on the non-domination level and crowding distance, and include the best $G - |P_t|$ designs to P_t. Otherwise, replace P_t by $(P_t \cup Q_i)$.

  - Step 1D. If $|P_t| = G$, stop and go to Step 2. Otherwise, $O = O - Q_i$, set $i = i + 1$ and go to Step 1B.

- Step 2. From P_t, select at random G pairs of designs with replacement. For each pair, use a tournament selection that keeps the better design to select G parents. Perform one-point crossover and mutation to generate G offspring designs.

- Step 3. $t = t + 1$, repeat Steps 1 and 2. and stop when a stopping rule is met.

The population size G used for this study is 100, the mutation rate is 1% and the search stops after 500 generations.

INITIAL DESIGNS

For initial designs, we include well-known fMRI designs and random designs. The m-sequence-based designs included are m-sequences or designs constructed from m-sequences (Barker, 2004; Godfrey, 1993; MacWilliams and Sloane, 1977). They are introduced to fMRI by Buračas and Boynton (2002) to facilitate the estimation of the HRF. Mixed designs are formed by concatenating a fraction of a block design with a fraction of an m-sequence-based design (or a random design). These designs are studied by Liu (2004) and they achieve intermediate efficiencies between detection and estimation.

7.6 References

Afsarinejad, K. (2001), "Trend-Free Repeated Measurement Designs," in *Moda 6 — Advances in Model-Oriented Design and Analysis*, eds. Atkinson, A. C., Hackl, P., and M uller, W. G., New York: Springer, pp. 1–12.

Atkinson, A. C., Donev, A., and Tobias, R. (2007), *Optimum Experimental Designs, with SAS*, Great Britain: Oxford U Pr.

Barker, H. A. (2004), "Primitive Maximum-Length Sequences and Pseudo-Random Signals," *Transactions of the Institute of Measurement & Control*, 26, 339–348.

Bose, M. and Stufken, J. (2007), "Optimal Crossover Designs When Carryover Effects Are Proportional to Direct Effects," *Journal of Statistical Planning and Inference*, 137, 3291–3302.

Buračas, G. T. and Boynton, G. M. (2002), "Efficient Design of Event-Related fMRI Experiments Using M-Sequences," *NeuroImage*, 16, 801–813.

Cabeza, R. and Kingstone, A. (2006), *Handbook of Functional Neuroimaging of Cognition*, Cognitive Neuroscience, Cambridge, Mass.: MIT Press, 2nd ed.

Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Planning and Inference*, 21, 191–208.

Chaloner, K. and Verdinelli, I. (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, 10, 273–304.

Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics*, 24, 586–602.

Cox, D. R. (1951), "Some Systematic Experimental Designs," *Biometrika*, 38, 312–323.

Dale, A. M. and Buckner, R. L. (1997), "Selective Averaging of Rapidly Presented Individual Trials Using fMRI," *Human Brain Mapping*, 5, 329–340.

Danskin, J. M. (1967), *The Theory of Max-Min and Its Application to Weapons Allocation Problems*, konometrie Und Unternehmensforschung, Berlin, New York: Springer-Verlag.

Deb, K. (2001), *Multi-Objective Optimization Using Evolutionary Algorithms*, Wiley-Interscience Series in Systems and Optimization, Chichester ; New York: John Wiley & Sons, 1st ed.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002), "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, 6, 182–197.

Dem'yanov, V. F. and Malozemov, V. N. (1974), *Introduction to Minimax*, New York,: Wiley.

de Pasquale, F., Del Gratta, C., and Romani, G. L. (2008), "Empirical Markov Chain Monte Carlo Bayesian Analysis of fMRI Data," *NeuroImage*, 42, 99–111.

Dette, H. (1997), "Designing Experiments with Respect to 'Standardized' Optimality Criteria," *Journal of the Royal Statistical Society Series B-Methodological*, 59, 97–110.

Dror, H. A. and Steinberg, D. M. (2006), "Robust Experimental Design for Multivariate Generalized Linear Models," *Technometrics*, 48, 520–529.

Fedorov, V. V. and Hackl, P. (1997), *Model-Oriented Design of Experiments*, New York: Springer.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., and Turner, R. (1998), "Event-Related fMRI: Characterizing Differential Responses," *NeuroImage*, 7, 30–40.

Godfrey, K. (1993), *Perturbation Signals for System Identification*, New York: Prentice Hall.

Gotwalt, C. M., Jones, B. A., and Steinberg, D. M. (2009), "Fast Computation of Designs Robust to Parameter Uncertainty for Nonlinear Settings," *Technometrics*, 51, 88–95.

Handwerker, D. A., Ollinger, J. M., and D'Esposito, M. (2004), "Variation of BOLD Hemodynamic Responses across Subjects and Brain Regions and Their Effects on Statistical Analyses," *Neuroimage*, 21, 1639–1651.

Henson, R. N. A. (2007), "Efficient Experimental Design for fMRI," in *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, eds. Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and D., P. W., London: Academic, 1st ed., pp. 193–210.

Huang, Y. C. and Wong, W. K. (1998), "Sequential Construction of Multiple-Objective Optimal Designs," *Biometrics*, 54, 1388–1397.

Kao, M.-H. (2009), "Multi-Objective Optimal Experimental Designs for ER-fMRI Using MATLAB," *Journal of Statistical Software*, to appear.

Kao, M.-H., Mandal, A., Lazar, N., and Stufken, J. (2009), "Multi-Objective Optimal Experimental Designs for Event-Related fMRI Studies," *NeuroImage*, 44, 849–856.

Kao, M.-H., Mandal, A., and Stufken, J. (2009), "Optimal Design for Event-Related Functional Magnetic Resonance Imaging Considering Both Individual Stimulus Effects and Pairwise Contrasts," *Statistics and Applications (Special Volume in Honour of Professor Aloke Dey)*, to appear.

Kempton, R. A., Ferris, S. J., and David, O. (2001), "Optimal Change-over Designs When Carry-over Effects Are Proportional to Direct Effects of Treatments," *Biometrika*, 88, 391–399.

Lazar, N. A. (2008), *The Statistical Analysis of Functional MRI Data*, Statistics for Biology and Health, New York: Springer.

Liefvendahl, M. and Stocki, R. (2006), "A Study on Algorithms for Optimization of Latin Hypercubes," *Journal of Statistical Planning and Inference*, 136, 3231–3247.

Lindquist, M. A. (2008), "The Statistical Analysis of fMRI Data," *Statistical Science*, 23, 439–464.

Lindquist, M. A. and Wager, T. D. (2007), "Validity and Power in Hemodynamic Response Modeling: A Comparison Study and a New Approach," *Human Brain Mapping*, 28, 764–784.

Liu, T. T. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part II: Design of Experiments," *NeuroImage*, 21, 401–413.

Liu, T. T. and Frank, L. R. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part I: Theory," *NeuroImage*, 21, 387–400.

Liu, T. T., Frank, L. R., Wong, E. C., and Buxton, R. B. (2001), "Detection Power, Estimation Efficiency, and Predictability in Event-Related fMRI," *NeuroImage*, 13, 759–773.

Lu, Y., Bagshaw, A. P., Grova, C., Kobayashi, E., Dubeau, F., and Gotman, J. (2006), "Using Voxel-Specific Hemodynamic Response Function in EEG-fMRI Data Analysis," *NeuroImage*, 32, 238–247.

MacWilliams, F. J. and Sloane, N. J. A. (1977), *The Theory of Error Correcting Codes*, Amsterdam ; New York New York: North-Holland Pub. Co. ; sole distributors for the U.S.A. and Canada, Elsevier/North-Holland.

Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., and Ciuciu, P. (2008), "A Fully Bayesian Approach to the Parcel-Based Detection-Estimation of Brain Activity in fMRI," *NeuroImage*, 41, 941–969.

Menz, M. M., Neumann, J., Muller, K., and Zysset, S. (2006), "Variability of the BOLD Response over Time: An Examination of within-Session Differences," *NeuroImage*, 32, 1185–1194.

Shukla, P. K. and Deb, K. (2007), "On Finding Multiple Pareto-Optimal Solutions Using Classical and Evolutionary Generating Methods," *European Journal of Operational Research*, 181, 1630–1652.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of Experimental Design in fMRI: A General Framework Using a Genetic Algorithm," *NeuroImage*, 18, 293–309.

Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. (2005), "Accounting for Nonlinear BOLD Effects in fMRI: Parameter Estimates and a Model for Prediction in Rapid Event-Related Studies," *NeuroImage*, 25, 206–218.

Woods, D. C., Lewis, S. M., Eccleston, J. A., and Russell, K. G. (2006), "Designs for Generalized Linear Models with Several Variables and Model Uncertainty," *Technometrics*, 48, 284–292.

Zhang, C. M. and Yu, T. (2008), "Semiparametric Detection of Significant Activation for Brain fMRI," *Annals of Statistics*, 36, 1693–1725.

CHAPTER 8

CONCLUSION

The rather complex experimental settings and assumptions make finding optimal ER-fMRI designs a fertile, challenging research area. Due to the popularity and high cost of ER-fMRI experiments, designs that help render valid and precise statistical inference are in great demand, and an efficient approach for obtaining them is called for.

Here, we develop an efficient approach for finding optimal ER-fMRI designs. Our first focus is on the popular linear model framework that involves two linear models for the two common statistical objectives, namely estimation and detection. We rigorously formulate our statistical models and define design criteria. We search for designs optimizing the design criterion via a knowledge-based GA that we propose. Our GA makes use of well known ER-fMRI designs and it is shown to outperform other methodologies. A MATLAB program implementing this algorithm is also developed.

When focusing on detecting activation, our GA obtains designs similar to block designs, which are well known good designs for detection. While block designs are included as part of initial designs of our GA, the designs that we obtain do not necessarily look similar to the initial ones. For example, our GA achieves a design sequence having no zeros (the control) when focusing only on pairwise contrasts between stimulus types, whereas all initial block designs contain zeros.

When aiming at estimating the HRFs, our GA obtains designs that are better than m-sequences. It is known that m-sequences can yield high estimation efficiencies for a particular situation; i.e., the response is with no trend or drift, the noise is uncorrelated and the interest

lies only in individual stimulus effects. Moving away from this rather uncommon situation, we recommend using our GA to search for good designs.

In addition, we demonstrate that our GA can achieve designs having advantageous compromises between the two competing statistical objectives. These designs are especially useful when both estimation and detection are of interest, and they outperform other designs currently in use by researchers. Our GA can also accommodate psychological constraints and customized requirements when searching for optimal ER-fMRI designs. We show that our GA is much better than its main competitor, the GA of Wager and Nichols (2003), in terms of CPU time and achieved design efficiencies when statistical and practical issues are considered simultaneously .

We adopt our approach to the case where both individual stimulus effects and pairwise contrasts are of interest. Optimal designs are searched for with different weights assigned to these two common interests of fMRI researchers. We observe that designs for individual stimulus effects retain a reasonable efficiency when the interest lies only in pairwise contrasts; they are robust with respect to a change in interests. On the other hand, designs that are optimal for pairwise contrasts can be inefficient for estimating individual stimulus effects. When focusing on detection problems, we also note that the block designs achieved under Case I (white noise and the response has no trend or drift) have block sizes smaller than those obtained under Case II (correlated noise and the response has a second-order polynomial drift). Similar observations are also discussed in Henson et al. (2002).

When researchers are more interested in pairwise contrasts, the stimulus frequency (the relative frequency of each stimulus type in a design sequence) of an efficient ER-fMRI design is larger; the number of zeros in the design sequence is thus smaller. The stimulus frequency decreases when more weight is assigned to individual stimulus effects. The stimulus frequencies of designs achieved by our GA agree with the approximated optimal stimulus frequencies derived by Liu and Frank (2004). In addition to stimulus frequency, our GA also helps to find the best sequence of stimuli, which is another important factor for design selection. A

future research of interest is to develop an improved algorithm by incorporating the optimal stimulus frequency.

In addition, we search for optimal designs when multiple scanning sessions are considered. Several algorithms are developed and compared. For this particular application, we can search for optimal design over a constraint design space; a design in this constraint space is formed by cyclically permuting stimulus types of a short design for the first scanning session to form other short designs for subsequent scanning sessions. Searching over this constraint space yields designs that are comparable to those obtained by searching over the entire space. We also observe that designs obtained by ignoring session effects can be inefficient when session effects are present. This is especially true when the response has a second-order polynomial drift and the noise is correlated. Taking into account session effects at the design stage is thus important.

Moreover, we consider hybrid algorithms which combine concepts of simulated annealing and GAs. We compare the performance of the proposed algorithms in finding optimal designs when the warm-up period of an MR scanner is taken into account. Through simulations, we demonstrate that our GA outperforms the proposed hybrid algorithms. Again, our GA is shown to be reliable in finding optimal ER-fMRI designs. This particular case study also indicates that, with a warm-up period of 16 seconds, designs with stimuli being presented in the warm-up period are only slightly better than designs with no stimuli in that period.

We also investigate ER-fMRI designs for nonlinear models. This is crucial since a nonlinear model allows us to consider both estimation and detection in a unified setting; it provides a more natural way to take into account the uncertainty in the shape of the HRF when focusing on detection problems. We find locally optimal designs for experiments involving one stimulus type. We observe that a design having a high design efficiency in detection under an HRF can be inefficient under a different HRF. To have designs that work relatively well over a variety of HRFs, we also consider finding pseudo-Bayesian designs under a given prior distribution on the parameters controlling the shape of the HRF. In addition, we

adopt techniques from multi-objective optimization problems to generate a set of designs for researchers to choose from based on their goals and needs.

For the study on designs for nonlinear models, we consider an HRF model that is already built in the popular software, SPM2 (http://www.fil.ion.ucl.ac.uk/spm). We follow Wager et al. (2005) to allow two free parameters at the design stage. However, there are other parameters as well. Increasing number of parameters is a topic of future research. Other HRF models are also proposed in the literature (e.g. Woolrich et al., 2004; Genovese, 2000). Studying designs with these HRF models is also a future research of interest.

Here, we focus on two common statistical goals, psychological constraints and customized requirements. We define individual design criteria for evaluating the quality of designs with respect to these objectives, and define multi-objective criteria as convex combinations of normalized individual criteria. Designs optimizing the multi-objective criterion are then searched for using search algorithms. As indicated in Chapter 2, we can also consider other objectives as long as corresponding design criteria can be defined. After a proper normalization, these design criteria can be incorporated to form MO criteria. Corresponding multi-objective optimal designs can then be searched for by search algorithms.

8.1 REFERENCES

Genovese, C. R. (2000), "A Bayesian Time-Course Model for Functional Magnetic Resonance Imaging Data," *Journal of the American Statistical Association*, 95, 691–703.

Henson, R. N. A., Price, C. J., Rugg, M. D., Turner, R., and Friston, K. J. (2002), "Detecting Latency Differences in Event-Related BOLD Responses: Application to Words Versus Nonwords and Initial Versus Repeated Face Presentations," *Neuroimage*, 15, 83–97.

Liu, T. T. and Frank, L. R. (2004), "Efficiency, Power, and Entropy in Event-Related fMRI with Multiple Trial Types: Part I: Theory," *NeuroImage*, 21, 387–400.

Wager, T. D. and Nichols, T. E. (2003), "Optimization of Experimental Design in fMRI: A General Framework Using a Genetic Algorithm," *NeuroImage*, 18, 293–309.

Wager, T. D., Vazquez, A., Hernandez, L., and Noll, D. C. (2005), "Accounting for Nonlinear BOLD Effects in fMRI: Parameter Estimates and a Model for Prediction in Rapid Event-Related Studies," *NeuroImage*, 25, 206–218.

Woolrich, M. W., Behrens, T. E. J., and Smith, S. M. (2004), "Constrained Linear Basis Sets for HRF Modelling Using Variational Bayes," *NeuroImage*, 21, 1748–1761.