DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

ONDERZOEKSRAPPORT NR 9831

EXPLOITING CORRELATION IN THE CONSTRUCTION OF *D*-OPTIMAL RESPONSE SURFACE DESIGNS

by

P. GOOS

M. VANDEBROEK

Katholieke Universiteit Leuven

Naamsestraat 69, B-3000 Leuven

ONDERZOEKSRAPPORT NR 9831

EXPLOITING CORRELATION IN THE CONSTRUCTION OF *D*-OPTIMAL RESPONSE SURFACE DESIGNS

by

P. GOOS

M. VANDEBROEK

D/1998/2376/31

Exploiting Correlation in the Construction of *D*-optimal Response Surface Designs

P. Goos
M. Vandebroek

Abstract

Cost considerations and difficulties in performing completely randomized experiments often dictate the necessity to run response surface experiments in a bi-randomization format. The resulting compound symmetric error structure not only affects estimation and inference procedures but it also has severe consequences for the optimality of the designs used. For this reason, it should be taken into account explicitly when constructing the design. In this paper, an exchange algorithm for constructing *D*-optimal bi-randomization designs is developed and the resulting designs are analyzed. Finally, the concept of bi-randomization experiments is refined, yielding very efficient designs, which, in many cases, outperform *D*-optimal completely randomized experiments.

1 Introduction

In cases where factor levels are difficult to change or to control, conducting a completely randomized design (CRD) is impractical and can be highly inconvenient and very costly. Typical examples of such factors are pressure, humidity and process temperature. Rather than conducting a CRD in which pressure has to be moved back and forth according to the randomization scheme, executing experimental runs with equal pressure successively will be preferred by the experimenter.

Anderson and McLean [1] describe an experiment from the steel industry in which the effect of temperature and orientation within the furnace on strength of three alloys was investigated. Temperature has four experimental levels, while the factor orientation has only two levels: random and aligned orientation. A single replicate of a CRD would imply $3 \times 4 \times 2 = 24$ independent heatings. However, all experimental runs for which the temperature was at the same level were conducted simultaneously in the same oven, implying only four independent heatings. This experiment suffers from a restricted randomization because assigning a temperature level to an experimental run determines the furnace in which the run will be executed. The

resulting experimental design is a bi-randomization design (BRD), since it possesses two separate randomization procedures. Firstly, the temperature levels are assigned randomly to the larger experimental units, called whole plots. Next, the whole plots are divided into smaller experimental units, called sub-plots. In the second randomization procedure, the levels of the remaining factors alloy type and orientation are assigned at random to the sub-plot experimental units.

The set of experimental variables in a bi-randomization experiment is thus divided in two groups. The n_w design variables that are difficult or costly to change and to control will be denoted by $z_1, z_2, \ldots, z_{n_w}$ or simply by $\mathbf{z}$ and will be referred to as the whole plot variables. The remaining n_s design variables are the sub-plot variables $x_1, x_2, \ldots, x_{n_s}$ or $\mathbf{x}$. In the example above, temperature is the whole plot variable of the experiment, whereas alloy type and orientation are the sub-plot variables. The BRD has two types of experimental units called whole plots and sub-plots and hence two separate randomization procedures. Firstly, each of the w unique combinations of $\mathbf{z}$ is assigned randomly to a whole plot, thereby generating the whole plot error variance. Suppose, for instance, that a BRD has two whole-plot variables with three levels each, then the experiment involves nine whole plots. The second randomization consists of assigning the combinations of $\mathbf{x}$ to the sub-plots, generating the sub-plot error variance. The jth observation within the ith whole plot can be written as

$$\begin{aligned} Y_{ij} &= \beta_0 + \boldsymbol{\gamma}'\mathbf{z}_i + \boldsymbol{\beta}'\mathbf{x}_{ij} + \mathbf{z}_i'\boldsymbol{\Gamma}\mathbf{z}_i + \mathbf{z}_i'\boldsymbol{B}\mathbf{x}_{ij} + \delta_i + \varepsilon_{ij}, \\ &= \mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{ij})\boldsymbol{\tau} + \delta_i + \varepsilon_{ij}, \end{aligned} \tag{1}$$

where $\mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{ij})$ represents the polynomial expansion of the experimental variables and the $p \times 1$ vector $\boldsymbol{\tau}$ contains the p model parameters. The whole plot error δ_i and the sub-plot error ε_{ij} are assumed to be independent and identically normally distributed with zero mean and variance σ_δ^2 and σ_ε^2 respectively. Using this assumption, we derive from (1) that

$$\begin{aligned} \sigma^2(Y_{ij}) &= \sigma_\delta^2 + \sigma_\varepsilon^2, & \\ \sigma(Y_{ij}, Y_{ij'}) &= \sigma_\delta^2, & j \neq j', \\ \sigma(Y_{ij}, Y_{i'j'}) &= 0, & i \neq i'. \end{aligned}$$

Any two observations within the same whole plot are correlated, whereas any two observations from different whole plots are independent. This implies that the variance-covariance matrix of the observations Y_{ij} for any given whole plot has compound symmetry. The variance-covariance matrix of the s_i observations within the ith

whole plot is the $s_i \times s_i$ matrix

$$\begin{aligned}\mathbf{V}_i &= \begin{bmatrix} \sigma_\delta^2 + \sigma_\varepsilon^2 & \sigma_\delta^2 & \cdots & \sigma_\delta^2 \\ \sigma_\delta^2 & \sigma_\delta^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\delta^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\delta^2 & \cdots & \sigma_\delta^2 & \sigma_\delta^2 + \sigma_\varepsilon^2 \end{bmatrix}, \\ &= \sigma_\varepsilon^2 \mathbf{I}_{s_i \times s_i} + \sigma_\delta^2 \mathbf{1}_{s_i \times 1} \mathbf{1}'_{s_i \times 1}, \\ &= \sigma_\varepsilon^2 (\mathbf{I}_{s_i \times s_i} + d \mathbf{1}_{s_i \times 1} \mathbf{1}'_{s_i \times 1}),\end{aligned} \tag{2}$$

with d the variance ratio $\sigma_\delta^2/\sigma_\varepsilon^2$, $\mathbf{I}_{s_i \times s_i}$ the s_i-dimensional identity matrix, and $\mathbf{1}_{s_i \times 1}$ a vector of ones. Since observations from different whole plots are independent, the variance-covariance matrix of all $n = \sum_{i=1}^{w} s_i$ observations is given by the block diagonal matrix

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}_w \end{bmatrix}. \tag{3}$$

1.1 Crossed and non-crossed BRDs

Crossed BRDs or split-plot designs differ from non-crossed BRDs in that every combination of levels of x appears in each whole plot. Each whole plot then has an equal number of sub-plots, namely $s = n/w$, as well as equal levels of $\mathbf{x}$. In non-crossed BRDs, each whole plot may have a different number of sub-plots and the levels of the sub-plot variables need no longer be identical across whole plots. In general, the design matrices $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \ldots, \mathbf{X}'_w)'$ of crossed and non-crossed BRDs can be written as

$$\begin{bmatrix} \mathbf{f}'(\mathbf{z}_1, \mathbf{x}_1) \\ \mathbf{f}'(\mathbf{z}_1, \mathbf{x}_2) \\ \vdots \\ \mathbf{f}'(\mathbf{z}_1, \mathbf{x}_s) \\ \cdots\cdots\cdots \\ \mathbf{f}'(\mathbf{z}_2, \mathbf{x}_1) \\ \mathbf{f}'(\mathbf{z}_2, \mathbf{x}_2) \\ \vdots \\ \mathbf{f}'(\mathbf{z}_2, \mathbf{x}_s) \\ \cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots \\ \mathbf{f}'(\mathbf{z}_w, \mathbf{x}_1) \\ \mathbf{f}'(\mathbf{z}_w, \mathbf{x}_2) \\ \vdots \\ \mathbf{f}'(\mathbf{z}_w, \mathbf{x}_s) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{f}'(\mathbf{z}_1, \mathbf{x}_{11}) \\ \mathbf{f}'(\mathbf{z}_1, \mathbf{x}_{12}) \\ \vdots \\ \mathbf{f}'(\mathbf{z}_1, \mathbf{x}_{1s_1}) \\ \cdots\cdots\cdots\cdots \\ \mathbf{f}'(\mathbf{z}_2, \mathbf{x}_{21}) \\ \mathbf{f}'(\mathbf{z}_2, \mathbf{x}_{22}) \\ \vdots \\ \mathbf{f}'(\mathbf{z}_2, \mathbf{x}_{2s_2}) \\ \cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots \\ \mathbf{f}'(\mathbf{z}_w, \mathbf{x}_{w1}) \\ \mathbf{f}'(\mathbf{z}_w, \mathbf{x}_{w2}) \\ \vdots \\ \mathbf{f}'(\mathbf{z}_w, \mathbf{x}_{ws_w}) \end{bmatrix}$$

respectively. Designs which fall within the category of crossed BRDs are the two- and three-level factorial designs. Included within the category of non-crossed BRDs are some two-level fractional factorial designs, the central composite design (CCD) and Box-Behnken designs.

1.2 Estimation and inference

The error structure of the BRD plays an important role in model estimation and editing. Under the assumption of normal errors, the maximum likelihood estimator of the unknown model parameters $\boldsymbol{\tau}$ is given by the generalized least squares (GLS) estimation equation

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{4}$$

The variance-covariance matrix of the estimators is

$$\mathrm{Var}(\hat{\boldsymbol{\tau}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{5}$$

However, (4) and (5) cannot be used directly since the variances σ_δ^2 and σ_ε^2 of the whole plot and sub-plot errors are not known. For crossed BRDs and some specific first order non-crossed BRDs with interactions, ordinary and generalized least squares prove to be equivalent, i.e.

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y},$$

implying that, in these cases, error variance knowledge is no longer necessary for model estimation purposes. For model editing, as well as for both model estimation and editing with most first order and with second order non-crossed designs, knowledge of the whole plot and sub-plot error variance remains essential. Approximate t-tests can be formed by dividing the estimated coefficients by the estimated standard errors, which are obtained by substituting the estimated error variances into equation (5). Error variance estimates are thoroughly described by Letsinger, Myers and Lentner [9]. They also recommend restricted maximum likelihood (REML) for error variance and model estimation because of its robustness across various values of d and because it is also a good estimation option when smaller designs and near full second order models are used. REML estimates for the variance components σ_ε^2 and σ_δ^2, and from them $\hat{\boldsymbol{\tau}}$, can be calculated using the MIXED procedure in SAS. For a comprehensive guide on the MIXED procedure, see Littell et al. [10].

The risks of improper analysis of BRDs are pointed out by Box and Jones [4] and by Davison [5], who extend the results of Kempthorne [7]. By using a BRD, a loss of precision in estimation of whole plot coefficients is incurred, while the opposite is true for the sub-plot coefficients and the whole plot by sub-plot interactions. Analysis of a split plot design as a CRD can therefore lead to erroneously considering whole plot effects as significant and sub-plot effects as insignificant.

1.3 Efficiency of designs

Typically, a standard design or a design generated by a statistical package is used to perform a bi-randomization experiment. However, these designs were developed to be applied in completely randomized experiments and do not take into account the bi-randomization error structure. Therefore, they are less efficient than designs specifically constructed for bi-randomization experiments.

By far the most frequently used efficiency criterion to evaluate designs is the D-optimality criterion. This criterion is a direct function of $\text{Var}(\hat{\tau})$. For a CRD the D-criterion value is given by $|\mathbf{X}'\mathbf{X}|/\sigma_\varepsilon^2$. Since σ_ε^2 is a constant, it does not affect the efficiency of a design. When an experiment is conducted under a bi-randomization structure, the D-criterion value $|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$ depends on σ_ε^2 and σ_δ^2 through $\mathbf{V}$. This result suggests that D-optimal CRDs are no longer D-optimal as a bi-randomization experiment. In fact, designing a completely randomized experiment only consists of determining the design points. Designing a bi-randomization experiment simultaneously involves choosing the number of whole plots and the number of sub-plots within each whole plot, i.e. determining the structure of the variance-covariance matrix $\mathbf{V}$. Neglecting the bi-randomization structure of a response surface experiment may therefore lead to poor designs and consequent analysis. The dependence of design efficiency on the bi-randomization structure was pointed out by Letsinger, Myers and Lentner [9], who compare D- and Q-efficiencies of commonly used first and second order response surface designs under various variance ratios. Davison [5] also proves that, while Q-optimality is not affected for first order crossed BRDs only, D-optimality is not influenced for crossed BRDs of any order.

In the next section, we will develop an exchange algorithm to construct D-optimal BRDs. In Section 3, a number of D-optimal BRDs is investigated and compared to the classic D-optimal CRDs. The concept of bi-randomization designs is then refined in Section 4, yielding designs which usually are more efficient than a completely randomized experiment.

2 Constructing D-optimal BRDs

A vast literature on the construction of exact D-optimal CRDs can be found. Early approaches involved direct maximization of $|\mathbf{X}'\mathbf{X}|$ by mathematical programming techniques. More recently, attention has focused on a number of methods which take into account some of the special characteristics of the design problem. For large numbers of observations, Kiefer [8] suggested rounding off the approximate D-optimal design to obtain the exact design. Welch [11] used branch-and-bound to find optimal designs. Most of the remaining approaches however may be classified as exchange algorithms.

In the first part of this section, we will describe the main features of exchange algorithms developed to construct D-optimal CRDs. Next, we will show how a similar approach can be adopted in the construction of D-optimal BRDs.

2.1 Exchange algorithms for CRDs

Exchange algorithms typically begin with a non-singular n-point starting design and then add and delete one or more observations in order to achieve increases in the determinant $|\mathbf{X}'\mathbf{X}|$. Design points are chosen from a predefined set of candidate or support points which cover the entire design region. The first well-known exchange algorithm was developed by Fedorov [6]. Atkinson and Donev [2] have adapted Fedorov's algorithm in order to speed it up. Key property of the information matrix of a completely randomized design is that it can be written as a sum of outer products of the polynomial expansion of its design points $\mathbf{x}_i$:

$$\mathbf{X}'\mathbf{X} = \sum_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}'(\mathbf{x}_i). \tag{6}$$

This property enables the experimenter to evaluate addition and/or deletion of design points at a low computational cost. In fact, adding a new observation to the experiment adds another outer product to the information matrix:

$$\mathbf{X}'_{new}\mathbf{X}_{new} = \sum_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}'(\mathbf{x}_i) + \mathbf{f}(\mathbf{x}_{new})\mathbf{f}'(\mathbf{x}_{new}).$$

The D-criterion value and the inverse of the information matrix can then be updated as follows:

$$|\mathbf{X}'_{new}\mathbf{X}_{new}| = |\mathbf{X}'\mathbf{X}|\{1 + \mathbf{f}'(\mathbf{x}_{new})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_{new})\},$$

$$(\mathbf{X}'_{new}\mathbf{X}_{new})^{-1} = (\mathbf{X}'\mathbf{X})^{-1} - \frac{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_{new})\mathbf{f}'(\mathbf{x}_{new})(\mathbf{X}'\mathbf{X})^{-1}}{1 + \mathbf{f}'(\mathbf{x}_{new})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_{new})}.$$

These equations are especially important for the construction of a starting design, which is improved through iteratively exchanging design and support points. To evaluate an exchange of a design point $\mathbf{x}_d$ and a support point $\mathbf{x}_s$ the exchange algorithms of Fedorov and Atkinson and Donev use the following formula:

$$\begin{aligned} |\mathbf{X}'_{new}\mathbf{X}_{new}| = |\mathbf{X}'\mathbf{X}|\Big\{&[1 + \mathbf{f}'(\mathbf{x}_s)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_s)][1 - \mathbf{f}'(\mathbf{x}_d)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_d)] \\ &+ [\mathbf{f}'(\mathbf{x}_d)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_s)]^2\Big\}. \end{aligned} \tag{7}$$

In order to speed up the algorithm, the BLKL algorithm of Atkinson and Donev does not evaluate all possible exchanges. Instead, it only considers the L support points with the highest prediction variance $\mathbf{f}'(\mathbf{x}_s)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_s)$ as candidates to enter the design and the K design points with the lowest prediction variance

$\mathbf{f}'(\mathbf{x}_d)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_d)$ as candidates to be removed from the design. From (7) can be seen that it is probable that the best possible exchange is among those considered. Small values of $[\mathbf{f}'(\mathbf{x}_d)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_s)]^2$ for the exchanges considered might disturb this approach. The algorithm is stopped when no favorable exchange can be found any more. In order to avoid being stuck in a local optimum, more than one starting design should be generated and improved. To generate starting designs, the algorithm randomly chooses a number of points from the set of support points and then adds the support points that yield the highest increase of the D-criterion value until an n-point design has been constructed.

The exchange algorithms can be easily adapted to develop D-optimal designs in the presence of variance heterogeneity because the variance-covariance matrix $\mathbf{V}$ of the observations is still diagonal. In that case, the information matrix can be expressed as a weighted sum of outer products:

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \sum_i \frac{\mathbf{f}(\mathbf{x}_i)\mathbf{f}'(\mathbf{x}_i)}{\sigma_i^2}. \tag{8}$$

2.2 An exchange algorithm for BRDs

From the outline of the exchange algorithms for the construction of CRDs emerges that it is crucial to write the information matrix as a sum of outer products in order to use the attractive update formulae. Since the variance-covariance matrix of the observations in a bi-randomization experiment is block diagonal instead of diagonal as in the homoscedastic or heteroscedastic CRD case, the information matrix of a BRD can not be written as in equations (6) and (8). However, we have from equation (2) that

$$\mathbf{V}_i = \sigma_\varepsilon^2(\mathbf{I}_{s_i\times s_i} + d\mathbf{1}_{s_i\times 1}\mathbf{1}'_{s_i\times 1}),$$

and therefore

$$\mathbf{V}_i^{-1} = \frac{1}{\sigma_\varepsilon^2}(\mathbf{I}_{s_i\times s_i} - \frac{d}{1+s_i d}\mathbf{1}_{s_i\times 1}\mathbf{1}'_{s_i\times 1}). \tag{9}$$

From (3), we have that

$$\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{V}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}_w^{-1} \end{bmatrix}.$$

Hence

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \sum_{i=1}^{w} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i. \tag{10}$$

Combining (9) and (10), we obtain

$$\begin{aligned}
\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} &= \sum_{i=1}^{w} \mathbf{X}_i'\mathbf{V}_i^{-1}\mathbf{X}_i, \\
&= \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{w} \mathbf{X}_i'(\mathbf{I}_{s_i\times s_i} - \frac{d}{1+s_i d}\mathbf{1}_{s_i\times 1}\mathbf{1}_{s_i\times 1}')\mathbf{X}_i, \\
&= \frac{1}{\sigma_\varepsilon^2}\sum_{i=1}^{w}\Big(\mathbf{X}_i'\mathbf{X}_i - \frac{d}{1+s_i d}(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})'\Big), \qquad (11) \\
&= \frac{1}{\sigma_\varepsilon^2}\Big(\sum_{i=1}^{w}\sum_{j=1}^{s_i}\mathbf{f}(\mathbf{z}_i,\mathbf{x}_{ij})\mathbf{f}'(\mathbf{z}_i,\mathbf{x}_{ij}) - \sum_{i=1}^{w}\frac{d}{1+s_i d}(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})'\Big).
\end{aligned}$$

It should be clear that this expression generalizes equation (6). In fact, a CRD has zero variance ratio ($d = 0$) and has no whole plot variables $\mathbf{z}$. However, in the BRD case, the w weighted outer products $d/(1 + s_i d)(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})'$ are subtracted from the sum of the outer products of the design points' polynomial expansions. $(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})$ is a p-dimensional vector containing the sums of the columns of $\mathbf{X}_i$.

Since it expresses the information matrix as a sum and difference of outer products, equation (11) now allows us to evaluate the effect of changes in the design of the experiment on the information matrix. Adding an observation to the ith whole plot will add an extra outer product to the information matrix, but it will also affect the part of the design matrix that corresponds to the ith whole plot, that is $\mathbf{X}_i$. The latter will affect the information matrix in two ways. Firstly, $\mathbf{X}_i'\mathbf{1}_{s_i\times 1}$ will change and, secondly, the number of sub-plots within the ith whole plot s_i will increase by one. Let $\mathbf{X}^*$, $\mathbf{V}^*$ and $\mathbf{X}_i^{*\prime}$ denote the new design matrix, the corresponding variance-covariance matrix and the part of the new design matrix corresponding to the ith whole plot respectively. The relationship between the information matrix before and after adding the observation is given by

$$\begin{aligned}
\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^{*\prime} &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \\
&\quad + \mathbf{f}(\mathbf{z}_i,\mathbf{x}_{(i,s_i+1)})\mathbf{f}'(\mathbf{z}_i,\mathbf{x}_{(i,s_i+1)}) \\
&\quad + \frac{d}{1+s_i d}(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})' \qquad (12) \\
&\quad - \frac{d}{1+(s_i+1)d}(\mathbf{X}_i^{*\prime}\mathbf{1}_{(s_i+1)\times 1})(\mathbf{X}_i^{*\prime}\mathbf{1}_{(s_i+1)\times 1})'.
\end{aligned}$$

Adding an observation in a new whole plot is a special case of (12) where $s_i = 0$, $\mathbf{X}_i = \mathbf{0}$ and $\mathbf{X}_i^* = \mathbf{f}'(\mathbf{z}_i,\mathbf{x}_{i1})$. Equation (12) then simplifies to

$$\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^{*\prime} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} + \frac{1}{1+d}\mathbf{f}(\mathbf{z}_{w+1},\mathbf{x}_{(w+1,1)})\mathbf{f}'(\mathbf{z}_{w+1},\mathbf{x}_{(w+1,1)}). \qquad (13)$$

The update formulae for the deletion of an observation are given in Appendix A and can be obtained by analogous reasoning. From them, and from equations (12) and (13), we may conclude the information matrix of a BRD is modified by adding and/or subtracting weighted outer products when design points are added to or deleted from the design. Therefore, the formulae from Section 2.1 can be used to update the determinant and the inverse of the information matrix. This enables us to evaluate the effect of adding and deleting design points on the D-criterion value of the design at a low computational cost.

We used these results to develop an exchange algorithm for the construction of D-optimal BRDs. The structure of the algorithm is completely analogous to the exchange algorithms for CRDs. However, the complexity of the update formulae and hence, of the construction of the starting design and of the evaluation of the exchange of design and support points has increased considerably. Moreover, in contrast with the BLKL exchange algorithm of Atkinson and Donev [2], all possible exchanges are evaluated because expressions similar to $[\mathbf{f}'(\mathbf{x}_d)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}_s)]^2$ in equation (7) are now exceedingly present in the exchange formula. For this reason, the prediction variance is no longer a good indicator of the desirability to add a candidate point to or to remove a design point from the design. Practical restrictions, e.g. on the number of whole plots or on the number of sub-plots within each whole plot, can be easily taken into account. The design construction algorithm also finds application in design augmentation problems.

3 D-optimal BRDs

Using our exchange algorithm, we have constructed D-optimal BRDs for models with different numbers of variables, different numbers of whole plot factors and sub-plot variables under various variance ratios. The D-optimal BRDs will be compared to properly and improperly conducted CRDs with respect to D- and A-efficiency. It is also examined to what extent misspecification of the variance ratio deteriorates the design efficiency.

3.1 Geometric features of D-optimal BRDs

The geometric properties of the D-optimal BRDs will be illustrated by means of two three-variable bi-randomization design problems. D-optimal BRDs for one and for two whole plot variables will be constructed. In all examples, design points were chosen from the 3^k factorial design, with k the number of variables of interest.

The design matrices of D-optimal BRDs and CRDs for first order models with and without interactions are identical. The only difference between both lies in the fact

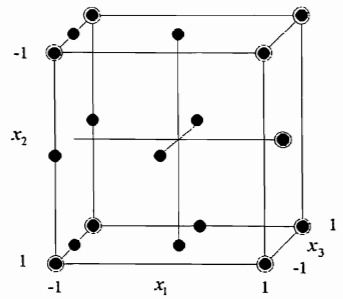
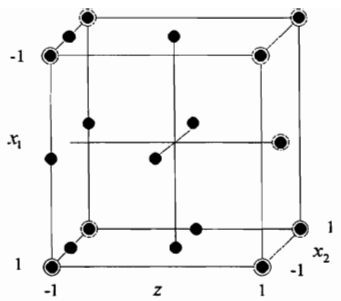


Figure 1: D-optimal 27-point CRD for the full quadratic model in three variables (• is a design point, ○ is a replicated design point).

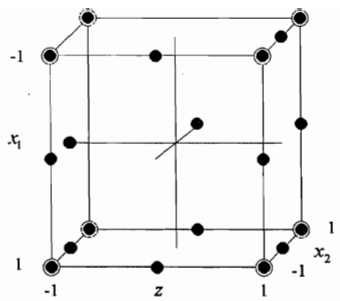
that, in a bi-randomization experiment, the assignment of the whole plot and sub-plot variables to the columns of the design matrix matters. For second order models, the design matrices of D-optimal BRDs and CRDs typically differ.

Consider the full quadratic model in three variables. The D-optimal 27-point CRD is displayed in Figure 1, in which the three variables of interest are denoted by x_1, x_2 and x_3. The D-optimal CRD can be computed by many statistical packages, by the BLKL algorithm of Atkinson and Donev [3] and by specifying $d = 0$ in our design construction algorithm. This design can be properly used in a completely randomized experiment (PCRD) or it can be improperly used in a bi-randomization experiment (ICRD). In the former case, all experimental runs are statistically independent and the CRD, which maximizes $|\mathbf{X}'\mathbf{X}|$, is the optimal design. In the latter case, all runs within the same whole plot are correlated and the optimal design should maximize $|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$. If a bi-randomization experiment with one whole plot variable is preferred, the optimal design is found in Figure 2. BRD1, BRD2, BRD3 and BRD4 are D-optimal for $d \leq 0.3959$, $0.3959 \leq d \leq 0.4727$, $0.4727 \leq d \leq 5.7306$, and $d \geq 5.7306$ respectively. In the geometric representations, the whole plot variable is denoted by z, while the two sub-plot variables are denoted by x_1 and x_2. It turns out that, for $d \leq 0.3959$, the CRD from Figure 1 is optimal on the condition that the whole plot variable is assigned to the horizontal axis. For larger d, the design matrices of the CRD and the D-optimal BRDs differ. From Figure 2, we see that the number of observations at $z = 0$ decreases as the variance ratio increases. The distribution of the observations across the sub-plot levels strongly resembles that of the CRD. Computational results for full quadratic models with one whole plot variable indicate that these conclusions remain valid if the number of sub-plot variables differs from two.

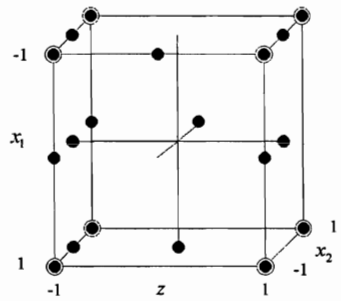
If two whole plot variables are used instead of one, the D-optimal designs look totally different. 27-point D-optimal BRDs for small, moderate and large d for this design problem are shown in Figure 3. The two whole plot variables and the sub-plot variable are denoted by z_1, z_2 and x respectively. Compared with the CRD in Figure 1,



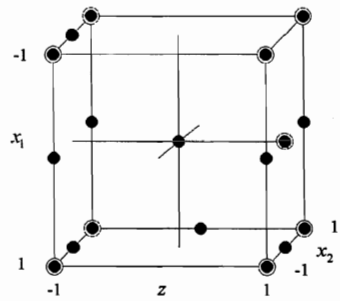
(a) BRD1



(b) BRD2

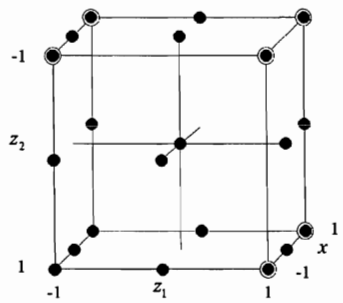


(c) BRD3

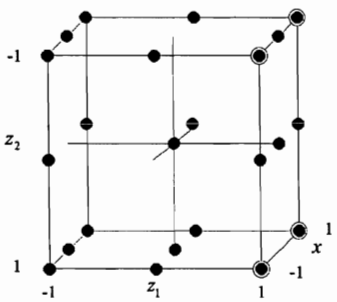


(d) BRD4

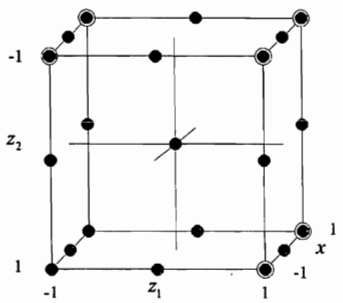
Figure 2: *D*-optimal 27-point designs for a full quadratic model in one whole plot variable and two sub-plot variables (• is a design point, o is a replicated design point).



(a) BRD5



(b) BRD6



(c) BRD7

Figure 3: *D*-optimal 27-point BRDs for a full quadratic model in two whole plot variables and one sub-plot variable (• is a design point, o is a replicated design point).

the BRDs have less observations in the cornerpoints of the design region and they all have an observation in the center point. The BRDs only differ in the location of one or two design points. However, no clear pattern in the designs can be observed.

This section has illustrated that D-optimal designs usually differ from the D-optimal CRD for second order models. Both are equal for first order models with and without interactions, as well as for second order models in one whole plot variable when the variance ratio is small. Nonetheless, even then the CRDs should be used with care since the assignment of variables matters. The next section computes to what extent D- and A-efficiency of experiments can be improved by taking into account the bi-randomization error structure in the design construction stage. Since the variance ratio is generally unknown, it is important to know how a given BRD behaves under different variance ratios. Therefore, we have also investigated the effect of variance ratio misspecification on the design efficiency.

3.2 D- and A-efficiency

In this section, we will compare the D- and A-efficiency of the designs from Figures 1, 2 and 3 by means of the determinant and the trace of their information matrix respectively. The D-optimal BRD maximizes $|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$ for a given design problem, while the A-optimal BRD minimizes $\text{tr}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})$. Unlike D-optimality, the concept of A-optimality does not take into account the covariances among estimates, but only their variances. Unless stated otherwise, reported results are relative to the D-criterion value obtained by ICRD. In this way, the improvement, generated by using D-optimal BRDs instead of the D-optimal CRD in a bi-randomization experiment, can readily be displayed. For each design considered, we have computed the D- and A-criterion values for values of d between 0 and 10, holding the variance $\sigma_\varepsilon^2 + \sigma_\delta^2$ equal to one. This allows us to assess the impact of misspecification of the variance ratio on the efficiency of the designs generated.

Figures 4 and 5 show the D-efficiencies of the BRDs for the examples from Section 3.1. Both figures contain the relative efficiency of the 3^3 factorial as well. The horizontal reference in each figure displays the efficiency of the ICRD, which would be obtained by ignoring the structure of a bi-randomization experiment in the design construction stage. The relative D-efficiencies of the 27-point BRDs for a full quadratic model in one whole plot variable and two sub-plot variables are displayed in Figure 4. It shows that BRD1, BRD2, BRD3 and BRD4 are D-optimal for $d \leq 0.3959$, $0.3959 \leq d \leq 0.4727$, $0.4727 \leq d \leq 5.7306$, and $d \geq 5.7306$ respectively. It turns out that BRD1 is more efficient than ICRD for any strictly positive d. Both designs are equivalent at $d = 0$. BRD2, BRD3 and BRD4 are inferior to both ICRD and BRD1 for small d, but become substantially better as the variance ratio increases. From the figure can also be seen that the efficiency gain is robust against misspecification of the variance ratio, except for small variance ratios. Slightly mis-

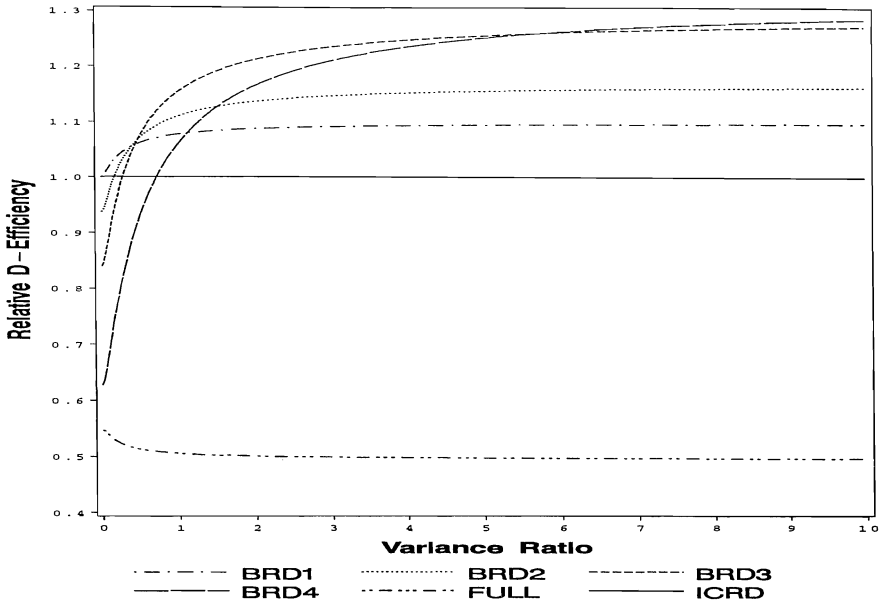


Figure 4: Relative D-efficiencies of the 27-point CRD and BRDs from Figures 1 and 2, and the 3^3 factorial for the full quadratic model in one whole plot variable and two sub-plot variables under different variance ratios.

specifying d typically results in the same design, and thus in the same efficiency. For instance, specifying any d from the interval $]0.4727, 5.7306[$ will lead to BRD3. Severe misspecification can lead to a different design, but unless a variance ratio close to zero is overestimated, the resulting BRD will be more efficient than ICRD. The 3^3 factorial turns out to be a poor alternative to the BRDs and even to ICRD.

Figure 5 shows a similar picture for the BRDs with two whole plot variables. At $d = 0$, ICRD is D-optimal, but for strictly positive d, it is overtaken by the BRDs. Like in the one whole plot variable case, the efficiency gain is robust to variance ratio misspecification. BRD7 is optimal for moderate and large d and highly robust against misspecification. Only overestimating a very small variance ratio might result in a BRD which is worse than ICRD. The 3^3 factorial remains a bad alternative to the BRDs, but has become better than ICRD at larger values of d.

In Figure 6, the D-efficiency of proper use of the CRD (PCRD) is compared to improper use in a bi-randomization experiment with one and with two whole plots (ICRDW1 and ICRDW2 respectively). The D-efficiency of the best BRD with one and two whole plots (BRDW1 and BRDW2 respectively) is displayed as well. D-efficiencies are relative to the D-criterion value of PCRD. For small variance ratios, conducting a PCRD is more efficient than a bi-randomization experiment. However, as d rises, the opposite is true. When d exceeds unity, conducting a BRD with one whole plot variable becomes more efficient than PCRD. When d grows larger than 3,

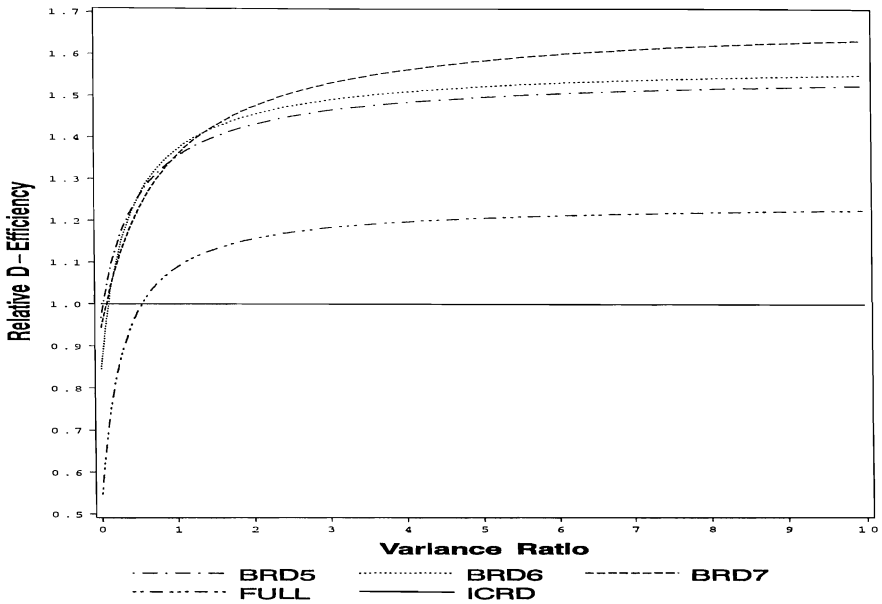


Figure 5: Relative D-efficiencies of the 27-point CRD and BRDs from Figures 1 and 3, and the 3^3 factorial for the full quadratic model in two whole plot variables and one sub-plot variable under different variance ratios.

the same goes for BRDs with two whole plot variables. The efficiency of the ICRDs closely follows that of the corresponding BRDs. Except for small variance ratios, conducting a BRD with one whole plot variable is clearly much more efficient than using two whole plot variables. This is due to the fact that a BRD with one whole plot variable has more nonzero off-diagonal elements in its variance-covariance matrix and therefore, it benefits more quickly from the positive effect of larger correlations. These results illustrate that inducing correlation between observations may have a beneficial effect on D-optimality. For saturated designs, correlated observations always lead to better D-criterion values. The proof is given in Appendix B.

Figures 7 and 8 show the relative A-efficiencies of the D-optimal BRDs and the factorial design. From both figures can be seen that the 3^3 factorial outperforms ICRD and all BRDs. Figure 7 displays the superiority of ICRD and BRD1 over BRD2, BRD3 and BRD4 for one whole plot variable when the variance ratio is small. However, when two whole plot variables are used, ICRD has a lower A-efficiency than BRD5 and BRD6, but is more efficient than BRD7 when d is small. This is shown in Figure 8. However, in both the one and two whole plot variable case, the A-efficiency of the BRDs is smaller than that obtained by properly conducting a CRD.

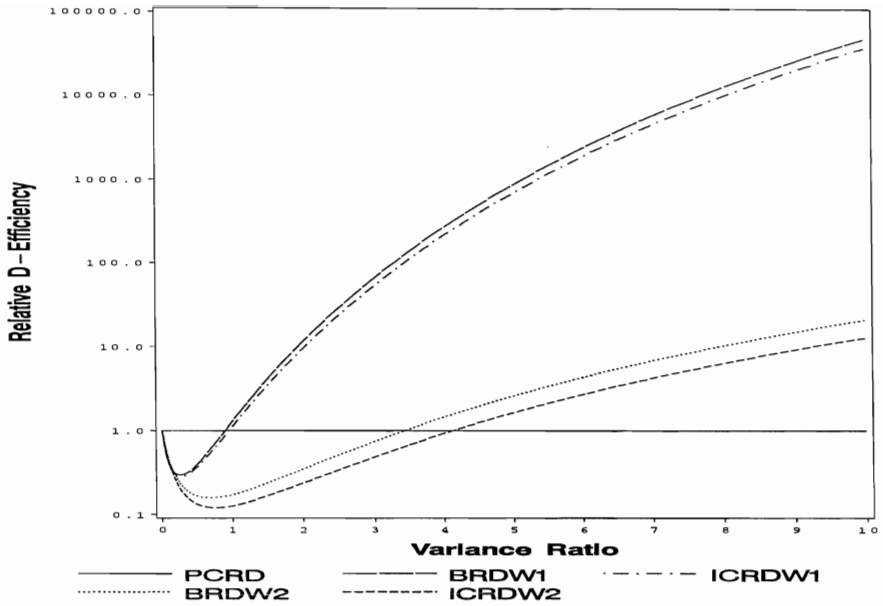


Figure 6: Comparison of the D-efficiency of the BRDs from Section 3.1 and proper and improper use of the CRD for the full quadratic model in three variables.

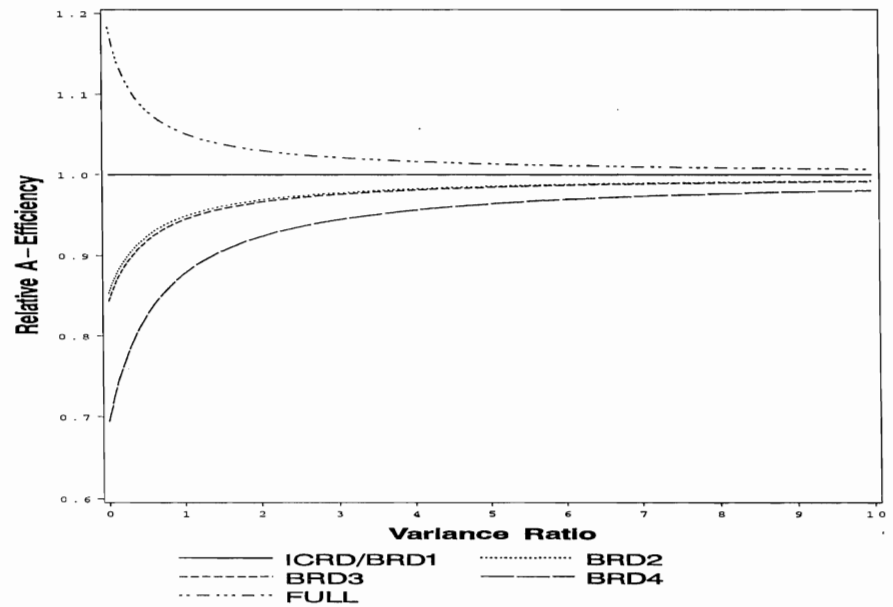


Figure 7: Relative A-efficiencies of the 27-point CRD and BRDs from Figures 1 and 2, and the 3^3 factorial for the full quadratic model in one whole plot variable and two sub-plot variables under different variance ratios.

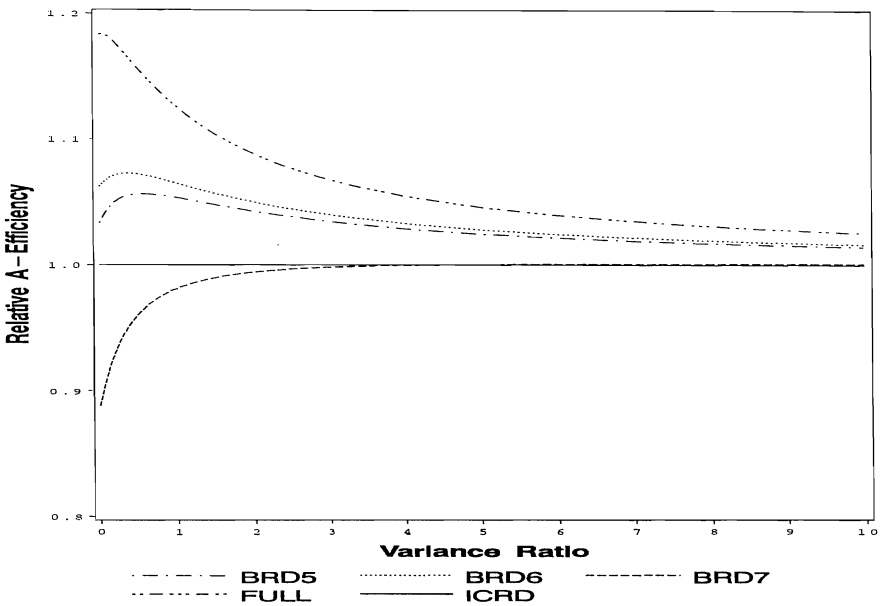


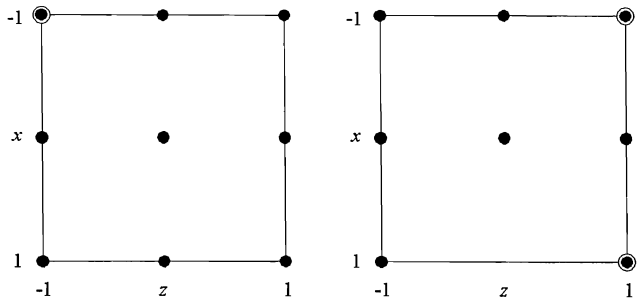
Figure 8: Relative A-efficiencies of the 27-point CRD and BRDs from Figures 1 and 3, and the 3^3 factorial for the full quadratic model in two whole plot variables and one sub-plot variable under different variance ratios.

3.3 21×21 grid

So far, we have only considered the points of the 3^k factorial design as candidate design points. As a result, each whole plot variable had at most three levels. However, we have also computed D-optimal BRDs using a finer grid on the experimental region. The resulting designs possess larger numbers of whole plots, but nevertheless, they strongly resemble the designs obtained by using the candidate points of the full factorial. Besides -1, 0 and +1, the design construction algorithm also chooses ± 0.1 and ± 0.9 as whole plots levels. The optimal sub-plot levels remain at 0 and ± 1. This result suggests that the optimal whole plot levels are indeed at the 0 and ± 1 levels, but that not all observations at a certain whole plot level should be put in the same whole plot. Instead, it strongly recommends the use of more whole plots in order to decrease the number of correlations among observations. In the next section, we relax one of the model assumptions in order to investigate this conjecture and attempt to construct better BRDs.

4 More efficient BRDs

In this section, we will show how the D-optimality of experiments can be further improved by refining the concept of bi-randomization. So far, we have assumed that there is a one to one relation between the combinations of whole plot factor levels and the whole plots. Now, we relax this assumption. Like before, observations with-



(a) CRD/BRD8 (b) BRD9

Figure 9: D-optimal BRDs for a full quadratic model in one whole plot and one sub-plot variable (• is a design point, o is a replicated design point).

in one whole plot all have the same whole plot factor level combination. However, a given whole plot factor level combination can occur in more than one whole plot. This relaxation gives rise to substantially better bi-randomization experiments, both in terms of D- and A-efficiency. Moreover, it turns out that in many instances conducting a bi-randomization experiment is statistically more efficient than properly conducting a completely randomized one. Since, in addition, bi-randomization experiments can be carried out more easily and are typically far less costly, they are economically efficient as well.

Using our design construction algorithm, we have computed D-optimal BRDs under the new assumption. In order to avoid confusion with the BRDs from Section 3, we will call them refined bi-randomization designs (RBRDs). Firstly, we examine their features and compare them with D-optimal BRDs for the same model. Next, D- and A-efficiencies are compared and it is shown that RBRDs are much more efficient than PCRDs and BRDs. Since the whole plot levels no longer solely determine the whole plot in which an observation is run, only part of the information about the experimental setup is contained within the geometric representation of a RBRD. Instead, the design matrix is used. Whole plots will be separated by a dotted line.

4.1 Features

It should be clear to the reader that one of the possible outcomes of our design construction algorithm for RBRDs is a completely randomized experiment. In many cases however, RBRDs with correlated observations possess a better D-criterion value. If one factor serves as a whole plot factor, it is D-optimal to drop PCRDs in favour of RBRDs, which have a restricted randomization. Consider a 10-point design for a full quadratic model in one whole plot and one sub-plot variable. For $d \leq 0.7011$, $0.7011 \leq d \leq 0.9113$ and $d \geq 0.9113$, the D-optimal RBRDs are given

by

$$\begin{bmatrix} -1 & -1 \\ -1 & +1 \\ \cdots\cdots & \\ -1 & 0 \\ \cdots\cdots & \\ -1 & +1 \\ \cdots\cdots & \\ 0 & -1 \\ \cdots\cdots & \\ 0 & 0 \\ \cdots\cdots & \\ 0 & +1 \\ \cdots\cdots & \\ +1 & 0 \\ \cdots\cdots & \\ +1 & -1 \\ +1 & +1 \end{bmatrix}, \quad \begin{bmatrix} -1 & -1 \\ -1 & +1 \\ \cdots\cdots & \\ -1 & 0 \\ -1 & +1 \\ \cdots\cdots & \\ 0 & -1 \\ \cdots\cdots & \\ 0 & 0 \\ \cdots\cdots & \\ 0 & +1 \\ \cdots\cdots & \\ +1 & 0 \\ \cdots\cdots & \\ +1 & -1 \\ +1 & +1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -1 & -1 \\ -1 & +1 \\ \cdots\cdots & \\ -1 & 0 \\ \cdots\cdots & \\ 0 & 0 \\ \cdots\cdots & \\ 0 & +1 \\ \cdots\cdots & \\ +1 & -1 \\ +1 & +1 \\ \cdots\cdots & \\ +1 & -1 \\ +1 & 0 \\ +1 & +1 \end{bmatrix} \tag{14}$$

respectively. In the sequel of this paper we will refer to these three designs as RBRD1, RBRD2 and RBRD3 respectively. The D-optimal CRD and the D-optimal BRDs for the same model are displayed in Figure 9. For $d \leq 2.8026$, the D-optimal coincides with the D-optimal CRD. In this example, the RBRDs possess more whole plots than the BRDs. Since both types of design contain the same design points, this is the only difference between them.

In general, the RBRDs have a couple of striking features. Firstly, the number of whole plots decreases as the variance ratio increases. The 10-point designs shown in (14) have eight, seven and six whole plots. Apparently, the higher the correlation between observations within the same whole plot, the better it is to group more experimental runs and thereby to induce more correlated observations. Otherwise, the lower the correlation, the more the optimal design will tend to a completely randomized experiment. Secondly, observations at the zero levels of the whole plot variables are assigned to separate whole plots, such that whole plots containing more than one observation only occur at $\mathbf{z} = \pm\mathbf{1}$. These properties are exhibited in (14). Both results apply for design problems with one and two whole plot variables.

When two whole plot variables are used instead of one, using a PCRD is D-optimal for small d, in which case the number of whole plots in the experiment equals the number of observations. However, as the variance ratio rises, D-optimal RBRDs group observations within the same whole plots, thereby increasing the number of correlated observations and decreasing the number of whole plots in the experiment.

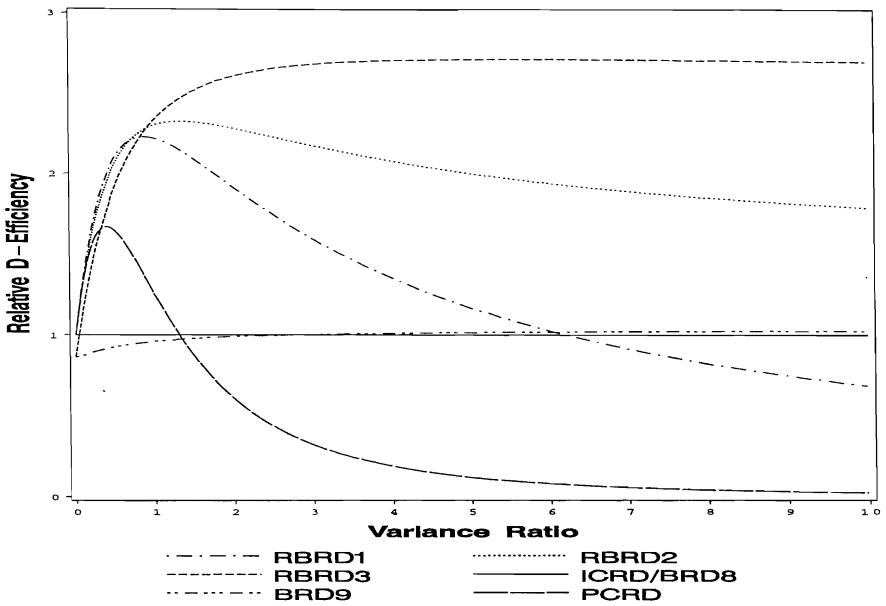


Figure 10: Relative D-efficiencies of the 10-point RBRDs and BRDs from Section 4.1 for the full quadratic model in one whole plot and one sub-plot variable under different variance ratios.

4.2 D- and A-efficiencies

In this section, the alternative experimental setups for the full quadratic model in one whole plot variable and one sub-plot variable are compared as to their D- and A-efficiency. It is also examined to what extent the 27-point designs from Section 3.1 can be improved by refining the concept of bi-randomization experiment.

Figures 10 and 11 display the relative D- and A-efficiencies of the RBRDs and the BRDs from Section 4.1, as well as the efficiency obtained by conducting a PCRD. Again, the variance $\sigma_\varepsilon^2 + \sigma_\delta^2$ is held fixed to one. Figure 10 shows that for small variance ratios, RBRD1 yields a better D-criterion value than PCRD, while RBRD2 and RBRD3 give a worse. However, both RBRD2 and RBRD3 become more efficient than PCRD and RBRD1 as d grows. While RBRD1 is inferior to ICRD/BRD8 and BRD9 for large values of d, RBRD2 and RBRD3 remain superior to both alternatives. It is clear that a substantial gain in D-efficiency can be realized by using RBRDs instead of a PCRD and BRDs. As can be seen from Figure 11, the same conclusion goes for A-efficiency. Slight or moderate misspecification of the variance ratio has no substantial consequences for the efficiency of the RBRDs.

The results for the one and two whole plot variable 27-point RBRDs show a similar picture. Figure 12 extends Figure 6 and compares proper and improper use of the CRD (PCRD and ICRD respectively) with the best BRDs and RBRDs for $0 \leq d \leq 10$ for both a full quadratic model in one and in two whole plot variables.

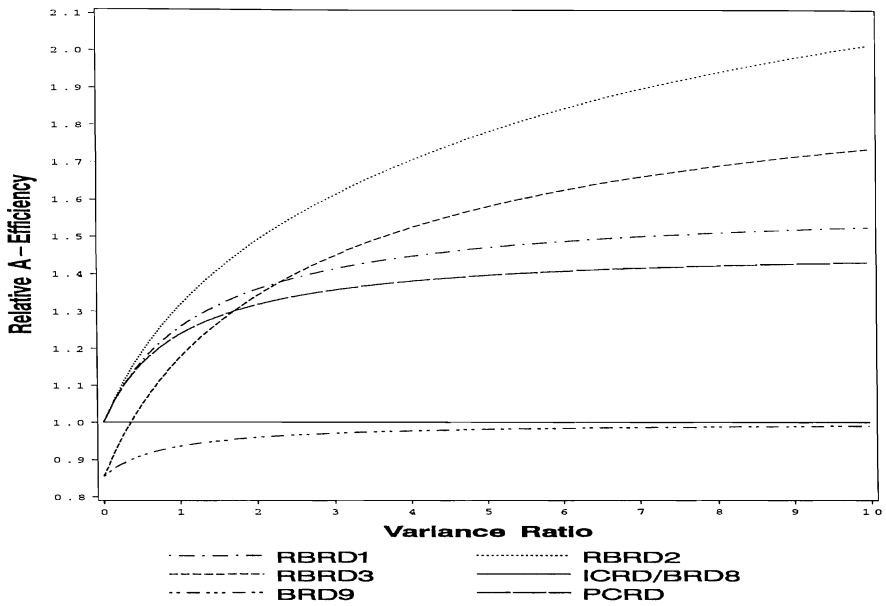


Figure 11: Relative A-efficiencies of the 10-point RBRDs and BRDs from Section 4.1 for the full quadratic model in one whole plot and one sub-plot variable under different variance ratios.

Designs used in a bi-randomization experiment with one (two) whole plot variable(s) are denoted by the extension W1 (W2). Figure 12 illustrates the improvement in D-efficiency that can be realized by using RBRDs instead of alternative experimental setups. For both one and two whole plot variables, the RBRD is much more efficient than the corresponding BRD. In contrast with the BRDs, the RBRDs are more efficient than PCRD at any value of the variance ratio.

From these examples can be seen that experiments can be made more efficient by using RBRDs instead of PCRDs. Firstly, the ease by which the experiment is conducted is increased since D-optimal RBRDs usually possess considerably less whole plots. Secondly, statistical efficiency of experimentation is improved substantially, as was illustrated by comparing the D- and A-criterion values of the alternative experimental setups. Moreover, the efficiency gain turns out to be robust against variance ratio misspecification, which is a desirable property from a practical point of view.

5 Conclusion

Standard response surfaces designs and design construction algorithms become inadequate if an experiment is conducted under a bi-randomization error structure. The use of bi-randomization experiments has been inspired by practical difficulties,

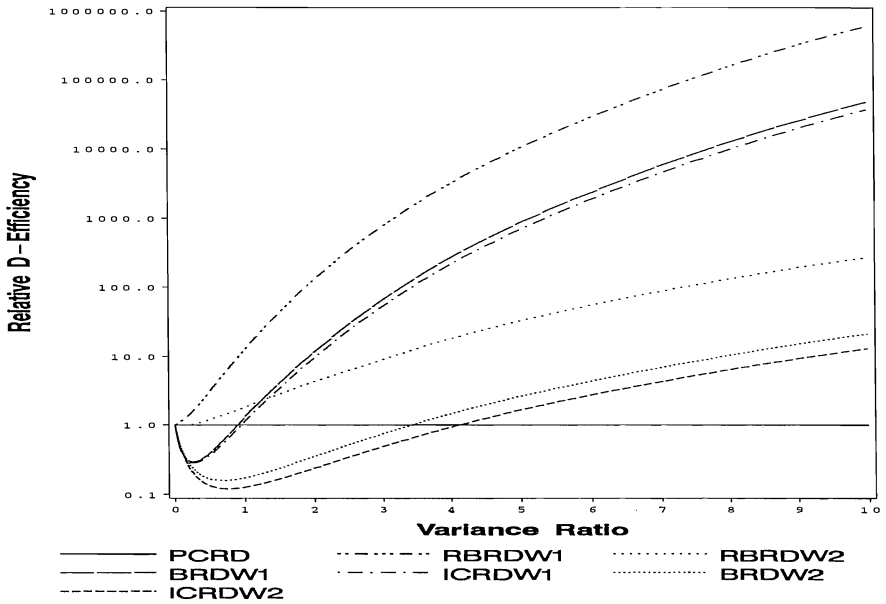


Figure 12: Comparison of the D-efficiency of the BRDs from Section 3.1, the RBRDs and proper and improper use of the CRD for the full quadratic model in three variables.

cost considerations and limited resource availability. This paper provides an algorithm to construct D-optimal bi-randomization designs. Moreover, it extends the concept of bi-randomization designs by allowing more flexibility in the composition of the whole plots. The resulting refined bi-randomization designs are robust against misspecification of the variance ratio and do not only outperform the traditional bi-randomization designs, but in many cases also the completely randomized experiments. Thanks to this increase in efficiency, smaller experiments can be carried out without losing any information. In addition, refined bi-randomization designs are easier to run than completely randomized designs. Finally, the design construction algorithm can cope with practical problems such as design augmentation and restrictions on the number of whole plots.

Acknowledgement

Peter Goos is supported by a doctoral research grant from the Fund for Scientific Research - Flanders. The authors would also like to thank Alexander Donev for his comments on earlier versions of this work.

Appendices

Appendix A. Update formulae for the information matrix of a BRD

Deleting the jth observation within the ith whole plot affects the information matrix of the BRD as follows:

$$\begin{aligned}\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^{*\prime} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} & \\ - \mathbf{f}(\mathbf{z}_i, \mathbf{x}_{ij})\mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{ij}) & \\ + \frac{d}{1+s_i d}(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})(\mathbf{X}_i'\mathbf{1}_{s_i\times 1})' & \\ - \frac{d}{1+(s_i-1)d}(\mathbf{X}_i^{*\prime}\mathbf{1}_{(s_i-1)\times 1})(\mathbf{X}_i^{*\prime}\mathbf{1}_{(s_i-1)\times 1})'. &\end{aligned}$$

When the deleted observation is the only observation within the ith whole plot, $s_i = 1$, $\mathbf{X}_i = \mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{i1})$ and $\mathbf{X}_i^* = \mathbf{0}$. The information matrix then becomes

$$\mathbf{X}^{*\prime}\mathbf{V}^{*-1}\mathbf{X}^{*\prime} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} - \frac{1}{1+d}\mathbf{f}(\mathbf{z}_i, \mathbf{x}_{i1})\mathbf{f}'(\mathbf{z}_i, \mathbf{x}_{i1}).$$

Appendix B. Saturated designs with correlated observations

Let $\mathbf{X}$ denote the design matrix of a given experiment. If a completely randomized experiment is conducted, the D-criterion value is given by $|\mathbf{X}'\mathbf{X}|$. If the experiment is conducted as a bi-randomization experiment, the observations within each whole plot are correlated. Let $\mathbf{R}$ denote the correlation matrix of the observations. The D-criterion value of this experiment can be written as

$$|\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}| = |\mathbf{R}^{-1}||\mathbf{X}'\mathbf{X}|.$$

Since $|\mathbf{R}| \leq 1$ and, consequently, $|\mathbf{R}^{-1}| \geq 1$, the D-criterion value for the bi-randomization experiment is larger than or equal to the value for the completely randomized experiment. Equality holds only if the correlation matrix $\mathbf{R}$ is diagonal. It is clear that the above proof holds for any correlation matrix.

References

[1] V.L. Anderson and R.A. McLean. *Design of Experiments: A Realistic Approach.* Marcel Dekker, Inc., New York, 1974.

[2] A.C. Atkinson and A.N. Donev. The construction of exact D-optimum experimental designs with application to blocking response surface designs. *Biometrika*, 76:515–526, 1989.

[3] A.C. Atkinson and A.N. Donev. *Optimum Experimental Designs.* Oxford U.K.: Clarendon Press, 1992.

[4] G.E.P. Box and S.P. Jones. Split-plot designs for robust product experimentation. *Journal of Applied Statistics*, 19:3–26, 1992.

[5] J.J. Davison. *Response Surface Designs and Analysis for Bi-Randomization Error Structures.* Ph.d. thesis, Virginia Polytechnic Institute and State University, 1995.

[6] V.V. Fedorov. *Theory of Optimal Experiments.* New York: Academic Press, 1972.

[7] O. Kempthorne. *The Design and Analysis of Experiments.* New York: Wiley, 1952.

[8] J. Kiefer. The role of symmetry and approximation in exact design optimality. In S.S. Gupta and J. Yachel, editors, *Statistical Design Theory and Related Topics*, pages 109–118. New York: Academic Press, 1971.

[9] J.D. Letsinger, R.H. Myers, and M. Lentner. Response surface methods for bi-randomization structures. *Journal of Quality Technology*, 28:381–397, 1996.

[10] R.C. Littell, G.A. Milliken, W.W. Stroup, and R.D. Wolfinger. *SAS System for Mixed Models.* SAS Institute Inc., Cary, NC, 1996.

[11] W.J. Welch. Branch-and-bound search for experimental designs based on D-optimality and other criteria. *Technometrics*, 24:41–48, 1982.