Model-Robust Two-Level Designs Using Coordinate Exchange Algorithms and a Maximin Criterion

Byran J. Smucker[1]*, Enrique del Castillo[2], and James L. Rosenberger[3]

[1]Department of Statistics

Miami University, Oxford OH 45056

[2]Department of Industrial & Manufacturing Engineering

The Pennsylvania State University, University Park PA 16802

[3]Department of Statistics

The Pennsylvania State University, University Park PA 16802

Abstract

We propose a candidate-list-free exchange algorithm which facilitates construction of exact, model-robust, two-level experiment designs. In particular, we investigate two model spaces previously considered in the literature. The first assumes that all main effects and an unknown subset of two-factor interactions are active, but that the experimenter knows the number of active interactions. The second assumes that an unknown subset of the main effects, and all associated two-factor interactions, are active. Previous literature utilizes two criteria for design construction: First, maximize the number of estimable models; then, differentiate between designs equivalent in estimability by choosing the design with the highest average $\mathcal{D}$-efficiency. We adopt a similar strategy, but (1) do not impose orthogonality or factor level balance constraints, resulting

*corresponding author, absmucker@gmail.com

in generally equal or larger numbers of estimable models, and (2) use a flexible secondary criterion which maximizes the minimum $\mathcal{D}$-efficiency. We provide results for many situations of interest.

Keywords: $\mathcal{D}$-efficiency, exact experimental design, optimal design, two-level design, model-robust, maximin

1 Introduction

Among two-level experiment designs used to fit models that include two-factor interactions, resolution V fractional factorial designs are the classical gold standard due to their ability to estimate all two-factor interactions (assuming larger-order interactions are negligible). However, as the number of factors grows, resolution V designs become quite large and thus may be prohibitively expensive to run. For instance, with 8 factors, a resolution V design requires 64 runs. Resolution IV designs are smaller alternatives (as few as 16 runs for 8 factors), but they do not allow estimation of all two-factor interactions. In particular, if it is unknown which main effects and/or two-factor interactions will be active, the experimenter will only be able to estimate the true model in the serendipitous case that the truly active effects happen to be the ones the design can estimate.

An alternative design strategy, presented in two different contexts by Li and Nachtsheim (2000) and Loeppky et al. (2007), considers design construction in these situations as a problem of model-robustness. In short, they propose a set of models for which estimability is desired, then search through some space of designs to (1) maximize the number of estimable models in the set; and (2) maximize the average $\mathcal{D}$-efficiency (defined in §4) with respect to all models.

Here, we present a new set of algorithms, based upon coordinate exchange (Meyer and Nachtsheim 1995), which construct designs with generally higher or equal numbers of estimable models than those in the literature, and use a secondary, maximin criterion to encourage high efficiency with respect to the models. We are able to make substantial estimability gains because unlike Li and Nachtsheim (2000) and Loeppky et al. (2007) we

do not require either orthogonality or factor level balance (that is, each column in the design having the same number of high and low values). The maximin criterion optimizes the worst-case efficiency with respect to the specified set of models and easily admits a generalization which allows the specification of model interest levels, providing additional flexibility should some models be favored over others.

This article is arranged as follows. In the next section, §2, we outline the model-robustness problem and review the literature. In §3 we give a description of classic exchange algorithms and discuss the model spaces of interest and the approaches given in the literature. In §4, we present our approach and algorithms. §5 presents results and comparisons with designs in the literature. We conclude in §6 with a discussion.

2 Model-Robustness and the Literature

2.1 The Setting

Assume that an experiment is to be designed for which there is a single response and k factors of interest. A general linear regression model with p parameters, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, is to be fit, where $\mathbf{y}$ is an n-vector, $\mathbf{X}$ is the $n \times p$ expanded design matrix, $\boldsymbol{\beta}$ is the $p \times 1$ vector of unknown parameters, and $\boldsymbol{\epsilon}$ is also an n-vector with independent elements, $E(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \sigma^2$. Assume also that the classical least squares criterion will be used to estimate $\boldsymbol{\beta}$, in which case the estimator is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ with $Var(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. The inverse of $Var(\hat{\boldsymbol{\beta}})$ is called the information matrix, denoted as $\mathbf{M} = (\mathbf{X}'\mathbf{X})/\sigma^2 = [Var(\hat{\boldsymbol{\beta}})]^{-1}$. At the design stage, σ^2 is generally not known but since it is a constant multiple of $\mathbf{X}'\mathbf{X}$ we take it to be 1 without loss of generality.

Furthermore, let $\mathcal{X}$ be the design space, Ξ be the set of all possible designs and $\xi_n \in \Xi$ be a discrete, n-point design which can be represented as

$$\xi_n = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \dots & \mathbf{x}_d \\ n_1 & n_2 & \dots & n_d \end{pmatrix} \tag{1}$$

where $n = \sum_{i=1}^{d} n_i$ is the total number of runs, and n_i is the number of runs performed at design point $\mathbf{x}_i$, $i = 1, \ldots, d$.

An optimal design strategy chooses a design by optimizing a function of the information matrix. There are many design optimality criteria, some of which focus on precise parameter estimation and some which focus on precise prediction. For the former, the $\mathcal{D}$-criterion minimizes the volume of the confidence ellipsoid about the parameter estimates (assuming σ^2 is known) by maximizing $\phi(\mathbf{M}) = |\mathbf{M}|$, where $|\cdot|$ is the matrix determinant. While this criterion has several favorable characteristics, its computational superiority—due to quick matrix- and determinant-updating formulae—is a preeminent positive property. However, since in the present setting two-level designs will be used to determine significant factors and interactions, the $\mathcal{D}$-criterion's focus on estimation is appropriate. In this article, we build our methods upon the $\mathcal{D}$-criterion.

Implicit in the structure of $\mathbf{X}$ and by extension $\mathbf{M}$, however, is not only a design but also a model form, $f(\mathbf{x})$, which is a $p \times 1$ expanded design vector that is usually assumed given though in reality is rarely known. Consequently, we represent the information matrix as $\mathbf{M}_f(\xi_n)$, though we may express it as $\mathbf{M}_f$ in the interest of simplicity. Thus, $\phi(\mathbf{M}_f(\xi_n))$ depends upon the model form f, and an optimal design for one model is not in general optimal for another. Later, we discuss in detail the sorts of model forms considered in this work as well as a solution to this dependence.

2.2 Review of Model-Robust Design Literature

The dependence of optimal designs on the form of the model was first noted by Box and Draper (1959). Much of the subsequent work in model-robustness emphasizes asymptotic designs and either followed in the steps of Box and Draper by attempting to protect against a larger, potential model (e.g. Montepiedra and Fedorov 1997; Fang and Wiens 2003), utilized constrained designs which optimize a function of a primary model subject to others with a certain efficiency level (e.g. Stigler 1971; Studden 1982; Dette and Franke 2000), or used as a criterion a function of a known class of models (e.g. Läuter 1974; Cook and Nachtsheim

1982; Dette 1990; Dette and Franke 2001). Pukelsheim and Rosenberger (1993) discussed several of these model-robust strategies.

Model-robust methodology for exact designs has a smaller, though still significant, literature. Welch (1983) and Fang and Wiens (2000) both utilized mean squared error approaches. DuMouchel and Jones (1994) used a Bayesian paradigm to provide protection against "potential" terms while retaining a good design for "primary" terms, and Jones and Nachtsheim (2011) use the same primary/potential paradigm to construct designs whose alias matrices have desirable properties subject to efficiency constraints on the primary model. Heredia-Langner et al. (2004) and Smucker et al. (2011) optimized functions of a user-specified set of models to provide designs robust for any model in the set. Tsai and Gilmour (2010) use an approximation to $\mathcal{A}_s$ optimality (where the $\mathcal{A}_s$ criterion minimizes the sum of a subset of the parameter estimate variances) to achieve model-robustness for all possible subsets of a specified largest model. Model-robustness has also been considered when only main effects and two-factor interactions are of interest. In an article that predated Li and Nachtsheim (2000) and Loeppky et al. (2007) which were discussed at the outset, Cheng et al. (1999) studied the connection between minimum aberration in regular fractional factorial designs and model-robustness.

Several authors have studied the maximin criterion in the context of experiment design, including Thibodeau (1977), who utilized a result that shows $\mathcal{D}$-efficiency is always at least as large as $\mathcal{G}$-efficiency for continuous designs. Imhof and Wong (2000) give a graphical method to find asymptotic maximin designs when two different optimality criteria are used. Dette and Franke (2001) explicitly characterize continuous maximin designs in the specific case of polynomial regression on $[-1, 1]$, where they maximize the minimum efficiency with respect to an assumed polynomial of degree m as well as efficiencies related to a goodness-of-fit criterion. Berger and Tan (2004) used a maximin criterion for mixed models.

In contrast to all the work above, this work seeks to provide a useful, flexible tool for experimenters to construct exact, completely randomized, two-level designs which allow comprehensive estimability and, once that is established, maximize the worst-case efficiency.

3 Building Blocks for the New Approach

In this section we review exchange algorithms, since our procedures utilize them. We also describe two model spaces, and the approaches taken to design with them in mind, that arise in contexts that would typically suggest fractional factorial designs.

3.1 Classic Exchange Algorithms

The idea that buttresses the original exchange algorithm (Fedorov 1972) is simple. Starting with a nonsingular design, consider exchanges between each design point (i.e. each row) and each member of a candidate list of points covering the design space, $\mathcal{X}$. Choose the exchange which most increases the determinant of the information matrix, and repeat until convergence. The computational requirements of this brute force method are lessened somewhat by a quick determinant-updating formula. But even with this mathematical device, Fedorov's procedure will sputter for large problems, since each iteration searches the candidate list n times but only makes a single exchange. Consequently, many alternatives have been proposed (Mitchell 1974; Cook and Nachtsheim 1980; Johnson and Nachtsheim 1983; Meyer and Nachtsheim 1995; Atkinson et al. 2007). Using the maximin criterion we describe, any of these algorithms could be generalized to the model-robust context.

The coordinate-exchange algorithm (Meyer and Nachtsheim 1995), however, is particularly attractive because it does not require a candidate list, which if unchecked grows exponentially as the size of the problem increases. This candidate-list-free property is a crucial advantage and the reason we use it as the basis for our algorithms. Specifically, and following Meyer and Nachtsheim (1995), our algorithm considers the exchange between a current design coordinate $\tilde{x}_{ij}$ and $\tilde{x}_j \in \mathcal{X}_j$ (where $\mathcal{X}_j$ is the portion of the design space associated with the j^{th} factor). We also use $\tilde{x}_{i,-j}$ to denote the portion of the design point $\mathbf{x}_i$ that is unaffected by the exchange. Then, we can write, for any point $\mathbf{x}$, $f(\mathbf{x}) = \begin{pmatrix} f_1(\tilde{x}_j) \\ f_2(\tilde{x}_{-j}) \end{pmatrix}$. We also define

$$A_f = (1 - V_f(\mathbf{x}_i, \xi_n))\mathbf{M}_f^{-1} + \mathbf{M}_f^{-1} f(\mathbf{x}_i) f^{\mathrm{T}}(\mathbf{x}_i) \mathbf{M}_f^{-1} \tag{2}$$

where $V_f(\mathbf{x}_i, \xi_n) = f^{\mathrm{T}}(\mathbf{x}_i)\mathbf{M}^{-1}(\xi_n)f(\mathbf{x}_i)$ is the prediction variance at $\mathbf{x}_i$. A_f is partitioned according to f as $A_f = \begin{pmatrix} A_{f11} & A_{f12} \\ A_{f21} & A_{f22} \end{pmatrix}$. Then, Meyer and Nachtsheim (1995) showed that the multiplicative difference in the determinant given such a coordinate exchange can be expressed as

$$\Delta_f^{ij}(\tilde{x}_{ij}, \tilde{x}_j, \tilde{x}_{i,-j}) = f_1^{\mathrm{T}}(\tilde{x}_j)A_{f11}f_1(\tilde{x}_j) + a^{\mathrm{T}}f_1(\tilde{x}_j) + c, \tag{3}$$

where $a = 2A_{f12}f_2^{\mathrm{T}}(\tilde{x}_{i,-j})$ and $c = f_2^{\mathrm{T}}(\tilde{x}_{i,-j})A_{f22}f_2(\tilde{x}_{i,-j}) + (1 - V_f(\mathbf{x}_i, \xi_n))$ do not change on account of the exchange. We also use an updating formula for the information matrix:

$$\mathbf{M}_f(\tilde{\xi}_n) = \mathbf{M}_f(\xi_n) + \mathbf{F}_1\mathbf{F}_2^{\mathrm{T}} \tag{4}$$

where $\mathbf{F}_1 = [-f(\mathbf{x}_j), f(\mathbf{x})]$ and $\mathbf{F}_2 = [f(\mathbf{x}_j), f(\mathbf{x})]$.

Note that exchange algorithms are heuristics which do not necessarily converge to a global optimum. Thus, the algorithms should be run multiple times from a variety of randomly generated initial designs in order to find a near-optimal design.

3.2 The MEPI Model Space

Li and Nachtsheim (2000) presented what they called *model-robust factorial designs*. These designs are robust for the model space that consists of k main effects and g of the possible two-factor interactions (called the "main effects plus interaction", or MEPI, model space). For instance, if $k = 5$ and $g = 3$, there are $\binom{5}{2} = 10$ possible two-factor interactions and consequently $\binom{10}{3} = 120$ different models with all five main effects and three two-factor interactions. This model space can be roughly justified via the sparsity-of-effects principle: Though many factors and interactions are possible, it is likely that only a few are active.

The design construction strategy of Li and Nachtsheim (2000) uses a primary and secondary criterion. The principal goal is to maximize the *estimation capacity* (EC), which is defined as the proportion of models that are estimable. To distinguish among designs which possess the same EC, a secondary criterion is considered: maximize something similar to

the average $\mathcal{D}$-efficiency with respect to all models in the set. The procedure used is the so-called restricted columnwise-pairwise algorithm in which iterations are made through the design columns, and in each column an element is randomly chosen and exchanges with all other elements are considered. The swap which most increases the estimation capacity is executed (i.e. the exchange that results in the largest number of estimable models); if more than one results in the highest estimation capacity, the one is chosen that maximizes the average $\mathcal{D}$-efficiency. Since this algorithm makes columnwise exchanges, factor level balance is enforced; i.e. each column in the design has the same number of high and low values.

3.3 The Projective Model Space

Loeppky et al. (2007) consider what has been called the projective model space (PMS), also based upon the sparsity-of-effects principle. Assume that $h \leq k$ of the main effects are active, along with the associated $\binom{h}{2}$ two-factor interactions. This results in $\binom{k}{h}$ possible models for any particular h. If such model spaces are defined for all $h \leq k$, the *projection estimation capacity* (PEC) sequence can be defined, which, for a particular design, gives the proportion of estimable models for all $h \leq k$. For designs that have the same PEC sequence, Loeppky et al. (2007) also use a criterion similar to average $\mathcal{D}$-efficiency as a secondary criterion.

Loeppky et al. (2007) are able to find compelling alternatives to typical designs. For instance, assuming $k = 7$ factors and an anticipated model consisting of main effects and two-factor interactions, the minimum aberration 2^{7-1} resolution VII design has 64 runs and a PEC sequence of $(1, 1, 1, 1, 1, 1, 1)$. Because the resolution VII design is run-intensive, one might consider instead the minimum aberration 2^{7-2} resolution IV design. This has 32 runs but a PEC sequence of $(1, 1, 1, .971, .857, .571, 0)$, which implies that if 4 or more factors are active, the experimenter will be unable to estimate all projective models. In contrast, Loeppky et al. (2007) constructed a 20-run design with a PEC sequence of $(1, 1, 1, 1, 1, 0, 0)$ and a 24-run design with a PEC sequence of $(1, 1, 1, 1, 1, 1, 0)$.

To find designs which perform well for each element in the sequence, Loeppky et al.

proposed two approaches: "bottom-up" and "top-down". The bottom-up algorithm seeks to sequentially maximize the PEC sequence starting with the first element that is not theoretically guaranteed to be 1. The top-down method optimizes elements in descending order, retaining as constraints each previously optimized value. Most of the designs in Loeppky et al. (2007) were constructed via a computationally simple algorithm which uses projections of Hadamard matrices. While computationally feasible and ensuring orthogonality, it is orthogonality itself that is ultimately a limiting factor in the effectiveness of the designs in terms of their own criteria. This has been noted elsewhere in the literature in a slightly different context (see Xu et al. 2009).

4 Coordinate Exchange Algorithms for Model-Robust Two-Level Experiment Designs

We consider both the main effect plus interaction (MEPI) and projective model spaces (PMS) using algorithms based upon coordinate exchange. We adopt a two-pronged approach which in many respects is similar to the approaches of Li and Nachtsheim (2000) and Loeppky et al. (2007). However, there are at least three important differences: (1) The algorithms are based on coordinate exchange and as such require no candidate list of design points; (2) The algorithms impose no orthogonality or factor level balance constraints; and (3) The secondary criterion is to maximize the minimum $\mathcal{D}$-efficiency instead of maximizing the average $\mathcal{D}$-efficiency.

An overview of the algorithms is as follows:

1. Initialize the algorithm.

   (a) Construct the model space.

   - If MEPI model space, specify sample size (n) and number of two-factor interactions (g). Form the set of models.
   - If projective model space, specify sample size (n) and the number of factors

associated with the largest element of the PEC sequence (ℓ). Form the set of models.

(b) Construct a randomly chosen initial design.

2. Improve the design by maximizing the estimation capacity (EC), until convergence.

   - Consider exchanges of each element of each design point; e.g. if the first factor in the first run has a value of $+1$, assess the effect of an exchange with -1.
   - For each exchange, use (4) to update the information matrix for possible models.
   - If a coordinate exchange results in a higher EC, it is executed.
   - Continue until convergence.

3. If the $EC = 1$ from Step 2, the design is further improved by increasing the minimum efficiency over all possible models, until convergence. If $EC < 1$, this step is omitted.

   - This is accomplished by considering exchanges of each element of each design point.
   - Equation (3) is used on each information matrix determinant to determine if the maximum $\mathcal{D}$-efficiency is increased.

Steps 2 and 3 of the algorithm employ coordinate exchange. In Step 2, the objective function is EC; in Step 3, the objective function is the minimum $\mathcal{D}$-efficiency (defined below). In the EC optimization, exchanges are considered for each element of a design row. The exchange is made that results in the largest number of estimable models. This is roughly accomplished by updating the information matrices with respect to all models in the model space and determining (via the condition number) whether the determinant is greater than zero (see Supplementary Material for more details).

This procedure presents a computational tradeoff, depending largely on the number of runs in the design. If the number of runs is large enough, the first stage (EC optimization) tends to be fast because it is easy to find a design that estimates all models in the model

space. In this case, most of the computational energy is expended in the second stage (maximin optimization) in an attempt to improve the efficiencies with respect to each model. On the other hand, if the number of runs is small, it is difficult or impossible to find a design with $EC = 1$, in which case the algorithm expends all its effort in this stage and completely skips the second. Illustrative examples are given in the next section. Once the specification of a model space occurs, the algorithms are essentially the same whether the model space is MEPI or projective. These algorithms could easily be extended to work with other model spaces (e.g., supersaturated, Jones et al. 2009).

The maximin portion of the algorithm (Step 3) works on $\mathcal{D}$-efficiencies, which we define now. Assuming ξ_n^* is the optimal n-run design for model f, the $\mathcal{D}$-efficiency for a design ξ_n with respect to model f is defined as:

$$E_{\mathcal{D},f}(\xi_n) = \left(\frac{|\mathbf{M}_f(\xi_n)|}{|\mathbf{M}_f(\xi_n^*)|} \right)^{1/p} \tag{5}$$

where $\mathbf{M}_f(\xi_n)$ is the information matrix with respect to model f and design ξ_n, $\mathbf{M}_f(\xi_n^*)$ is the information matrix for the exact n-point design optimal for f alone (practically, this "optimal" design will often be found by using a discretized version of the design space), and p is the number of parameters for model f. Note that if there is strong prior knowledge regarding certain models in the model space, our procedure can easily and effectively account for this by defining $v_f \in (0, 1]$ as the weight for model f and using it to define a *generalized* $\mathcal{D}$*-efficiency*: $G_f(\xi_n) = E_{\mathcal{D},f}(\xi_n)/v_f$. If G_f is utilized within the algorithm instead of $E_{\mathcal{D},f}$, the prior model knowledge will be reflected in the model-robust design.

The $\mathcal{D}$-efficiency in (5) requires the calculation of the n-point $\mathcal{D}$-optimal designs for each of the models in the set. This can be accomplished using standard software (e.g. $SAS^{\circledR}$ software's PROC OPTEX, MATLAB's *cordexch* function, JMP, Minitab, DesignExpert). For small and even moderately-sized model sets this is not a significant hurdle to overcome, but when the number of models grows into the thousands, it may be desirable to use a measure of efficiency that does not require individually optimal designs. Li and Nachtsheim

(2000) noted that the individually optimal two-level designs are often nearly orthogonal. This leads to the following efficiency measure, which is a lower-bound for the $\mathcal{D}$-efficiency defined in (5):

$$E^o_{\mathcal{D},f}(\xi_n) = \left(\frac{|\mathbf{M}_f(\xi_n)|}{n^p}\right)^{1/p} = \frac{|\mathbf{M}_f(\xi_n)|^{1/p}}{n}. \tag{6}$$

In what follows, we will assume that we are using (5) unless otherwise noted, though we have found little difference between the designs based on these two types of efficiencies.

For a more detailed explanation of the algorithms, see the Supplementary Material accompanying this article.

5 Examples

In this section we present empirical results which illustrate our methodology, for both the MEPI and projective model spaces. In the process, we compare ours with results found in the literature. We demonstrate that, in terms of estimation capacity (EC), when possible (i.e. when the designs in the literature have less than 100% estimation capacity) our designs are better. This improvement is a result of relaxations on the part of our procedures: Unlike Li and Nachtsheim (2000), we do not require factor level balance in the case of the MEPI model space; unlike Loeppky et al. (2007) for the projective model space, we do not require orthogonality. Consequently, our designs are consistently equal or superior to theirs. We note that—as we discuss in the final section—there is a price to pay for these relaxations, primarily in terms of ease of analysis and interpretability. In addition to estimation capacity, we give results for the maximin secondary criterion when appropriate. We also discuss computational efficiency at the end of this section.

Unless otherwise noted, the efficiencies given in the tables are calculated using (5). Also, the results are based upon 20 algorithm tries, each one from a randomly generated starting design.

5.1 Designs for the MEPI Model Spaces

The model space addressed by Li and Nachtsheim (2000) assumes a two-level n-run design with k active main effects but only g active two-factor interactions (2fi's). This leads to a model space with $r = \binom{\binom{k}{2}}{g}$ total models.

Most of the designs in this section are for experiments considered by Li and Nachtsheim (2000), allowing direct comparisons to be made. We describe the designs using several measures. First, the estimation capacity is the proportion of estimable models in the model space. Then, for all EC-optimal designs (i.e. designs that have an estimation capacity not exceeded by any other design), we compare in terms of their average and minimum $\mathcal{D}$-efficiency with respect to all of the estimable models.

In the tables in this subsection, we compare the following: Our candidate-list-free-EC-optimal (CLF-EC) designs (those for which only the EC criterion—not maximin—has been optimized); our candidate-list-free-Maximin (CLF-Maximin) designs; the Model-Robust Factorial Designs (MRFD) of Li and Nachtsheim (2000); and, in certain cases, traditional fractional factorial designs (FFD). Note that when the estimation capacity for the CLF-EC design is less than 1, we do not calculate maximin designs because testing suggests that little or no efficiency gains can be realized.

Tables 1 and 2 give the results of various designs for 12-run and 16-run experiments, respectively. For some situations, such as $\{n = 12, k = 5, g = 5\}$ in Table 1, our procedure allows all models to be estimated whereas the algorithm of Li and Nachtsheim (2000) does not. In others, such as $\{n = 12, k = 7, g = 3\}$, neither Li and Nachtsheim's design (.992) or ours (.997) are able to achieve 100% estimation capacity. In cases that both procedures give designs with $EC = 1$, our designs have larger minimum efficiencies, and the difference is often substantial (e.g., {n=12,k=6,g=3}). Note that in some examples—for instance $\{n = 16, k = 7, g = 3\}$—the maximin design even has a higher average $\mathcal{D}$-efficiency than the MRFD.

When $EC < 1$, a simple remedy would be to include more runs in the experiment, which will enable additional models to be estimated. Since we are using an algorithmic approach,

Table 1: Comparison of designs robust for the MEPI model space with $n = 12$, k factors, g possible two-factor interactions, and a model space with r elements.

k	g	r	*Designs*	EC	$\overline{E}_{\mathcal{D}}$	$\min E_{\mathcal{D}}$
5	2	45	CLF-EC	1	.877	.780
			CLF-Maximin	1	.983	.970
			MRFD	1	.967	.910
5	4	210	CLF-EC	1	.888	.720
			CLF-Maximin	1	.867	.741
			MRFD	.995	-	-
5	5	252	CLF-EC	1	.836	.730
			CLF-Maximin	1	.836	.730
			MRFD	.976	-	-
6	3	455	CLF-EC	1	.850	.649
			CLF-Maximin	1	.871	.725
			MRFD	1	.867	.667
7	3	1330	CLF-EC	.997	.748	.532
			MRFD	.992	-	-
9	2	630	CLF-EC	.994	.672	.411
			MRFD	.986	-	-

we can add as few as one or two additional runs. For instance, if a 13th run is added to the $\{n = 12, k = 9, g = 2\}$ experiment, all of the formerly inestimable models become estimable. This flexibility also highlights a potential drawback of these algorithmically-generated designs: Orthogonality is not explicitly considered as a criterion. Our designs, as well as that of Li and Nachtsheim (2000), sacrifices orthogonality for better estimation properties.

In Section 4, we defined two separate versions of $\mathcal{D}$-efficiency. One, (5), depends upon knowing the optimal designs for each model individually; the other, (6), removes this requirement. All of our designs in this subsection were constructed using the first definition of $\mathcal{D}$-efficiency. However, it appears that little-to-nothing is compromised if the simpler measure is used, based on several tests that we have conducted. For example, for the $\{n = 12, k = 5, g = 5\}$ experiment, the design found using (6) resulted in the same minimum and mean $\mathcal{D}$-efficiencies as the design found using (5). For the $\{n = 16, k = 9, g = 2\}$ experiment, the minimum $\mathcal{D}$-efficiency was about one percentage point lower for the design

Table 2: Comparison of designs robust for the MEPI model space with $n = 16$, k factors, g possible two-factor interactions, and a model space with r elements.

k	g	r	*Designs*	EC	$\overline{E}_{\mathcal{D}}$	$\min E_{\mathcal{D}}$
6	3	455	CLF-EC	1	.770	.671
			CLF-Maximin	1	.898	.840
			FFD	.747	-	-
7	2	210	CLF-EC	1	.824	.740
			CLF-Maximin	1	.895	.838
			MRFD	1	.911	.756
			FFD	.900	-	-
7	3	1330	CLF-EC	1	.753	.636
			CLF-Maximin	1	.878	.788
			MRFD	1	.865	.723
			FFD	.771	-	-
8	2	378	CLF-EC	1	.761	.599
			CLF-Maximin	1	.868	.788
			MRFD	1	.881	.711
			FFD	.889	-	-
10	2	990	CLF-EC	1	.642	.494
			CLF-Maximin	1	.770	.657
			MRFD	1	.798	.598
			FFD	.178	-	-

based upon (6). Other examples give similar results.

5.2 Designs for the Projective Model Space

In the second type of model space under consideration—the so-called projective model space—we assume that only a subset of the k factors, along with the associated two-factor interactions, are active. Loeppky et al. (2007) define the projection estimation capacity (PEC) sequence $(p_1, p_2, \ldots, p_k)$ such that the h^{th} element in the sequence is the proportion of estimable models for the model space composed of $\binom{k}{h}$ models, each possessing h main effects and $\binom{h}{2}$ two-factor interactions. For instance, if $k = 8$, p_5 would be the estimable proportion of the $\binom{8}{5}$ models which are composed of five of the eight main effects and the associated $\binom{5}{2}$ two-factor interactions. Practically speaking, there is a PEC sequence element ℓ which is the largest projection supported by a given sample size n. For instance,

if $n = 20$ and $k = 10$, there are $1 + 6 + \binom{6}{2} = 22$ terms if we project down to 6 factors. Only when the projection is to 5 factors (16 terms) or fewer are there enough degrees of freedom to estimate all parameters. Thus, in this example, $\ell = 5$. As shorthand, we generally give the PEC sequence as $(p_1, p_2, \ldots, p_\ell)$, and denote the model space associated with ℓ as $\mathcal{L}$.

Here, we give selected results for the PMS when $n = 16$ (Table 3), $n = 20$ (Table 4), $n = 24$ (Table 5), and $n = 28$ (Table 6). Most examples are designs for which our procedure can estimate all models; for experiments in which not all elements of the PEC sequence are 1, several possible remedial measures are outlined below. As in the MEPI case, our designs are indicated in the tables as either CLF-EC (where the estimation capacity is with respect to $\mathcal{L}$) or CLF-Maximin (where maximin is with respect to $\mathcal{L}$). The designs of Loeppky et al. (2007) are denoted PEC, PEC-TD (PEC-Top-Down), or PEC-BU (PEC-Bottom-Up); see §3.3. Furthermore, the "r" column gives the number of models in $\mathcal{L}$, and the average and minimum $\mathcal{D}$-efficiencies given in these tables are with respect to $\mathcal{L}$ as well. We also display the relevant portions of the PEC sequence for comparison (unshown PEC sequence elements can be assumed to be 1). As before, we do not pursue maximin designs for those EC-optimal designs whose largest PEC sequence element is less than 1; and we only give the efficiency measures for the designs with the highest estimation capacity for $\mathcal{L}$.

We were unable to replicate some of the estimation capacities given in Loeppky et al. (2007), even after significant effort including interaction with the first author. In certain cases we just give the results as presented in Loeppky et al.; in others, we give our calculations. The discrepancies do not affect the conclusions based upon the comparisons between designs. We have included appropriate footnotes in the tables.

For our designs we use a top-down approach, in the sense that we optimize with respect to the models in $\mathcal{L}$. Sometimes, this results in designs that are uniformly better than either the bottom-up or top-down designs given in Loeppky et al. (2007). But at other times, particularly for designs with fewer runs, a top-down approach results in a significantly higher estimation capacity for the largest element of the PEC sequence, but inferior estimability for one or more of the smaller PEC sequence elements. For instance, in Table 3, the CLF-

Table 3: Comparison of designs robust for the projective model space when $n = 16$, k factors, a model space with r elements, and a portion of the projective estimation capacity sequence (p_3, p_4, p_5).

k	r	*Designs*	(p_3, p_4, p_5)	$\overline{E}_{\mathcal{D}}$	$\min E_{\mathcal{D}}$
6	6	CLF-EC	$(1, 1, .833)$	.623	.545
		PEC*	$(1, 1, .333)$	-	-
10	252	CLF-EC	$(.967, .838, .516)$	.553	.545
		PEC*	$(1, .986, .214)$	-	-
14	2002	CLF-EC	$(.967, .814, .348)$	.575	.545
		PEC*	$(1, .979, .224)$	-	-

* Indicates that the associated PEC sequence is given as reported in Loeppky et al. (2007), though we were unable to replicate it.

Maximin 16-run, 10-factor design, when projected down to five factors, can estimate 51.6% of the 252 models, considerably higher than the design given by Loeppky et al. However, their design has higher estimation capacities for the third and fourth element of the PEC sequence.

Several strategies could be employed to solve this problem. One would be to simply allow more than two levels to be considered, which would eliminate the estimability problem altogether. Another would be to implement a bottom-up procedure using our algorithm. Yet another would be to choose, based upon the results given in Loeppky et al. (2007) or this article, another element of the PEC sequence to optimize. For instance, if we take the 16-run, 10-factor experiment and optimize with respect to the fourth element of the PEC sequence instead of the fifth, our algorithm finds an estimation capacity of 100% for the fourth element (compared with 98.6% in Loeppky et al.; see Table 3). A final alternative would be to increase the number of runs in the experiment. For instance, in Table 4 when there are 10 factors our algorithm fails to find a design with 100% estimation capacity for all elements in the PEC sequence. However, setting the number of runs to 21 results in a design with a PEC sequence of $(1, 1, 1, 1, 1)$. We do not pursue these issues further here.

In surveying the results, it is clear that estimation capacities for $\mathcal{L}$ are generally higher for our designs, unless $EC = 1$ in which case the minimum efficiencies are higher. For instance,

Table 4: Comparison of designs robust for the projective model space when $n = 20$, k factors, a model space with r elements, and a portion of the projective estimation capacity sequence p_5.

k	r	*Designs*	p_5	$\overline{E}_{\mathcal{D}}$	$\min E_{\mathcal{D}}$
6	6	CLF-EC	1	.632	.545
		CLF-Maximin	1	.933	.924
		PEC	1	.905	.903
8	56	CLF-EC	1	.700	.536
		CLF-Maximin	1	.771	.667
		PEC*	.929	-	-
10	252	CLF-EC	.988	.727	.545
		PEC*	.857	-	-
12	792	CLF-EC	.955	.735	.536
		PEC*	.825	-	-

* Indicates that the PEC sequence is given as reported in Loeppky et al. (2007), though we were unable to replicate it.

when $n = 28$ (Table 6) and $k = 8$ we realize a significant minimum $\mathcal{D}$-efficiency increase over the PEC designs (.808 versus .678 and .643) and even a higher average $\mathcal{D}$-efficiency. An anomaly in Table 5 for $k = 8$, in which the PEC design outperforms our maximin design in terms of minimum $\mathcal{D}$-efficiency (.643 to .614), demonstrates that 20 algorithm tries is not always sufficient to find the optimal design (a subsequent run of another 20 algorithm tries gave a maximin design with a minimum efficiency of .679).

5.3 Algorithmic Performance

In this subsection, we give some computational results to demonstrate the capabilities and limitations of our algorithms. Note that the times given in the following tables are in terms of minutes/algorithm try.

In Table 7, we give computational results for a variety of designs using our algorithm with the MEPI model space. As can be seen, our implementation can handle small-to-medium sized problems (in terms of r, the number of models), but as the number of models gets larger, it becomes computationally expensive. Li and Nachtsheim (2000) give some

Table 5: Comparison of designs robust for the projective model space when $n = 24$, k factors, a model space with r elements, and a portion of the projective estimation capacity sequence (p_4, p_5, p_6).

k	r	Design	(p_4, p_5, p_6)	$\overline{E}_{\mathcal{D}}$	$\min E_{\mathcal{D}}$
6	1	CLF-EC	$(1, 1, 1)$	.468	.468
		CLF-Maximin	$(1, 1, 1)$	1	1
		PEC-TD	$(1, 1, 1)$	.864	.864
		PEC-BU	$(1, 1, 1)$	.778	.778
8	28	CLF-EC	$(1, 1, 1)$	.618	.498
		CLF-Maximin	$(1, 1, 1)$	.695	.614
		PEC-TD**	$(1, 1, 1)$	.762	.643
		PEC-TD*	$(1, 1, .964)$	-	-
		PEC-BU	$(1, 1, .786)$	-	-
10	210	CLF-EC	$(1, 1, .976)$	.638	.468
		PEC-TD**	$(1, .988, .833)$	-	-
		PEC-TD*	$(1, .988, .814)$	-	-
		PEC-BU	$(1, 1, 0)$	-	-
12	924	CLF-EC	$(1, 1, .922)$	.644	.468
		PEC-TD**	$(.994, .957, .748)$	-	-
		PEC-TD*	$(.993, .957, .705)$	-	-
		PEC-BU	$(1, 1, 0)$	-	-

* Indicates that the PEC sequence is given as reported in Loeppky et al. (2007), though we were unable to replicate it.

** Indicates our calculation of the PEC sequence, which differs from that given in Loeppky et al. (2007).

computational results for this model space as well, and ours compare favorably to theirs. Any specific comparisons, however, are tenuous at best given the passage of time and difference in platform (our implementation is MATLAB; theirs is Java).

Table 8 addresses the projective model space and it is clear here as well that as the number of models explodes, so does the computation time. (In this table, "r" represents the number of models in $\mathcal{L}$.) Loeppky et al. (2007) do not give computation times, but it is likely their methods are much faster than our procedures, though constraints imposed by their methods limit the quality of their results.

The most serious computational hurdle for our algorithms clearly is the rapid growth of the size of relevant model spaces. Because of this, the methods developed in this work

Table 6: Comparison of designs robust for the projective model space when $n = 28$, k factors, a model space with r elements, and a portion of the projective estimation capacity sequence (p_5, p_6).

k	r	Design	(p_5, p_6)	$\overline{E}_{\mathcal{D}}$	$\min E_{\mathcal{D}}$
8	28	CLF-EC	$(1, 1)$	.592	.447
		CLF-Maximin	$(1, 1)$	.837	.808
		PEC-TD	$(1, 1)$	.807	.678
		PEC-BU	$(1, 1)$	.781	.643
10	210	CLF-EC	$(1, 1)$	.658	.516
		CLF-Maximin	$(1, 1)$	.765	.679
		PEC-TD*	$(1, 1)$	-	-
		PEC-BU*	$(1, 1)$	-	-
12	924	CLF-EC	$(1, 1)$	.697	.447
		CLF-Maximin	$(1, 1)$	.697	.544
		PEC-TD*	$(1, 1)$	-	-
		PEC-BU	$(1, .997)$	-	-

* Indicates that the PEC sequence is given as reported in Loeppky et al. (2007), though we were unable to replicate it. Note that since our calculations showed that the largest element of the PEC sequence was less than 1, we did not include efficiency results.

will be effective for small and medium size problems. Large model sizes stretching into the thousands will severely tax our procedures. At least two measures could be taken to speed up our algorithm as it is currently programmed: (1) use a compiled language—such as C++ or FORTRAN—instead of an interpreted language; (2) parallelize so that each algorithm try is conducted simultaneously on separate processors. Even with these improvements, however, large model spaces will likely be computationally prohibitive. Incidentally, the maximin criterion as implemented does not appear to play a major role in the computation time, compared to a model-averaging approach, since each criterion must deal with the same large model spaces.

Table 7: Assuming the MEPI model space, these are computation times for our procedure for a variety of experiments, where n is the sample size, k is the number of factors, g is the number of possible two-factor interactions, and r is the number of elements in the model space.

n	k	g	r	Minutes/Algorithm Try
12	5	2	45	0.060
12	7	3	1,330	7.750
12	9	2	210	3.998
16	7	2	210	0.566
16	7	3	1,330	3.816
16	7	4	5,985	37.252
16	8	2	378	1.334
16	9	2	630	2.497

6 Discussion

In this article we have presented and demonstrated a candidate list free exchange algorithm that finds model-robust designs for two-level experiments. These designs are constructed with respect to one of two model spaces previously studied in the literature: (1) models containing all main effects and a specified number of two-factor interactions; or (2) models containing a specified number of main effects and all associated two-factor interactions. The proposed methodology, applied to these model spaces, produces designs which often allow more models to be estimated and, via a secondary maximin criterion, provides worst-case efficiency protection for all the models in the set.

Another advantage of the maximin criterion is that prior information, about which effects are likely to be active, can be easily and effectively incorporated. In our experience, averaging criteria such as those employed by Li and Nachtsheim (2000) and Loeppky et al. (2007) are less sensitive to prior information. In addition, we have observed that designs based upon the maximin criterion tend to have less variable $\mathcal{D}$-efficiencies across all possible models than those based on an average efficiency criterion.

We have limited the scope of application of these procedures (i.e. model-robust design algorithms that utilize a maximin criterion on $\mathcal{D}$-efficiencies) to model spaces for which the size of each model is the same, in terms of the number of parameters. If the maximin crite-

Table 8: Assuming the projective model space, these are computation times for our procedure for a variety of experiments, where n is the sample size, k is the number of factors, and r is the number of elements in the sample space.

n	k	r	Minutes/Algorithm Try
16	6	6	0.023
16	10	252	5.246
16	14	2002	88.250
20	6	6	0.036
20	8	56	0.237
20	10	252	3.950
20	12	792	20.192
24	6	1	0.017
24	8	28	0.187
24	10	210	5.714
24	12	924	41.428

rion is used for model sets with models possessing widely varying numbers of parameters, the designs will tend to have similar efficiencies for the large and small models. Since the larger models are more difficult to estimate, as evinced by larger parameter estimate variances, designs with equal efficiencies result in much higher variances for the parameters in the large models than for those in the smaller models. In short, $\mathcal{D}$-efficiency is a poor comparative measure for the effectiveness of a design with respect to models of significantly differing sizes. Since the maximin criterion tends to produce designs with similar $\mathcal{D}$-efficiencies for different models, it amplifies this weakness of $\mathcal{D}$-efficiency. If different-sized models are involved in the model space, an averaging criterion is more appropriate.

One of the principal elements of our algorithms is that they do not restrict the design to be orthogonal or balanced. Furthermore, our approach can trivially be adapted to include more than two levels for each factor, or an arbitrary number of runs. For instance, instead of a minimum aberration (balanced, orthogonal) 16-run, two-level design, one could design an experiment to accommodate, for instance, three levels and 17 runs. This flexibility, while allowing gains in estimability and estimation efficiency, comes at a price. Traditionally stalwart design properties—orthogonality and balance—may be sacrificed by our approach, though perhaps not to a great extent (Li and Nachtsheim 2000, demonstrate that a lack

of orthogonality, for designs that do not enforce this property, does not tend to be severe). Thus, the simple analysis procedures and interpretations allowed by orthogonal designs, and the property that the intercept term is uncorrelated with balanced factors, are lost. The trade-off is one the experimenter should consider: Using our methods results in superior estimation; using Li and Nachtsheim (2000) or Loeppky et al. (2007) gives designs with somewhat less precision while retaining traditional properties.

Besides the relative computational advantages to the $\mathcal{D}$-criterion, there seems to be no reason that other criteria—for instance the average prediction variance ($\mathcal{IV}$) criterion—could not be used instead. In fact, the critical computational bottleneck for this procedure is not determinant-updating, but the size of the model set. In other words, if progress could be made to overcome the problem of exploding model size, algorithms based upon $\mathcal{IV}$- or $\mathcal{A}$- or $\mathcal{G}$-efficiencies would likely become viable.

Supplementary Materials

Supplementary materials are available from the first author which include the designs referred to in this article and the MATLAB code used to construct them.

Acknowledgements

The authors express their thanks to several anonymous referees, of this and another manuscript, that has resulted in significant additional insight into this work. The first author also gratefully acknowledges the Dissertation Fellowship Program of the U.S. Census Bureau, which has funded this work in part.

References

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

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

Box, G. E. P. and Draper, N. R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54(287), 622–654.

Cheng, C.-S., Steinberg, D. M., and Sun, D. X. (1999), "Minimum Aberration and Model Robustness for Two-Level Fractional Factorial Designs," *Journal of the Royal Statistical Society, Series B*, 61(1), 85–93.

Cook, R. D. and Nachtsheim, C. J. (1980), "A Comparison of Algorithms for Constructing Exact D-Optimal Designs," *Technometrics*, 22, 315–324.

— (1982), "An Analysis of the k-exchange Algorithm," Tech. rep., University of Minnesota, Dept. of Operations and Management Science.

Dette, H. (1990), "A generalization of $\mathcal{D}$- and $\mathcal{D}_1$-optimal design in polynomial regression," *Annals of Statistics*, 18, 1784–1804.

Dette, H. and Franke, T. (2000), "Constrained D- and D_1-Optimal Designs for Polynomial Regression," *The Annals of Statistics*, 28(6), 1702–1727.

— (2001), "Robust Designs for Polynomial Regression by Maximizing a Minimum of $\mathcal{D}$- and $\mathcal{D}_1$-Efficiencies," *The Annals of Statistics*, 29(4), 1024–1049.

DuMouchel, W. and Jones, B. (1994), "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model," *Technometrics*, 36(1), 37–47.

Fang, Z. and Wiens, D. P. (2000), "Integer-Valued, Minimax Robust Designs for Estimation and Extrapolation in Heteroscedastic, Approximate Linear Models," *Journal of the American Statistical Association*, 95(451), 807–818.

— (2003), "Robust regression designs for approximate polynomial models," *Journal of Statistical Planning and Inference*, 117, 305–321.

Fedorov, V. V. (1972), *Theory of Optimal Experiments*, Academic Press, New York, NY.

Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M., and Borror, C. M. (2004), "Model-Robust Optimal Designs: A Genetic Algorithm Approach," *Journal of Quality Technology*, 36(3), 263–279.

Imhof, L. and Wong, W. K. (2000), "A Graphical Method for Finding Maximin Efficiency Designs," *Biometrics*, 56(1), 113–117.

Johnson, M. E. and Nachtsheim, C. J. (1983), "Some Guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces," *Technometrics*, 25, 271–277.

Jones, B. and Nachtsheim, C. J. (2011), "Efficient Designs With Minimal Aliasing," *Technometrics*, 53(1), 62–71.

Jones, B. A., Li, W., Nachtsheim, C. J., and Ye, K. Q. (2009), "Model-robust supersaturated and partially supersaturated designs," *Journal of Statistical Planning and Inference*, 139, 45–53.

Läuter, E. (1974), "Experimental design in a class of models," *Mathematische Operationsforschung und Statistik*, 5, 379–398.

Li, W. and Nachtsheim, C. J. (2000), "Model-Robust Factorial Designs," *Technometrics*, 42(4), 345–352.

Loeppky, J. L., Sitter, R. R., and Tang, B. (2007), "Nonregular Designs with Desirable Projection Properties," *Technometrics*, 49(4), 454–467.

Meyer, R. K. and Nachtsheim, C. J. (1995), "The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs," *Technometrics*, 37(1), 60–69.

Mitchell, T. J. (1974), "An Algorithm for the Construction of D-Optimal Experimental Designs," *Technometrics*, 16, 203–210.

Montepiedra, G. and Fedorov, V. V. (1997), "Minimum bias designs with constraints," *Journal of Statistical Planning and Inference*, 63, 97–111.

Pukelsheim, F. and Rosenberger, J. L. (1993), "Experimental Designs for Model Discrimination," *Journal of the American Statistical Association*, 88(422), 642–649.

Smucker, B. J., del Castillo, E., and Rosenberger, J. L. (2011), "Exchange Algorithms for Constructing Model-Robust Experimental Designs," *Journal of Quality Technology*, 43(1).

Stigler, S. M. (1971), "Optimal Experimental Design for Polynomial Regression," *Journal of the American Statistical Association*, 66(334), 311–318.

Studden, W. J. (1982), "Some Robust-Type D-Optimal Designs in Polynomial Regression," *Journal of the American Statistical Association*, 77(380), 916–921.

Thibodeau, L. A. (1977), "Robust design for regression problems," Ph.D. thesis, School of Statistics, University of Minnesota.

Tsai, P.-W. and Gilmour, S. G. (2010), "A General Criterion for Factorial Designs under Model Uncertainty," *Technometrics*, 52(2), 231–242.

Welch, W. J. (1983), "A Mean Squared Error Criterion for the Design of Experiments," *Biometrika*, 70(1), 205–213.

Xu, H., Phoa, F. K. H., and Wong, W. K. (2009), "Recent Developments in Nonregular Fractional Factorial Designs," *Statistical Surveys*, 3, 18–46.