

Constrained Multi-objective Designs for Functional MRI Experiments via A Modified NSGA-II

Ming-Hung Kao

Arizona State University, Tempe, USA.

Abhyuday Mandal

University of Georgia, Athens, USA.

John Stufken

University of Georgia, Athens, USA.

Summary. Functional magnetic resonance imaging (fMRI) is an advanced technology for studying brain functions. Due to the complexity and high cost of fMRI experiments, high quality multi-objective (MO) fMRI designs are in great demand; they help to render precise statistical inference, and are keys to the success of fMRI experiments. Here, we propose an efficient approach for obtaining MO fMRI designs. In contrast to existing methods, the proposed approach does not require users to specify weights for the different objectives, and can easily handle constraints to fulfil customized requirements. Moreover, the underlying statistical models that we consider are more general. We can thus obtain designs for cases where brief, long or varying stimulus durations are utilized. The usefulness of our approach is illustrated using various experimental settings.

Keywords: Genetic algorithms, Hemodynamic response function, Multi-objective optimization, Design efficiency

1. Introduction

Functional magnetic resonance imaging (fMRI) is one of the most dominant brain mapping techniques. This pioneering technology has many important clinical potentials such as early identification of Alzheimer's Disease (Wierenga and Bondi, 2007) and pre-neurosurgical planning (Bookheimer, 2007); see also Brown (2007). In an fMRI experiment, an MRI scanner is used to noninvasively measure changes in the blood oxygenated level dependent (BOLD) signal due to mental tasks or stimuli (Ogawa et al., 1990). The measured fluctuations of the BOLD signal, or the BOLD time series, are analyzed to make statistical inference about brain functions. Two common statistical goals include 1) detection of active brain voxels (three-dimensional imaging unit), and 2) the estimation of the hemodynamic response function (HRF), a function of time describing the change of the BOLD signals evoked by one single stimulus. Detection allows to identify brain regions activated by the stimuli. Estimation helps to understand the effect of a stimulus on the brain. Here, we aim at obtaining good designs that help to achieve both goals while taking into account practical constraints. Other relevant issues and overviews of fMRI can be found in Lazar (2008) and Lindquist (2008).

Address for correspondence: Ming-Hung Kao, School of Mathematical and Statistical Sciences, Arizona State University, P.O. Box 871804, Tempe, AZ 85287
E-mail: mhkao@math.asu.edu

Well planned designs are crucial to the success of fMRI experiments. However, obtaining high quality fMRI designs is not an easy task. It requires careful consideration regarding study objectives, statistical models, psychological constraints, and MRI machine settings. Taking into account these practical issues, existing approaches utilize discrete optimization techniques to search over the enormous design space for good designs. Wager and Nichols (2003) propose a genetic algorithm (GA) framework that targets multi-objective (MO) designs achieving high statistical efficiencies in detection and estimation and avoiding possible psychological confounds such as anticipation and habituation. Their GA has been applied in many studies over the last few years (e.g., Summerfield et al., 2006; Wang et al., 2007; Rameson et al., 2010). Following this work, Kao et al. (2009a) develop a more efficient approach. They make use of current knowledge about fMRI designs to largely improve the efficiency of the GA search. Their approach is demonstrated to outperform previous methodologies.

Previous studies deal with the MO nature of fMRI experiments by using an objective function that is a weighted sum of criteria that target each objective separately. When weights are available, the GA of Kao et al. (2009a) is most efficient, and thus recommended. However, assigning weights can be arduous in practice (Deb, 2001; Ding et al., 2004). One reason is that the mapping between the assigned weights and the performance of the resulting design under the various criteria is usually unclear. The design may fail to meet the experimenter’s expectation (Marler and Arora, 2010, 2004). For example, assigning equal weights is common when seeking a design with equal efficiencies for all objectives. However, the result can be far from this requirement; see, e.g. Cook and Wong (1994) and Section 4 of this paper.

In this article, we consider an alternative approach for obtaining MO designs when information about weights is vague. The idea is to obtain not only one, but a class of diverse, (near) Pareto-optimal designs; a design is Pareto-optimal if no other designs perform better in one or more objectives while being equivalent in all other objectives. The designs in this class approximate the Pareto frontier, and hence, offer optimal trade-offs between different objectives. Experimenters can then scrutinize the characteristics of the obtained designs and select one that suits their needs best. To efficiently achieve such a design class, we propose a search algorithm that combines the merits of the GA of Kao et al. (2009a) and the nondominated sorting GA II (NSGA-II; Deb et al., 2002). The NSGA-II is popular for MO optimization problems. By incorporating the GA of Kao et al. (2009a), the proposed algorithm is a natural fit for fMRI design problems and can efficiently produce good MO designs for a wide spectrum of real-life experimental settings.

Our proposed approach belongs to the class of ‘a posteriori’ methods, in which a set of solutions is obtained for users to select from based on their preferences (Miettinen, 1999). The weighted sum method can also be used to achieve this. In our context, this is done by repeatedly using the GA of Kao et al. (2009a) to obtain designs for systematically changed weights. We compare our new approach with this weighted sum method, and demonstrate that our approach uses much less CPU time and achieves good designs. More importantly, the designs that we obtain have greater diversity in the objective space. This diversity facilitates design selection, and is viewed as an important property to achieve (Zitzler et al., 2003, and references therein). Furthermore, our approach can handle constraints easily and generates designs satisfying experimenters’ requirements. By contrast, the weighted sum method is clumsy in this regard. We also compare our approach with the original NSGA-II. With a slightly increased CPU time, our approach achieves designs with better ability to detect brain activation.

Moreover, the underlying statistical models that we consider are more general than those used in previous studies. Our models accommodate cases considering brief stimuli (~ 1 s), stimuli lasting several seconds, or stimuli with durations varying across types. Such cases are not uncommon. For example, Murphy et al. (2006) use a 1-s flashing checkerboard to detect active voxels and estimate the HRF. Whereas Birn et al. (2004) and Brendel et al. (2010) consider stimuli of varying durations. Martin et al. (2005) employ 5-s pictures and detect activated brain regions on stroke patients with aphasia. The models that we consider also allow cases where event-related regressors are used to replace epoch regressors as advocated by Mechelli et al. (2003). With these important improvements, our approach provides a flexible and powerful tool for obtaining good fMRI designs. A MATLAB program implementing this approach is available upon request from the first author.

The remainder of the article is organized as follows. In Section 2, we briefly introduce background information about fMRI designs. We then present our proposed approach in Section 3, including the statistical models, design criteria and search algorithm. In Section 4, our proposed approach is compared with the weighted sum method and the NSGA-II under various experimental settings. The paper closes with a brief discussion in Section 5. While terminology and notation are explained when introduced, for ease of reference we provide a list of selected items in an Appendix at the end of the paper.

2. Background and Terminology

An fMRI design is a sequence of mental stimuli (e.g. pictures or sounds) of one or more types interlaced with a control (e.g., rest or fixation). Such a sequence is presented to an experimental subject while an MRI machine scans the subject's brain to collect BOLD time series from each brain voxel. When being presented, each stimulus may last several milliseconds to a few seconds. Times between consecutive stimulus onsets may vary across stimuli; they are assumed to be multiples of a pre-specified time which is termed as the inter-stimulus interval (*ISI*) in this paper. The control fills in the time when no stimulus is being presented. Such an fMRI design is typically written as a finite sequence of finite numbers; e.g., $s = \{101210 \cdots 1\}$. A number q ($\neq 0$) at the k th position of a design indicates an onset of a q th-type stimulus at time $(k - 1)ISI$. A '0' means no stimulus onset at that time point.

Well known fMRI designs include block designs, m -sequences, random designs, mixed designs, permuted block designs, and clustered m -sequences. A block design is a patterned sequence in which stimuli of the same type are clustered. For example, a block design with block size four may consist of repetitions of $\{111122220000\}$. An m -sequence can be generated from primitive polynomials for a Galois field (MacWilliams and Sloane, 1977; Godfrey, 1993; Buračas and Boynton, 2002). Although systematically generated, m -sequences look rather random with no perceivable patterns. Random designs are randomly generated and typically have no perceivable patterns. A mixed design is obtained from concatenating a fraction of a block design with a fraction of an m -sequence or a random design. Permuted block designs are generated by repeatedly exchanging elements of a block design to make the resulting design more and more random. On the other hand, a clustered m -sequence is generated by clustering (through permutation) stimuli of the same type in an m -sequence. See Liu (2004) for detailed descriptions of these designs.

Different designs are recommended for different purposes. Friston et al. (1999) demonstrate a high performance of the block design in detecting activations. For estimating the

HRF, Dale (1999) advocates the use of random designs, and Buračas and Boynton (2002) recommend m -sequences due to their low autocorrelation property. Buxton et al. (2000), Liu et al. (2001), Liu and Frank (2004) and Liu (2004) further investigate efficiencies of designs with respect to these two dimensions. They observe that 1) block designs possess high efficiencies in detection, but have low estimation efficiencies, 2) random designs and m -sequences are efficient for estimating the HRF, but they are not as good as block designs in detection, and 3) designs containing both ‘blocky’ and ‘random’ components can reach good compromises between the two objectives; such designs can be selected from mixed designs, permuted block designs and clustered m -sequences. These authors conclude that there is a trade-off between efficiency for detection and efficiency for estimation. In addition, Liu and colleagues emphasize the importance of perceived randomness (or unpredictability) of designs. Patterned designs such as block designs are easy to predict by the experimental subject. They may give rise to psychological confounds such as anticipation and expectation; see also Dale (1999). Designs that are not patterned tend to avoid such confounds and are thus favorable.

While study objectives are important considerations at the design stage, other experimental conditions such as time to repetition (TR , time between two MRI scans of the same voxel), ISI , and duration of the experiment are also key factors. These conditions can vary across experiments. To achieve tailor-made designs for each unique experiment, Wager and Nichols (2003) propose a GA framework to search for good designs with respect to user-specific experimental settings. Following this GA framework, Kao et al. (2009a) take advantage of current knowledge about fMRI designs to develop a more efficient approach. Their approach has also been adapted to cases involving multiple scanning sessions (Kao et al., 2009b), and studies where both stimulus effects and pairwise comparisons are of interest (Kao et al., 2008). Maus et al. (2010a) also apply this approach for investigating the robustness of designs against mis-specified temporal autocorrelation of fMRI noise. They also use it to study the efficiency of D-optimal designs under the A -optimality criterion, and vice versa.

One difficulty of the previously mentioned GAs is that they require user-specific weights on study objectives for an MO approach. The interpretation of these weights is almost always unclear, so that there is almost never a meaningful choice. To tackle this issue, we propose an alternative approach. Our approach is described in the next section including the underlying statistical models, design criteria and a search algorithm; current knowledge about the mechanism behind the observed BOLD response, which underpins the statistical models, is also briefly presented.

3. Methodology

3.1. Dual General Linear Models

In an fMRI experiment, stimuli are presented to a subject in an order determined by a selected design. Each stimulus evokes neuronal firings at an activated brain voxel. Due to increased metabolic demands, oxygenated blood flows into the cerebral blood vessels around the active voxel, and changes the ratio of oxy- to deoxy-hemoglobin. This change affects the local magnetic field, and leads to a fluctuation in the signal intensity collected by an MRI scanner; see also Cabeza and Kingstone (2006). It takes a few seconds for the MRI signal intensity to rise and decay. The process is sluggish and is typically described by an HRF. When stimuli occur in close succession, the evoked HRFs overlap and accumulate to

form the BOLD time series, which also involves nuisance signals and noise.

Linear models are the most popular for analyzing the BOLD time series (Friston et al., 1995; Worsley and Friston, 1995; Lindquist, 2008). These models commonly assume a linear time invariant system, in which each stimulus of the same type evokes the same HRF throughout the experiment, and overlapping HRFs accumulate additively. For detection problems, the HRF is typically assumed to be the product of an unknown amplitude parameter, θ_q , and an assumed HRF shape, \mathbf{h}_q^* . Here, $q = 1, \dots, Q$, and Q is the total number of stimulus types. With these assumptions, the linear model that we consider for detection has the following form:

$$\mathbf{y} = \sum_{q=1}^Q \mathbf{X}_q \mathbf{h}_q^* \theta_q + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \quad (1)$$

Here, \mathbf{y} is a T -by-1 vector representing the BOLD time series of a voxel acquired every TR seconds. $\mathbf{X}_q \mathbf{h}_q^* \theta_q$ represents the contribution to the accumulated HRFs evoked by the q th-type stimulus. \mathbf{X}_q is a zero-one design matrix corresponding to occurrences of the q th-type stimulus (see Appendix I for further details). $\mathbf{S}\boldsymbol{\gamma}$ is a nuisance term representing the drift/trend of \mathbf{y} with a parameter vector $\boldsymbol{\gamma}$, and $\boldsymbol{\varepsilon}$ is correlated noise. The main focus of detection problems is on the amplitudes θ_q or, more generally, their linear combinations, $\mathbf{C}_\theta \boldsymbol{\theta}$, where \mathbf{C}_θ is a coefficient matrix for linear combinations of interest, and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_Q)'$. For example, if $\mathbf{C}_\theta = (1, -1, 0, \dots, 0)$, the focus is on the difference between the HRF amplitudes evoked by the first two stimulus types.

Model (1) is general in that it accommodates many practical situations. For example, when the study involves only brief stimuli, we could choose \mathbf{h}_q^* from the double-gamma function of SPM, a popular software package for fMRI (<http://www.fil.ion.ucl.ac.uk/spm/>). Specifically, we would take the j th element of \mathbf{h}_q^* as $h_{q,j}^* = g((j-1)\Delta T) / \max_t g(t)$, where

$$g(t) = \begin{cases} \frac{t^5 e^{-t}}{5!} - \frac{1}{6} \frac{t^{15} e^{-t}}{15!}, & t \in [0, 32]; \\ 0, & \text{otherwise} \end{cases}$$

is a double-gamma function, ΔT is the greatest real value making both $(ISI/\Delta T)$ and $(TR/\Delta T)$ integers, $j = 1, 2, \dots, 1 + \lfloor 32/\Delta T \rfloor$, and $\lfloor a \rfloor$ is the integer part of a ; see Figure 1(A).

When the stimulus lasts several seconds, \mathbf{h}_q^* may be chosen as a convolution of a boxcar function and the function $g(t)$. In such a case, $h_{q,j}^* = h((j-1)\Delta T) / \max_t h(t)$, where $h(t) = \int_0^t b(t-\tau)g(\tau)d\tau$ is the convolution of $b(t)$ and $g(t)$,

$$b(t) = \begin{cases} 1, & t \in [0, \tau_{dur}]; \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

is a boxcar function for a stimulus with duration τ_{dur} seconds, and $j = 1, \dots, 1 + \lfloor (32 + \tau_{dur})/\Delta T \rfloor$. Boxcar functions of 3-, 6- and 12-s are presented in Figure 1(B); their convolutions with $g(t)$, normalized to have a maximum of 1, are presented in Figure 1(C). We note that τ_{dur} should ideally be determined by the duration of the underlying neuronal activity evoked by the stimulus. However, this duration may not be known; see also Loh et al. (2008). Here, we follow a common practice to set τ_{dur} to the stimulus duration, and assume that the neuronal activity and stimulus durations are the same. In Section 4, we investigate the robustness of designs when these two durations do not match; i.e., τ_{dur} is mis-specified.

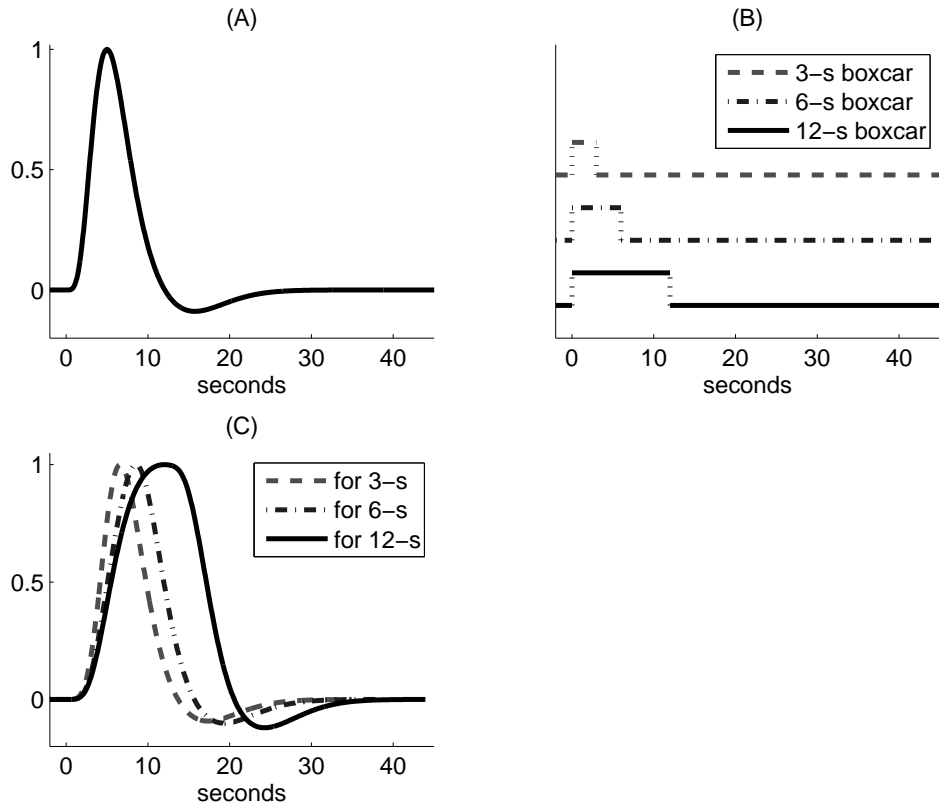


Fig. 1. (A) The function $g(t)$, normalized to have a maximum of 1. (B) The boxcar functions of 3-, 6- and 12-s. (C) The convolutions, normalized to have a maximum of 1, of the function $g(t)$ and the boxcar functions.

Model (1) can also be considered when one chooses to follow Mechelli et al. (2003) to consider event-related regressors in lieu of epoch-related regressors. This can be done by replacing the previously described boxcar function $b(t)$ by a spike train $\mathbb{1}_\Gamma(t)$ with, say, $\Gamma = \{0, 3, 6, \dots, 3\lfloor \tau_{dur}/3 \rfloor\}$; here, $\mathbb{1}_\Gamma(t)$ is 1 for $t \in \Gamma$, and 0 otherwise. We also note that \mathbf{h}_q^* can be different across stimulus types. This allows varying stimulus durations across types. Although the double-gamma function is popular, other HRF shapes can also be considered.

For estimating the HRF, the following linear model is considered (see also, Dale, 1999):

$$\mathbf{y} = \sum_{q=1}^Q \mathbf{X}_q \mathbf{h}_q + \mathbf{S}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}. \quad (3)$$

Here, \mathbf{h}_q is a vector of parameters describing the HRF evoked by a q th-type stimulus (Dale, 1999); the j th element of \mathbf{h}_q represents the height of the HRF at time $(j-1)\Delta T$ following a stimulus onset. All other terms are defined under model (1). The focus here is on estimating $\mathbf{C}_h \mathbf{h}$; \mathbf{C}_h is a linear combination matrix, and $\mathbf{h} = (\mathbf{h}'_1, \dots, \mathbf{h}'_Q)'$. When $\mathbf{C}_h = [\mathbf{I}_k, -\mathbf{I}_k, \mathbf{0}, \dots, \mathbf{0}]$, the interest lies in the difference between the HRFs incurred by the

first two stimulus types; \mathbf{I}_k is the k -by- k identity matrix, and k is the length of the vector \mathbf{h}_q , which is set to be the same as that of \mathbf{h}_q^* in Model (1).

3.2. Optimal Design Criteria

Following Kiefer (1959), the performance of a design for detecting activation is evaluated by some function of the covariance matrix $Cov(\mathbf{C}_\theta \hat{\boldsymbol{\theta}})$ of the generalized least squares estimator (GLSE), $\mathbf{C}_\theta \hat{\boldsymbol{\theta}}$. Similarly, the estimation efficiency is measured by some function of $Cov(\mathbf{C}_h \hat{\mathbf{h}})$, where $\mathbf{C}_h \hat{\mathbf{h}}$ is the GLSE of $\mathbf{C}_h \mathbf{h}$. Specifically, with models (1) and (3), these covariance matrices are

$$\begin{aligned} Cov(\mathbf{C}_\theta \hat{\boldsymbol{\theta}}) &= \mathbf{C}_\theta \{ \mathbf{H}' \mathbf{X}' \mathbf{V}' (\mathbf{I}_T - \mathbf{P}_{\mathbf{V}\mathbf{S}}) \mathbf{V} \mathbf{X} \mathbf{H} \}^{-1} \mathbf{C}_\theta'; \\ Cov(\mathbf{C}_h \hat{\mathbf{h}}) &= \mathbf{C}_h \{ \mathbf{X}' \mathbf{V}' (\mathbf{I}_T - \mathbf{P}_{\mathbf{V}\mathbf{S}}) \mathbf{V} \mathbf{X} \}^{-1} \mathbf{C}_h'. \end{aligned}$$

Here, \mathbf{H} is a diagonal matrix with the vectors $\mathbf{h}_1, \dots, \mathbf{h}_Q$ along the diagonal, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_Q)$, and \mathbf{V} is an assumed whitening matrix so that $\mathbf{V}\boldsymbol{\varepsilon}$ is white noise. $\mathbf{P}_A = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ is the orthogonal projection matrix onto the column space of \mathbf{A} , and \mathbf{A}^- is a generalized inverse of \mathbf{A} .

We consider the A -optimality criterion in our case studies. Designs with smaller average variances for the estimators of the linear combinations of the parameters of interest are said to be (A -)better (Bailey, 2007). For a design \mathbf{s} , the A -efficiency for detection is defined as:

$$F_d^*(\mathbf{s}) = \frac{F_d(\mathbf{s})}{F_d(\mathbf{s}_d^*)}, \text{ where } F_d(\mathbf{s}) = [\text{trace}\{Cov(\mathbf{C}_\theta \hat{\boldsymbol{\theta}})\}]^{-1},$$

and \mathbf{s}_d^* is a max- F_d design maximizing F_d . We use the GA of Kao et al. (2009a) to maximize F_d , and to obtain an approximate max- F_d design. For estimation, the A -efficiency of \mathbf{s} is:

$$F_e^*(\mathbf{s}) = \frac{F_e(\mathbf{s})}{F_e(\mathbf{s}_e^*)}, \text{ where } F_e(\mathbf{s}) = [\text{trace}\{Cov(\mathbf{C}_h \hat{\mathbf{h}})\}]^{-1}.$$

The design \mathbf{s}_e^* is a max- F_e design maximizing F_e , and can be approximated via the GA of Kao et al. (2009a). While we present results for A -optimality, our approach can also accommodate other optimality criteria, such as D -optimality. Note that the design criteria presented here are for a subset of the model parameters. Such criteria are sometimes termed as A_s - (or D_s -)optimality; see also Atkinson et al. (2007). For clarity, we omit the subscript and use the term A - (or D -)optimality in this article.

In addition to statistical efficiencies, experimenters may want to use a desired frequency for each stimulus type due to practical reasons (see Section 4.3). We take this requirement into account by considering the F_f^* -criterion of Kao et al. (2009a):

$$F_f^*(\mathbf{s}) = 1 - \frac{F_f(\mathbf{s})}{F_f(\mathbf{s}^0)}, \text{ where } F_f(\mathbf{s}) = \sum_{i=1}^Q [|n_i - nP_i|], \quad (4)$$

n_i is the number of the type- i stimulus in the sub-design of \mathbf{s} excluding all zeros, n is the length of the sub-design, P_i is the desired proportion of the i th stimulus type, \mathbf{s}^0 is the design containing only the stimulus type i with the smallest P_i , and $|a|$ is the greatest integer less than or equal to the absolute value of a . To fulfil the experimenter's requirement, we regard " $F_f^* \geq c_f$ " as a constraint for a given c_f while optimizing (F_d^*, F_e^*) . In the next subsection, we present an efficient algorithm to achieve this goal.

3.3. Search Algorithm

Our proposed algorithm combines the NSGA-II of Deb et al. (2002) and the GA of Kao et al. (2009a). It mimics Darwin’s theory of evolution to move through generations. Good parents are selected to reproduce offsprings, and with the survival-of-fittest principle, individuals of better fit survive to the next generation. The process, when repeated, ensures preservation of good traits and high-quality designs can be expected.

To determine the fitness of designs with respect to multiple objectives, we follow the NSGA-II to consider two measures, namely nondomination rank and crowding distance. These measures are functions of design efficiencies (i.e., F_i^* -values). Nondomination ranks are assigned to designs within each GA generation. The first rank is assigned to designs that are not dominated by any other designs in the same generation; a design \mathbf{s}_1 is said to be dominated by another design \mathbf{s}_2 if $F_i^*(\mathbf{s}_1) \leq F_i^*(\mathbf{s}_2)$ for both $i = d, e$ and $F_i^*(\mathbf{s}_1) < F_i^*(\mathbf{s}_2)$ for $i = d$ or e . Designs of the second nondomination rank are dominated by one or more first-ranked designs, but not by others. The subsequent ranks are assigned accordingly. When comparing designs, this measure is of the primary concern; it helps to move toward Pareto-optimality. An efficient way for rank assignments can be found in Deb et al. (2002), and is not repeated here.

When the constraint $F_f^* \geq c_f$ is used, the “constrained-domination” of Deb et al. (2002) is utilized for assigning nondomination ranks. A design \mathbf{s}_1 is said to be constrained-dominated by \mathbf{s}_2 if 1) both designs satisfy the constraint, and \mathbf{s}_1 is dominated by \mathbf{s}_2 , or 2) at least one \mathbf{s}_i fails to satisfy the constraint and $F_f^*(\mathbf{s}_1) < F_f^*(\mathbf{s}_2)$. We note that the constrained case is equivalent to the unconstrained one if $c_f = 0$.

As the secondary measure, the crowding distance is used to compare designs of the same nondomination rank. It helps to maintain diversity in the objective space in terms of design efficiencies (F_d^* and F_e^*). Designs with small crowding distances are close to their neighbors, and we would like to leave them out. To obtain the crowding distance, we first use F_d^* to sort the designs. If a design \mathbf{s} has the smallest or the largest F_d^* , we follow Deb et al. (2002) to set $d_d(\mathbf{s}) = \infty$ to avoid leaving out the designs on the boundaries. Otherwise, $d_d(\mathbf{s})$ is the absolute difference between the F_d^* -values of the first neighbors of \mathbf{s} . Repeat the procedure again by using F_e^* to obtain $d_e(\mathbf{s})$. The crowding distance of \mathbf{s} is $d_d(\mathbf{s}) + d_e(\mathbf{s})$; see also Deb et al. (2002).

With these two measures, we describe the proposed algorithm below.

1. *Initials (first generation)*. Obtain $2G$ initial designs to form the first generation, where G is an even number. The designs that we use include block designs, m -sequences, random designs, the max- F_d design and the max- F_e designs and combinations of the latter two. Calculate design efficiencies for these designs.
2. *Fitness & mating pool*. Compute the nondomination ranks and crowding distances for the designs in the current generation. Based on these two measures, select the G best designs from the current generation to form a mating pool.
3. *Stopping rule*. If a given number of generations is reached, terminate the search and report the G designs in the mating pool. Otherwise, continue to the next step.
4. *Selecting parents*. Obtain $G/2$ pairs of parents from the mating pool via a tournament selection. Specifically, four distinct designs are randomly selected from the mating pool to form two pairs of designs. Choose the better one from each of the two design pairs; the two winners continue as a pair of parents. This process is then repeated

$G/2$ times to form $G/2$ pairs of parents, each time starting with the selection of four designs from the original mating pool of G designs.

5. *Reproduction.* Use the paired parents to generate offsprings via crossover and mutation; each pair gives birth to two offspring designs. The crossover operator exchanges the corresponding fractions of each pair of designs based on a randomly selected cut-point. The mutation operator then perturbs a randomly selected portion α_m of elements of the G resulting designs. The selected elements are replaced by new ones generated from the discrete uniform distribution $\mathcal{U}\{0, 1, \dots, Q\}$.
6. *Next generation.* Obtain the design efficiencies of the G offspring designs. The G offsprings along with the G designs in the mating pool form the next generation. Go back to Step 2 and repeat the process.

We follow Kao et al. (2009a) to include well known fMRI designs in the initial generation. These designs provide ‘building blocks’ to facilitate the search. They can be generated systematically, but are not easy to achieve via random mechanisms (crossover and mutation). We also include max- F_d and max- F_e designs which are required for calculating F_d^* and F_e^* , respectively. Including them does not significantly increase the computational resource and is helpful. When implementing the algorithm, we set G to 100 and α_m to 1%. The search is terminated after 2,500 generations. While these algorithmic parameters may not be optimal, the achieved designs are satisfactory as presented in Section 4.

We note that the NSGA-II has a different procedure in the first iteration. In that algorithm, G designs are first randomly generated to produce another G designs via tournament selection, crossover and mutation. The $2G$ designs are combined to form ‘initial designs’ in Step 1. Other steps are then implemented to generate subsequent generations; see also Deb et al. (2002). When implementing this algorithm, we also set $G = 100$, and $\alpha_m = 1\%$, and terminate the search after 2,500 generations.

4. Case Studies

In this section, we use our approach to generate MO fMRI designs for various experimental settings. The results are compared with those of the weighted sum method and NSGA-II in terms of the diversity of the obtained designs in the objective space, achieved trade-offs, and CPU time spent. In addition, designs are selected from the obtained design classes using various selection criteria. These selected designs are compared to provide information about the ability of the three approaches in fulfilling the experimenter’s needs. We also use our approach to study the robustness of designs to mis-specified τ_{dur} in (2). Moreover, our approach is applied to obtain designs for the experiment of Brendel et al. (2010), who use a design obtained through random permutations. While such designs are not uncommon in practice, we demonstrate that our designs can significantly outperform them.

4.1. MO designs for brief stimuli

We first consider an experimental setting that is also studied by Kao et al. (2009a). The number of stimulus types is three ($Q = 3$) and all stimuli have a short duration. The length of the design sequence is 255. The *ISI* is 2s and so is the *TR*. Two study objectives of interest include detecting activation and estimating the HRF, and the linear combination matrices C_θ and C_h are identity matrices. The response is assumed to have a second-order

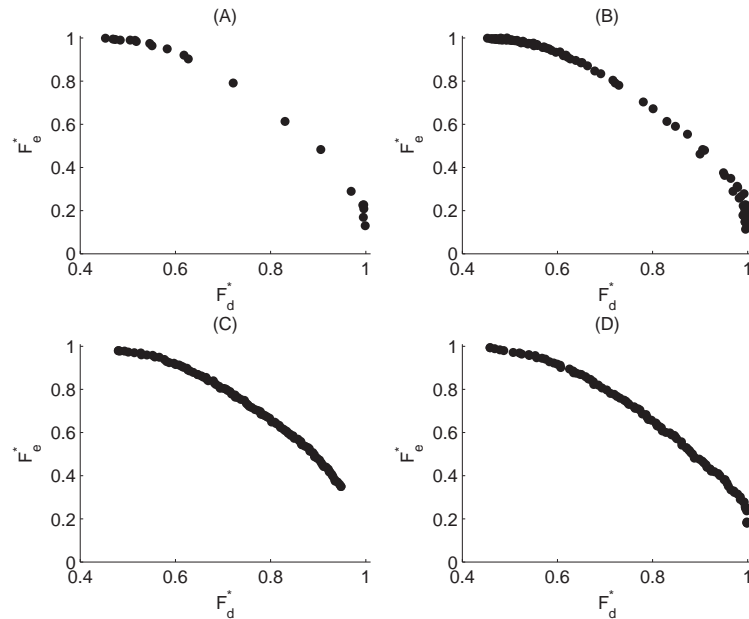


Fig. 2. The F_e^* - against F_d^* -values of the designs obtained from (A) WS-0.05, (B) WS-0.01, (C) the NSGA-II, and (D) our approach. The number of designs are: (A) 21; (B) 101; (C&D) 100.

polynomial drift. The noise follows an AR(1) process with a correlation coefficient $\rho = 0.3$. For $q = 1, 2, 3$, the HRF shape \mathbf{h}_q^* of model (1) is the normalized $g(t)$ discussed in Subsection 3.1.

Figure 2(A) presents the F_e^* - against F_d^* -values of the 21 designs obtained by the weighted sum method under this scenario. These designs are achieved by repeatedly implementing the MATLAB program of Kao (2009) to maximize $F = wF_d^* + (1 - w)F_e^*$ with w increased from 0 to 1 in steps of 0.05. Except for the stopping rule, the algorithmic parameters of the MATLAB program, including the population size (= 20), mutation rate (= 1%), and number of immigrants (= 4), are set to their default values (see Table 1 of Kao, 2009). The stopping rule that we choose is also built in the program and it terminates the search if there is no significant improvement in the F -value. In our experience, this stopping rule can save CPU time without sacrificing much in the achieved efficiency. Note that the mesh size of 0.05 is also considered in Kao et al. (2009a). For brevity, the weighted sum method with this mesh size is referred to as WS-0.05 henceforward.

A disadvantage of the WS-0.05 is observed in Figure 2(A) — the designs are clustered at the two ends of the approximate Pareto-frontier and are sparse in the middle. Users do not have many choices if designs with intermediate efficiencies are desired. The use of a smaller mesh size of 0.01 (referred to as WS-0.01) slightly improves the situation. However, clusters and gaps are still observed (see Figure 2B). Our approach and the NSGA-II yield designs with better diversity in the objective space. As shown in Figures 2(C) and (D), no large distances between consecutive designs are observed, and there are designs almost everywhere on the approximate Pareto-frontier. Experimenters can select a suitable design based on their preferences for a trade-off between the different objectives.

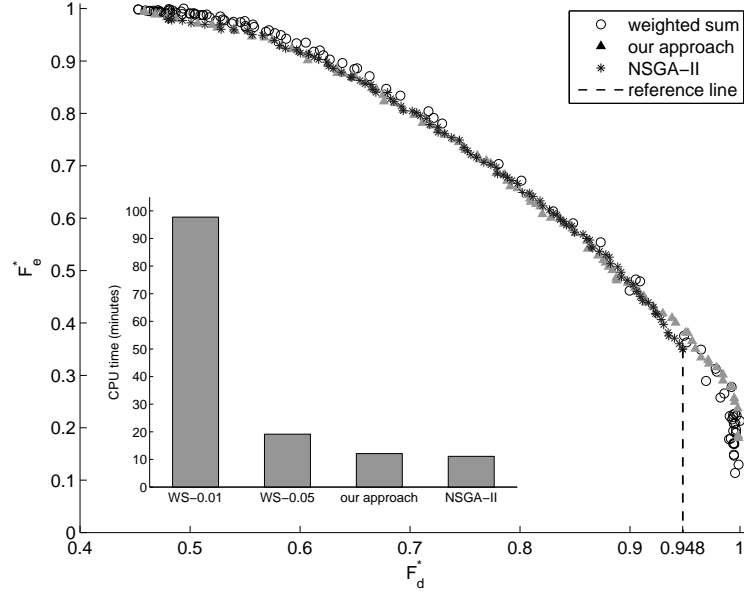


Fig. 3. The F_e^* - against F_d^* -values of the designs of different approaches and CPU times spent for obtaining these designs

Figures 2(C) and (D) also reveal that the maximal F_d^* -value achieved by the NSGA-II is slightly smaller than that attained by our approach. This is better manifested in Figure 3 where Figures 2(B)-(D) are overlaid; note that the design class in Figure 2(A) is a subset of that in Figure 2(B). As shown in Figure 3, the NSGA-II does not provide any design with $F_d^* > 0.948$. With the help of good initial designs, our approach is free from this drawback. Figure 3 also indicates that the weighted sum method can achieve designs with slightly better trade-offs; these designs are located slightly to the upper right of some designs of the other two approaches. While the weighted sum approach is slightly advantageous in this regard, it requires much more CPU time as also shown in Figure 3.

The CPU times presented in Figure 3 are obtained from implementing the different methods on a desktop computer with a 3.0 GHz Intel Pentium 4 quad-core processor. The weighted sum method is the most time consuming. With a mesh size of 0.05, it takes about 19 minutes to obtain 21 designs. About 98 minutes are required for the weighted sum method to achieve 101 designs. The NSGA-II spends about 11 minutes and our approach uses about 12 minutes to obtain 100 designs. The latter two approaches are more efficient.

In addition to the above comparisons, it is also of interest to see if the three approaches can fulfil the experimenters needs in design selection. One possible selection criterion is to find a design with balanced efficiencies ($F_d^* = F_e^*$). Such a design works equally well, relative to the best designs, in both dimensions. With this selection criterion, a design with $(F_d^*, F_e^*) = (0.744, 0.745)$ can be selected from the 100 designs of our approach. The NSGA-II achieves a design with $(F_d^*, F_e^*) = (0.745, 0.749)$. The weighted sum method yields $(F_d^*, F_e^*) = (0.729, 0.781)$, which is achieved by the WS-0.01 with the objective function $F = 0.62F_d^* + 0.38F_e^*$. Our approach and the NSGA-II produce designs with closer F_d^* - and F_e^* -values and better satisfy the selection criterion. We also note that, when equal weights

are assigned ($w = 0.5$), the weighted sum method does not yield a design with equal efficiencies for detection and estimation, but results in an (F_d^*, F_e^*) -value of $(0.618, 0.92)$.

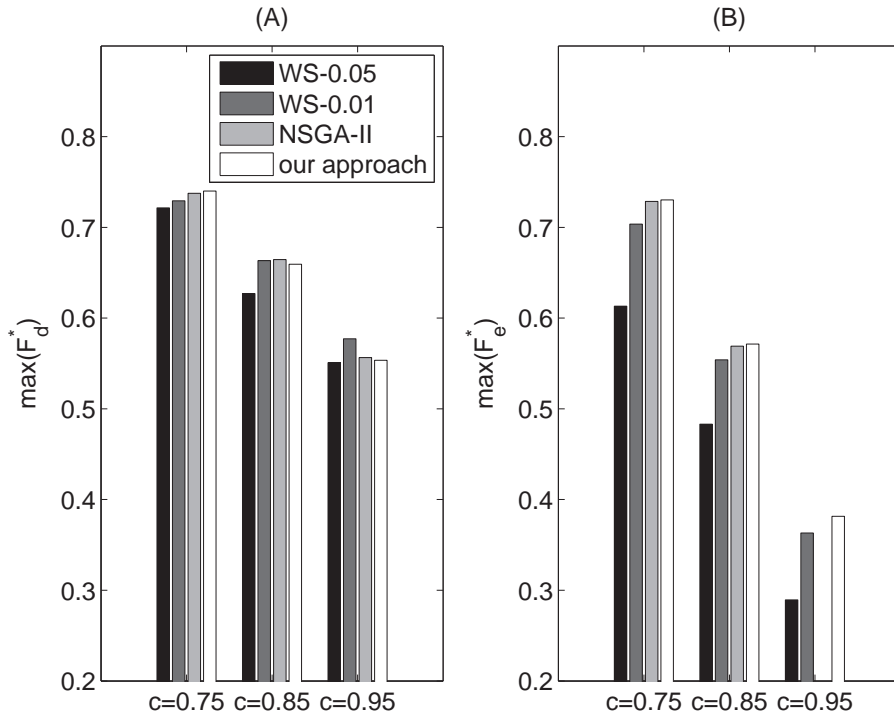


Fig. 4. (A) The maximum F_d^* achieved when $F_e^* \geq c$; (B) the maximum F_e^* achieved when $F_d^* \geq c$, $c = 0.75, 0.85$, and 0.95 .

Selecting constrained designs that maximize F_d^* (or F_e^*) subject to $F_e^* \geq c$ (or $F_d^* \geq c$) for a given c is also not uncommon. Figure 4 presents the $\max(F_i^*)$ achieved by different approaches when $F_j^* \geq 0.75, 0.85$ or 0.95 . The WS-0.05 consistently performs the worst, and the use of a finer mesh size improves the result. Although our approach and the NSGA-II use much less CPU time, Figure 4 demonstrates that these two methods achieve designs that are better or not much worse than those of the WS-0.01. While the NSGA-II performs similar to our approach in most cases, it fails to produce designs with $F_d^* \geq 0.95$ as indicated in the right most group of Figure 4(B). Our approach is therefore recommended.

4.2. MO designs for longer stimulus durations

For the second setting, stimuli that last several seconds are considered. We start with the case of two types ($Q = 2$) of 4-s stimuli. The length of the design is $L = 242$. The *ISI* is 4s and *TR* 2s. The HRF shape is the convolution of a 4-s boxcar function with the normalized $g(t)$; see Subsection 3.1. Other conditions are the same as for the experimental setting in Subsection 4.1.

Comparisons among our approach, the NSGA-II and weighted sum method are again conducted. The results convey similar information as in the previous subsection and are

thus omitted. Here, we focus on another important issue, namely the robustness of designs to mis-specification of the duration τ_{dur} of the boxcar function $b(t)$ in (2). This mis-specification may occur when there is a difference between the stimulus duration, which is commonly used to specify τ_{dur} , and the duration of the evoked neuronal activity, which is almost always uncertain and may vary across brain voxels. For example, in an experiment involving tasks requiring memory or decision-making, the neuronal activity might last longer than the stimulus presentation. In other instances, the neuronal activity evoked by a simple task might not last as long as the presentation duration of the task. Our results suggest that our obtained designs remain efficient if the difference between the mis-specified neural activity duration and the real one is no more than 2s, especially when the $ISI \leq 6s$. An efficiency loss is observed when the mis-specified duration is more than 4s apart from the real one. The amount of loss in the design efficiency increases with the ISI .

Figure 5 presents the performance of two sets of designs, namely designs for brief stimuli and those for 2-s stimuli ($\tau_{dur} = 2$) when the $ISI = 4s$. In Figure 5(A), we assume that the neuronal activity is brief; the τ_{dur} of 2 used to generate the second set of designs is thus mis-specified, and the first set of designs are suited to this particular case. As shown in the Figure, the Pareto fronts approximated by the two design sets are nearly indistinguishable. This indicates that the designs obtained for 2-s stimuli are very efficient when the neuronal activity is brief. In Figure 5(B), we assume that the actual neuronal activity duration is 2s. Again, the two designs sets perform similarly in this latter case. The designs generated for brief stimuli are therefore very efficient when the actual τ_{dur} is 2s.

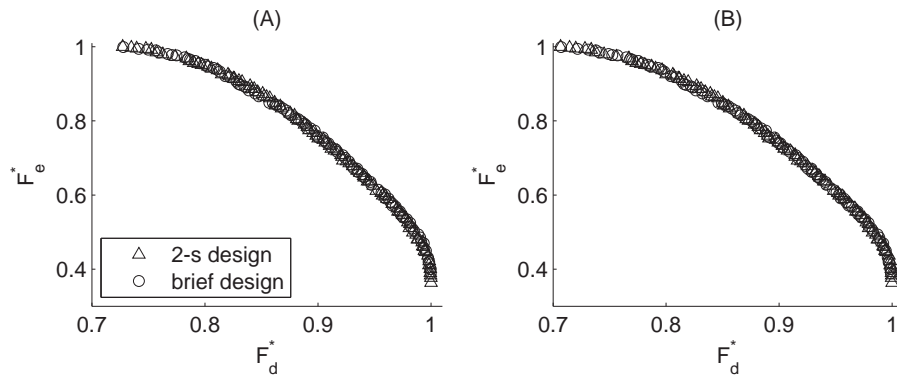


Fig. 5. F_e^* - against F_d^* -values of designs when the ISI is 4s and the actual neuronal activity duration is (A) brief, or (B) 2s.

To accommodate stimuli with duration longer than 2s, we increase the ISI to 6s and 8s. The F_e^* - against F_d^* -values of designs generated with 6-s ISI are presented in Figures 6(A-1) and (A-2), and those with 8-s ISI are in Figures 6(B-1) and (B-2). As shown in Figure 6(A-1), designs for 4-s stimuli yield similar F_d^* - and F_e^* -values as designs for brief stimuli when the actual neuronal activity duration is brief. On the other hand, Figure 6(A-2) indicates that designs for brief stimuli remain efficient when the actual neuronal activity duration is 4s. Similar results are observed when the ISI is 8s. The 4-s-stimulus designs perform quite well for brief neuronal activity; see Figure 6(B-1). We also observe that the designs for brief stimuli are efficient when the actual neuronal activity is 4s (not shown). Figure 6(B-1) also suggests that designs for 6-s stimuli can suffer an efficiency loss

if the neuronal activity evoked by the stimulus is brief. Figure 6(B-2) indicates that the brief-stimulus designs do not perform well when the underlying neuronal activity evoked by each stimulus lasts 6s. However, the 4-s stimulus designs achieve a good performance in this last scenario in Figure 6(B-2). Designs for 2-s stimuli are also studied, and are observed to perform similarly as brief-stimulus designs in all cases presented in Figure 6. For clarity, these 2-s-stimulus designs are omitted from Figure 6.

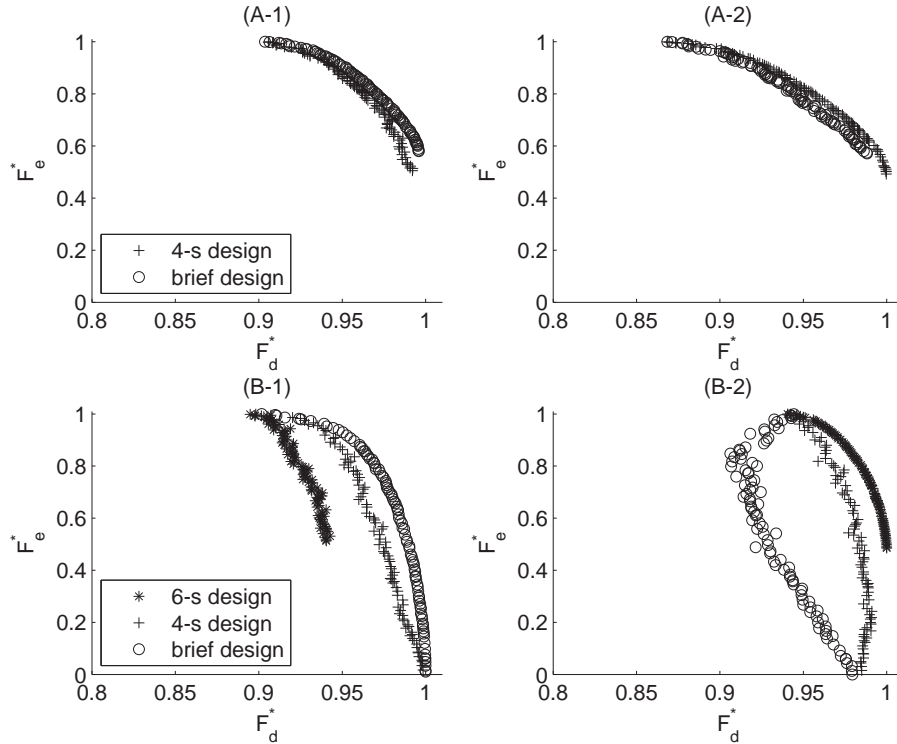


Fig. 6. F_e^* - against F_d^* -values of designs when (A-1) the actual neuronal activity duration is brief and the ISI is 6s, (A-2) the actual neuronal activity duration is 4s and the ISI 6s, (B-1) the actual neuronal activity duration is brief and the ISI is 8 s, and (B-2) the actual neuronal activity duration is 6s and the ISI 8s.

4.3. Constrained MO designs for varying stimulus durations

Following Brendel et al. (2010), we consider here three stimulus types with durations 6, 3.6 and 2 seconds, respectively. Specifically, the three stimulus types are 1) a 2-s preparatory auditory tone signal (S1) plus a 4-s, 2.5-Hz isochronous click train, 2) S1 plus a 1.6-s short click (2.5 Hz) train, and 3) S1 alone. They are respectively termed as the long click train (LCT), short click train (SCT) and no click train (NCT). The ISI is set to 9s and the design length to 67; it takes about 10 minutes to present the entire design. The MRI scanner scans each voxel every 1.5s ($TR = 1.5s$) to collect fMRI time series for detecting active brain voxels and estimating the HRFs evoked by the three types of stimulus.

The LCT and SCT involve clicks of sounds, namely strokes of a pen against a desk. During the experiment, the subject is asked to synchronize the syllable /ta/ to the clicks as close as possible in time. The experimenters decide to include more LCT trials, and expect that, with a higher number of LCT trials, the subject will be more engaged throughout the experiment. the number of occurrences of LCT is about three times as many as SCT and NCT. Therefore, in the notation of (4), $P_1 = \frac{3}{5}$ for LCT, $P_2 = \frac{1}{5}$ for SCT and $P_3 = \frac{1}{5}$ for NCT when calculating the F_f^* -value and we consider “ $F_f^* \geq 0.95$ ” as the constraint when seeking designs with good (F_d^*, F_e^*) -values.

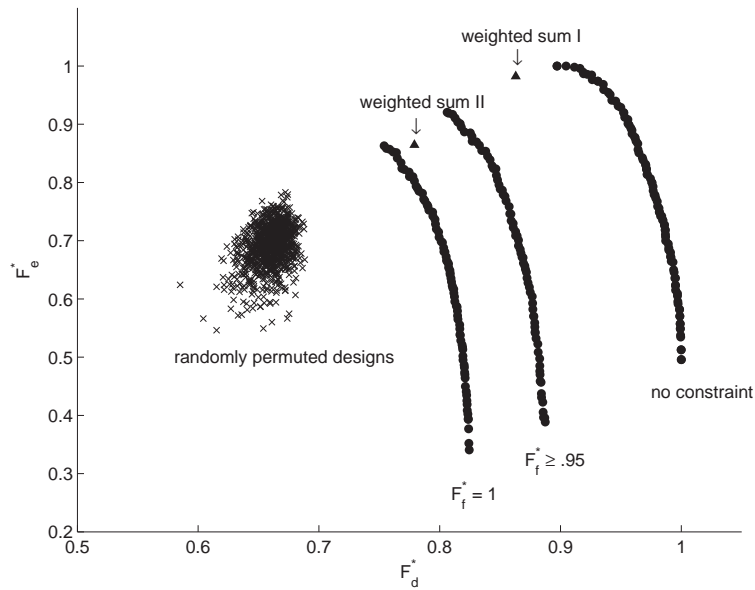


Fig. 7. F_e^* - against F_d^* -values of 1) 1000 random permutations of designs with 22 LCTs, 8 SCTs, 8 NCTs and 29 zeros; 2) MO designs of our approach with $F_f^* = 1$, $F_f^* \geq .95$, and $F_f^* \geq 0$; 3) MO designs of the weighted sum approach maximizing (I) $\frac{1}{3}F_d^* + \frac{1}{3}F_e^* + \frac{1}{3}F_f^*$, and (II) $\frac{1}{6}F_d^* + \frac{1}{6}F_e^* + \frac{2}{3}F_f^*$

Our approach does not have trouble in accommodating the constraint to obtain good MO fMRI designs. The efficiencies of the obtained designs are shown in Figure 7 (the cluster of dots that are labeled with $F_f^* \geq 0.95$). From these designs, we could, as an example, select the one with approximately equal F_d^* - and F_e^* -values. One can obviously make other choices based on needs or preferences. The design that we select is presented in Figure 8. Among the 67 elements, there are 27 ones (LCTs), 12 twos (SCTs), and 12 threes (NCTs); the remaining elements are zeros. While the frequencies of LCTs is not quite three times as much as that of SCTs or NCTs (we did after all not insist that $F_f^* = 1$), the design meets the spirit of the experimenters requirement, and achieves high efficiencies for estimation and detection. In addition, although the unpredictability of designs is not explicitly taken into account, the obtained design does not seem to have a perceivable pattern (except for a larger number of LCTs). This agrees with Liu (2004), where designs with high estimation efficiencies are observed to be unpredictable; see also a discussion in Section 5.

Figure 7 also presents the efficiencies of designs that are random permutations of a sequence with 22 LCTs, 8 SCTs, 8 NCTs and 29 zeros, which is how Brendel et al. (2010)

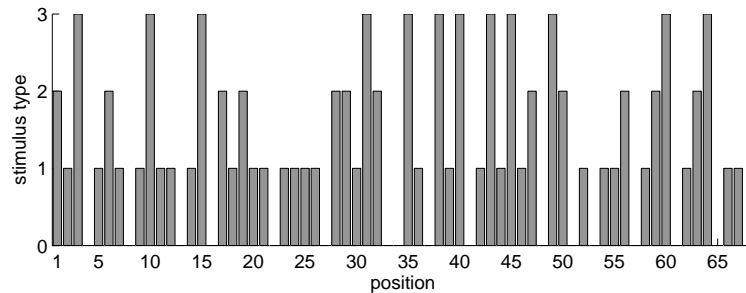


Fig. 8. The design obtained by our approach with $(F_d^*, F_e^*, F_f^*) = (0.847, 0.845, 0.953)$; this design has 27 LCTs (=1), 12 SCTs (=2), and 12 NCTs (=3).

obtained their design. Obtaining designs via random permutation is not uncommon in fMRI. As presented in Figure 7, random permutation does typically not result in designs with high efficiencies. Specifically, the F_d^* -values of the 1000 randomly permuted designs do not exceed 0.7 and their F_e^* -values are less than 0.8. Designs that we obtain are better.

Moreover, we modify the MATLAB program of Kao (2009) and use it to obtain designs for the current experimental setting. Appropriate weights of objectives are unknown, and we simply consider equal weights. The efficiencies of the obtained design are presented in Figure 7, where it is labeled as “weighted sum I”. This design achieves an F_f^* -value of 0.877; it fails to satisfy the requirement of $F_f^* \geq 0.95$. We then consider to maximize $F = \frac{1}{6}F_d^* + \frac{1}{6}F_e^* + \frac{2}{3}F_f^*$; i.e. putting more weight on F_f^* . The obtained design is labeled as “weighted sum II” in Figure 7. It achieves an F_f^* -value of 0.991. Although the constraint on stimulus frequency is satisfied, the (F_d^*, F_e^*) -value is sacrificed for such a high F_f^* -value. Designs with better (F_d^*, F_e^*) -values are obtained by our approach under the constraint; these designs are to the upper-right of the second weighted sum design. Moreover, assigning equal weights on estimation and detection tends to produce designs that are in favor of estimation. This is also observed in the previous subsections. While one can keep altering the weights for a more suitable design, this trial-and-error method is tedious and time consuming. Our approach is much easier to use and it saves time.

The efficiencies of designs that are obtained with a more strict constraint, $F_f^* = 1$, are also presented in Figure 7. As expected, these designs perform worse in estimation and detection due to a higher value of F_f^* . These designs can be considered when the experimenter demands a design completely satisfying the specified relative stimulus frequencies. In addition, the (F_d^*, F_e^*) -values achieved without imposing the constraint are also plotted in Figure 7. The stimulus frequency of these designs is close to the optimal stimulus frequency approximated by Liu and Frank (2004). Each symbol (0, 1, 2, or 3) occurs nearly equally often. Further investigations are required to confirm the optimal stimulus frequency for designs with long or varying stimulus durations.

5. Discussion

In this article, we present an efficient approach for obtaining MO designs for fMRI experiments. Our approach combines the NSGA-II and the GA of Kao et al. (2009a), and we consider more general statistical models. The proposed approach can accommodate many

real-life experimental settings to find a class of MO fMRI designs. With this class of designs, the trade-offs between study objectives can be explored and experimenters can choose a design from the class based on their needs.

We demonstrate the advantages of our approach over the popular weighted sum method and the NSGA-II using various experimental settings. Compared with the weighted sum method, our approach uses much less CPU time and achieves good designs that are nearly evenly distributed on the approximated Pareto front. Our case studies also show that designs obtained by the NSGA-II may not achieve F_d^* -values as high as those for designs obtained by our approach (e.g., Figure 4B). This drawback of the NSGA-II can be alleviated by increasing the number of generations. For example, using the same scenario as in Subsection 4.1, the NSGA-II achieves a maximal F_d^* value of 0.994 after 100,000 generations. However, this F_d^* -value is still slightly smaller than that achieved by our approach with 2,500 generations. Instead of investing much more CPU time, we take advantages of current knowledge about the performance of fMRI designs, and utilize good initial designs to greatly improve the efficiency of the search for high-quality fMRI designs. Our proposed approach largely saves resources.

While we consider A -optimality in our case studies, our approach can also accommodate the D -optimality criterion. The selection of the optimality criterion should be guided by the needs and preferences of the experimenter. The A -optimality criterion may be considered when minimizing the average variance of parameter estimators is of interest. On the other hand, D -optimality aims at minimizing the volume of a simultaneous confidence ellipsoid of the parameters; see Kao et al. (2009a), Maus et al. (2010a), and Maus et al. (2010b) for formulations of the D -optimality criterion for fMRI. Our approach can be easily extended to accommodate other optimality criteria as well.

When evaluating the performance of a design, we focus on the designs ability to detect activation, estimate the HRF, and meet requirements for the relative stimulus frequencies. Our approach can also accommodate additional performance criteria such as the counterbalancing criterion, F_c^* , for avoiding patterned, predictable designs; see Kao et al. (2009a) and Wager and Nichols (2003). In our experience, designs with high F_c^* -values tend to have high F_d^* -values, and hence, are not easy to predict. This observation agrees with Liu (2004). We also observe a trade-off relationship between F_c^* and F_d^* .

In our case studies, we use AR(1) noise with autocorrelation coefficient $\rho = 0.3$. Based on the results of Maus et al. (2010a), A -optimal designs obtained with $\rho = 0.3$ are quite robust against other $\rho \in [0, 0.5]$. Therefore, it is not unreasonable to choose $\rho = 0.3$ at the design stage. Alternatively, if a prior estimate of ρ is available, one could use that value; see also Wager and Nichols (2003).

Moreover, the HRF shape that we consider is the double-gamma function of SPM normalized to have a maximum of one. However, our approach is not restricted to this particular HRF shape and can accommodate other HRF shapes as well. We also assume the linear time invariant system; see Subsection 3.1. Some studies suggest that the assumption might not hold and that the BOLD response might have a nonlinear effect for some situations, e.g., when the stimuli are too close, say $< 4s$ (Wager et al., 2005; Soltysik et al., 2004). Wager and Nichols (2003) propose a simple way to deal with this nonlinear effect. They set a ceiling for the accumulated HRFs. Similar to Wager and Nichols (2003) and Kao (2009), our approach can be modified to find designs for cases where such a ceiling is imposed.

Acknowledgement

The research of Abhyuday Mandal was in part supported by NSF Grant DMS-09-05731, and that of John Stufken by NSF Grants DMS-07-06917 and DMS-10-07507. We thank Professor D. Majumdar for helpful discussions, and anonymous referees for raising questions that resulted in an improvement of this article.

References

- Atkinson, A. C., A. Donev, and R. Tobias (2007). *Optimum Experimental Designs, with SAS*. Great Britain: Oxford U Pr.
- Bailey, R. A. (2007). Designs for two-colour microarray experiments. *Applied Statistics* **56**, 365–394.
- Birn, R. M., R. W. Cox, and P. A. Bandettini (2004). Experimental designs and processing strategies for fMRI studies involving overt verbal responses. *NeuroImage* **23**, 1046–1058.
- Bookheimer, S. (2007). Pre-surgical language mapping with functional magnetic resonance imaging. *Neuropsychology Review* **17**, 145–155.
- Brendel, B., I. Hertrich, M. Erb, A. Lindner, A. Riecker, W. Grodd, and H. Ackermann (2010). The contribution of mesiofrontal cortex to the preparation and execution of repetitive syllable productions: An fMRI study. *NeuroImage* **50**, 1219–1230.
- Brown, G. (2007). Functional magnetic resonance imaging in clinical practice: Look before you leap. *Neuropsychology Review* **17**, 103–106.
- Buračas, G. T. and G. M. Boynton (2002). Efficient design of event-related fMRI experiments using m-sequences. *NeuroImage* **16**, 801–813.
- Buxton, R. B., T. T. Liu, A. Martinez, L. R. Frank, W. M. Luh, and E. C. Wong (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 A. Kingstone (2006). *Handbook of Functional Neuroimaging of Cognition* (2nd ed.). Cognitive Neuroscience. Cambridge, Mass.: MIT Press.
- Cook, R. D. and W. K. Wong (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* **89**, 687–692.
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping* **8**, 109–114.
- Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms* (1st ed.). Wiley-Interscience Series in Systems and Optimization. Chichester; New York: John Wiley & Sons.
- Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* **6**, 182–197.

- Ding, R., D. K. J. Lin, and D. Wei (2004). Dual-response surface optimization: A weighted mse approach. *Quality Engineering* **16**, 377 – 385.
- Friston, K. J., A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. R. Williams, R. S. J. Frackowiak, and R. Turner (1995). Analysis of fMRI time-series revisited. *NeuroImage* **2**, 45–53.
- Friston, K. J., E. Zarahn, O. Josephs, R. N. A. Henson, and A. M. Dale (1999). Stochastic designs in event-related fMRI. *NeuroImage* **10**, 607–619.
- Godfrey, K. (1993). *Perturbation Signals for System Identification*. New York: Prentice Hall.
- Goos, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments*. New York: Springer.
- Grinband, J., T. D. Wager, M. Lindquist, V. P. Ferrera, and J. Hirsch (2008). Detection of time-varying signals in event-related fmri designs. *Neuroimage* *43*(3), 509–520.
- Kao, M.-H. (2009). Multi-objective optimal experimental designs for ER-fMRI using MATLAB. *Journal of Statistical Software* **30**, 1–13.
- Kao, M.-H., A. Mandal, N. Lazar, and J. Stufken (2009a). Multi-objective optimal experimental designs for event-related fMRI studies. *NeuroImage* **44**, 849–856.
- Kao, M.-H., A. Mandal, and J. Stufken (2008). Optimal design for event-related functional magnetic resonance imaging considering both individual stimulus effects and pairwise contrasts. *Statistics and Applications* **6**(1&2), 235–256.
- Kao, M.-H., A. Mandal, and J. Stufken (2009b). Efficient designs for event-related functional magnetic resonance imaging with multiple scanning sessions. *Communications in Statistics-Theory and Methods* **38**, 3170–3182.
- Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **21**, 272–319.
- Lazar, N. A. (2008). *The Statistical Analysis of Functional MRI Data*. Statistics for Biology and Health. New York: Springer.
- Lindquist, M. A. (2008). The statistical analysis of fMRI data. *Statistical Science* **23**, 439–464.
- 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 L. R. Frank (2004). Efficiency, power, and entropy in event-related fMRI with multiple trial types: Part I: Theory. *NeuroImage* **21**, 387–400.
- Liu, T. T., L. R. Frank, E. C. Wong, and R. B. Buxton (2001). Detection power, estimation efficiency, and predictability in event-related fMRI. *NeuroImage* **13**, 759–773.
- Loh, J. M., M. A. Lindquist, and T. D. Wager (2008). Residual analysis for detecting mis-modeling in fMRI. *Statistica Sinica* *18*(4), 1421–1448.

- Lu, Y., A. P. Bagshaw, C. Grova, E. Kobayashi, F. Dubeau, and J. Gotman (2006). Using voxel-specific hemodynamic response function in EEG-fMRI data analysis. *NeuroImage* **32**, 238–247.
- MacWilliams, F. J. and N. J. A. Sloane (1977). *The Theory of Error Correcting Codes*. Amsterdam: North-Holland.
- Marler, R. T. and J. S. Arora (2004). Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization* **26**, 369–395.
- Marler, R. T. and J. S. Arora (2010). The weighted sum method for multi-objective optimization: New insights. *Structural and Multidisciplinary Optimization* **41**, 853–862.
- Martin, P. I., M. A. Naeser, K. W. Doron, A. Bogdan, E. H. Baker, J. Kurland, P. Renshaw, and D. Yurgelun-Todd (2005). Overt naming in aphasia studied with a functional MRI hemodynamic delay design. *NeuroImage* **28**, 194–204.
- Maus, B., G. J. P. van Breukelen, R. Goebel, and M. P. F. Berger (2010a). Robustness of optimal design of fMRI experiments with application of a genetic algorithm. *NeuroImage* **49**, 2433–2443.
- Maus, B., G. J. P. van Breukelen, R. Goebel, and M. P. F. Berger (2010b). Optimization of blocked designs in fmri studies. *Psychometrika* *75*(2), 373–390.
- Mechelli, A., R. N. A. Henson, C. J. Price, and K. J. Friston (2003). Comparing event-related and epoch analysis in blocked design fMRI. *NeuroImage* **18**, 806–810.
- Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. International Series in Operations Research & Management Science. Boston: Kluwer Academic Publishers.
- Murphy, K., V. Dixon, K. LaGrave, J. Kaufman, R. Risinger, A. Bloom, and H. Garavan (2006). A validation of event-related fMRI comparisons between users of cocaine, nicotine, or cannabis and control subjects. *American Journal of Psychiatry* **163**, 1245–1251.
- Ogawa, S., T. M. Lee, A. S. Nayak, and P. Glynn (1990). Oxygenation-sensitive contrast in magnetic-resonance image of rodent brain at high magnetic-fields. *Magnetic Resonance in Medicine* **14**, 68–78.
- Rameson, L. T., A. B. Satpute, and M. D. Lieberman (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage* **50**, 701–708.
- Soltysik, D. A., K. K. Peck, K. D. White, B. Crosson, and R. W. Briggs (2004). Comparison of hemodynamic response nonlinearity across primary cortical areas. *NeuroImage* **22**, 1117–1127.
- Summerfield, C., T. Egner, J. Mangels, and J. Hirsch (2006). Mistaking a house for a face: Neural correlates of misperception in healthy humans. *Cerebral Cortex* **16**, 500–508.
- Wager, T. D. and T. E. Nichols (2003). Optimization of experimental design in fMRI: A general framework using a genetic algorithm. *NeuroImage* **18**, 293–309.
- Wager, T. D., A. Vazquez, L. Hernandez, and D. C. Noll (2005). Accounting for nonlinear bold effects in fMRI: Parameter estimates and a model for prediction in rapid event-related studies. *NeuroImage* **25**, 206–218.

- Wang, Y., G. Xue, C. Chen, F. Xue, and Q. Dong (2007). Neural bases of asymmetric language switching in second-language learners: An ER-fMRI study. *NeuroImage* **35**, 862–870.
- Wierenga, C. and M. Bondi (2007). Use of functional magnetic resonance imaging in the early identification of Alzheimer’s disease. *Neuropsychology Review* **17**, 127–143.
- Worsley, K. J. and K. J. Friston (1995). Analysis of fMRI time-series revisited—again. *NeuroImage* **2**, 173–181.
- Zitzler, E., L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca (2003). Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation* **7**, 117–132.

Appendix I

The design matrix \mathbf{X}_q in model (1) is a zero-one matrix for the q th type stimuli; $q = 1, \dots, Q$. To obtain \mathbf{X}_q , we first construct a working matrix \mathbf{W}_q having entries 0 and 1. The first column of \mathbf{W}_q indicates the times of the onsets of the q th-type stimuli in terms of multiples of ΔT . The next columns are obtained by shifting the elements of the previous column one position down and adding a 0 at the top. The matrix \mathbf{X}_q is then obtained by deleting rows from \mathbf{W}_q , only keeping rows $(1 + (i - 1)m_{TR})$, where $m_{TR} = TR/\Delta T$ and $i = 1, 2, \dots, 1 + \lfloor (T - 1)/m_{TR} \rfloor$. Note that the dimension of \mathbf{X}_q is T -by- k , where $k = 1 + \lfloor 32/\Delta T \rfloor$ for brief (~ 1 s) stimuli or $k = 1 + \lfloor (32 + \tau_{dur})/\Delta T \rfloor$ for stimuli with a longer duration τ_{dur} seconds (Subsection 3.1).

As an illustrative example, we consider a design $\mathbf{s} = \{101210\dots1\}$ of brief stimuli with $TR = 2$ s and $ISI = 3$ s; thus, $\Delta T = 1$ s, $m_{TR} = 2$ and $m_{ISI} = ISI/\Delta T = 3$. The stimuli of the first type occur at time 0, $2ISI = 6\Delta T$, $4ISI = 12\Delta T$, and so on (see Section 2). The first column of \mathbf{W}_1 is therefore $\mathbf{w}_1 = (100000100000100000\dots100)'$ with 1 occurring at positions $1(= 0 + 1)$, $7(= 6 + 1)$, $13(= 12 + 1)$, and so on to indicate the onsets of the first-type stimuli. The j th column of \mathbf{W}_1 is $\mathbf{L}^{j-1}\mathbf{w}_1$, where $j = 1, \dots, 33$, and

$$\mathbf{L} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Similarly, the j th column of \mathbf{W}_2 is $\mathbf{L}^{j-1}\mathbf{w}_2$, where $\mathbf{w}_2 = (000000000100000000\dots000)'$. That the 10th element of \mathbf{w}_2 is 1 is a consequence of the second-type stimulus occurs at time $3ISI = 9\Delta T$. We then obtain \mathbf{X}_q by deleting even rows and keeping odd rows of \mathbf{W}_q ; $q = 1, 2$.

Appendix II

Following is a list of selected terminology and notation used in this paper.

BOLD: blood oxygenated level dependent.

GA: genetic algorithm.

fMRI or MRI: (functional) magnetic resonance imaging.

HRF: hemodynamic response function, a function of time describing the change of the MRI signals evoked by one single stimulus.

ISI: inter-stimulus interval, the pre-specified minimum time (in seconds) between the onsets of two consecutive stimulus presentations.

max- F_d design: the best design for detection that maximizes F_d .

max- F_e design: the best design for estimation that maximizes F_e .

MO: multi-objective.

NSGA-II: the nondominated sorting genetic algorithm II.

SPM: statistical parametric mapping, a popular software package for analyzing fMRI data.

TR: time to repetition, the pre-specified time (in seconds) between two consecutive MRI scans of the same voxel.

voxel: three-dimensional imaging unit; e.g. a $3 \times 3 \times 5$ mm^3 box.

WS-0.01: the weighted sum method with mesh size 0.01.

WS-0.05: the weighted sum method with mesh size 0.05.

L : design length.

Q : total number of stimulus types.

T : length of the fMRI time series \mathbf{y} .

ΔT : the greatest real value making both $(ISI/\Delta T)$ and $(TR/\Delta T)$ integers

m_{ISI} : $(ISI/\Delta T)$

m_{TR} : $(TR/\Delta T)$

θ_q : unknown amplitude (or maximal height) of the HRF for the q th-type stimuli.

\mathbf{h}_q^* : a vector of known function values of the assumed shape of the HRF for the q th-type stimuli.

\mathbf{h}_q : a vector of unknown parameters corresponding to the function values of the unknown HRF for the q th-type stimuli.

τ_{dur} : duration of the boxcar function $b(t)$ in (2).