Blocking response surface designs

Goos, P. and Donev, A.N.

2005

MIMS EPrint: **2007.162**

ISSN 1749-9097

Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2006) 1075 – 1088
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Blocking response surface designs

P. Goos[a,*], A.N. Donev[b]

[a]Department of Mathematics, Statistics and Actuarial Sciences, Faculty of Applied Economics, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium
[b]AstraZeneca, Mereside, Alderley Park, Macclesfield, SK10 4TG,UK

Received 16 February 2005; received in revised form 10 November 2005; accepted 10 November 2005
Available online 5 December 2005

Abstract

The design of experiments involving more than one blocking factor and quantitative explanatory variables is discussed, the focus being on two key aspects of blocked response surface designs: optimality and orthogonality. First, conditions for orthogonally blocked experiments are derived. Next, an algorithmic approach to compute D-optimal designs is presented. Finally, the relationships between design optimality and orthogonality in the context of response surface experiments are discussed in detail.

Keywords: Exchange algorithm; D-optimality; Fixed blocks; Random blocks; Orthogonality

1. Introduction

In this paper, we focus on the construction of response surface designs in blocks. Blocking is usually beneficial in experimental situations where it is possible to identify groups, or blocks, of experimental units, such that within blocks the experimental units are considerably more homogeneous than the blocks themselves. The present article considers experiments in which more than one source of heterogeneity is present and extends earlier work by Atkinson and Donev (1989), Cook and Nachtsheim (1989), Khuri (1992) and Goos and Vandebroek (2001). The variation between the blocks in the experiment is accounted for by including block effects in the statistical model. In most applications, as in the present paper, the block effects are considered to be nuisance parameters and the accuracy of their estimation is not important.

Depending on what randomisation has been used to form the blocks, their effects could be regarded as *random* or *fixed*. Typically, the blocks are considered random when they are selected from a population at random. However, the implementation of such a selection is often impossible or impractical. In such cases the block effects are usually considered fixed and treated as levels of one or more qualitative variables whose effects are not of interest. Some authors, e.g. Gilmour and Trinca (2000), though mention the possibility of treating the block effects as random as soon as the block labels are randomly assigned to the blocks, even when the block effects are not a random sample from a population. The issue of choosing between random and fixed blocks is also discussed by Ganju (2000).

* Corresponding author. Tel: +32 32 204 059; fax: +32 32 204 817.
E-mail address: peter.goos@ua.ac.be (P. Goos).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2005.11.003

In any case, the nature of the blocking variables has an important impact on the data analysis and therefore on the optimal design for blocked experiments. These topics receive detailed attention in Khuri (1992, 1994), Ganju (2000), Ganju and Lucas (2000) and Gilmour and Trinca (2000), who discuss the analysis of experiments with one blocking variable, and Goos (2002).

In many practical applications the experimenter may deal with both fixed and random blocking variables. Here are two examples.

Example 1. *Valve wear experiment*: In order to establish a measurable valve wear in an internal combustion engine, the latter would need to be run for a long time, typically more than 1000 h. The study required the effect of several variables on the valve wear to be modelled. Some of these variables were parameters of the engine setup, while others were related to the valves themselves (material, dimensions, coatings, etc.).

The wear characteristics of different cylinders in the engines were known to be different in a consistent way, e.g. the end cylinders run cooler than centre cylinders. Also, the scientists believed that the wear of a valve in one cylinder had a negligible effect on the wear of the valves in the other cylinders in the tested operating conditions. In the study, 6 cylinder engines were used, so there were 6 valve positions. Running different valve parameters for each of the 6 cylinders opened the prospect of a six-fold reduction of the total number of engine-hours of testing. This brought in the valve position as a first blocking variable. Although there were 6 different valve positions, the corresponding blocking variable acted at only three levels because valves at equal distances from the centre were similar. Another way to shorten the time of experimentation was to use several engines. This introduced a second blocking variable acting at as many levels as the number of engines used.

Example 2. *Food additives experiment*: An important problem faced by the research laboratory of a food additives producer was to find out how the yield of a starch extraction process depends on the water content of the dough, the flour extraction rate (or milling rate) and the temperature. For each observation in this experiment the raw material, wheat, was milled until the desired flour extraction rate (between 70% and 80%) was reached. Then, water was added after which the gluten in the dough started to agglomerate. The water content was measured by the water/flour ratio, which lay between 0.6 and 1.2. Dough was prepared at different temperatures between 10°C and 40°C. After some time, the gluten were extracted by sieving the mixture. The yield of this process was the percentage of gluten recovered from the wheat. In order to increase the yield of the process, enzymes were added to the dough. The producer used several suppliers of enzymes and the wheat came in different batches. The variation between the enzymes received from different suppliers, as well as between batches of wheat was considerable. Thus, there were two blocking variables in the experiment: suppliers and batches.

The complexity of the design problems in the examples necessitates the use of an algorithmic approach to construct optimum experimental designs for them. Several algorithms for the construction of optimally blocked response surface designs involving one blocking variable have been described in the literature. For instance, Atkinson and Donev (1989, 1992), Cook and Nachtsheim (1989), Miller and Nguyen (1994) and Trinca and Gilmour (2000) discuss the construction of designs with a fixed blocking variable. Goos and Vandebroek (2001) present an algorithm for computing optimal designs in the presence of one random blocking variable. Goos et al. (2005) review the optimal design of blocked experiments. Despite the huge interest in blocked designs, the case when there are several blocking variables has not received much attention. Exceptions are Gilmour and Trinca (2003) who discuss the row–column arrangement of factorial and composite designs, and Ankenman et al. (2003) who study the case where random blocks are formed by two nested factors.

In this paper, we present an algorithm that can be used to generate optimum designs for situations where fixed or random blocks are generated by several crossed blocking variables and where the block structure is dictated by the experimental situation. In addition, the conditions for orthogonally blocking response surface designs are extended to cases with two or more blocking variables and the trade-off between orthogonality and D-optimality is illustrated.

In the next section, the statistical model and the notation are introduced. The conditions for orthogonal blocking are derived in Section 3. The design construction algorithm is described in detail in Section 4. Finally, the algorithm is applied to an interesting problem involving two blocking and two quantitative experimental variables.

2. Models and design optimality

We are interested in the case where a response Y can be explained by a linear model with regressors being functions of, say, m variables. In addition, the experimenter can observe the response in groups, or blocks, of relatively homogeneous observations. We assume the experiment involves B blocking variables that have $b_1, b_2, \ldots, b_B$ levels, respectively; B_F of the blocking variables bring f fixed effects while B_R of them bring $r = \sum_{i=1}^{B_R} b_i$ random effects to the model, $B = B_R + B_F$. We use the subscripts R and F to point out that the block effects are treated as random or fixed, respectively.

In this general setting the model that has to be estimated can be written as

$$\begin{aligned} Y &= \mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\gamma} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \\ &= \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} \\ &= \mathbf{F}\boldsymbol{\xi} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}, \end{aligned} \tag{1}$$

where Y is a vector of n observations, $\mathbf{X}$ is an $n \times p$ matrix with rows corresponding to the values of the regressors for the individual observations, $\boldsymbol{\beta}$ is a vector of p regression parameters, $\mathbf{C}$ is an indicator matrix for the levels of the fixed blocking variables, $\boldsymbol{\gamma}$ is a vector of f fixed block effects, $\mathbf{Z}$ is an indicator matrix for the levels of the random blocking variables, $\boldsymbol{\delta}$ is a vector of r random effects, $\mathbf{W} = [\mathbf{C}\ \ \mathbf{Z}]$, $\boldsymbol{\alpha}$ is a vector of all block effects, $\mathbf{F} = [\mathbf{X}\ \ \mathbf{C}]$ and $\boldsymbol{\xi}$ collects all fixed effects of the ith random-blocking variable in the model. We assume that the vector $\boldsymbol{\varepsilon}$ contains the errors of the observations, which are independent and normally distributed with zero mean and variance σ_ε^2, and that the block effects are additive. The random block effects of the ith random-blocking variable in the model are assumed to be normally distributed with zero means and variances $\sigma_i^2,\ i = 1, 2, \ldots, B_R$. Also it is assumed that they are independent from each other and from $\boldsymbol{\varepsilon}$. In other words,

$$\boldsymbol{var}(\boldsymbol{\delta}) = \mathbf{G} = \operatorname{diag}\left\{\sigma_1^2\mathbf{I}_{b_1}, \sigma_2^2\mathbf{I}_{b_2}, \ldots, \sigma_{B_R}^2\mathbf{I}_{b_{B_R}}\right\},$$

where $b_i,\ i = 1, 2, \ldots, B_R$, denotes the number of levels of the ith random blocking variable. Hence

$$\operatorname{var}(\mathbf{y}) = \mathbf{V} = (\mathbf{Z}\mathbf{H}\mathbf{Z}^\mathrm{T} + \mathbf{I}_n)\sigma_\varepsilon^2, \tag{2}$$

where

$$\mathbf{H} = \operatorname{diag}\left\{\eta_1\mathbf{I}_{b_1}, \eta_2\mathbf{I}_{b_2}, \ldots, \eta_{B_R}\mathbf{I}_{b_{B_R}}\right\}$$

and

$$\eta_i = \frac{\sigma_i^2}{\sigma_\varepsilon^2}, \quad i = 1, 2, \ldots, B_R.$$

The generalised least squares (GLS) parameter estimator of the fixed effects ξ is

$$\hat{\boldsymbol{\xi}} = (\mathbf{F}^\mathrm{T}\mathbf{V}^{-1}\mathbf{F})^{-1}\mathbf{F}^\mathrm{T}\mathbf{V}^{-1}\mathbf{Y}, \tag{3}$$

which has variance–covariance matrix

$$\operatorname{var}(\hat{\boldsymbol{\xi}}) = \left(\mathbf{F}^\mathrm{T}\mathbf{V}^{-1}\mathbf{F}\right)^{-1}. \tag{4}$$

As the goal of this paper is to find designs that allow a precise estimation of $\boldsymbol{\beta}$, it is appropriate to focus on the part of (4) corresponding to $\hat{\boldsymbol{\beta}}$ and to use a D_s-optimality criterion that minimizes the corresponding determinant. However, since the block structure is assumed to be determined by the experimental situation, this is equivalent to maximizing

$$D = \det\left(\mathbf{F}^\mathrm{T}\mathbf{V}^{-1}\mathbf{F}\right). \tag{5}$$

This criterion is referred to as the D-optimality criterion. When no fixed blocking factors are involved in the experiment, $\boldsymbol{\xi}$ and $\mathbf{F}$ can be replaced by $\boldsymbol{\beta}$ and $\mathbf{X}$ in (3)–(5). The matrix $\mathbf{Z}$ can be partitioned when there are several random blocking variables: $\mathbf{Z} = \left[\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_{B_R}\right]$ where each $\mathbf{Z}_i$ assigns the observations of the experiment to a level of the ith random blocking variable. When, as in many practical design problems, an equal number of observations, say k, are obtained at each combination of the random blocking variables, so that $n=k\prod_{i=1}^{B_R} b_i$, the information matrix can be rewritten as

$$\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - \sum_{i=1}^{B_R}\eta_i c_i \mathbf{X}^{\mathrm{T}}\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\mathbf{X} + d\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right)\sigma_\varepsilon^{-2}$$
$$= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - \sum_{i=1}^{B_R}\eta_i c_i \sum_{j=1}^{b_i}\mathbf{X}_{ij}^{\mathrm{T}}\mathbf{1}_{a_i}\mathbf{1}_{a_i}^{\mathrm{T}}\mathbf{X}_{ij} + d\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right)\sigma_\varepsilon^{-2}, \tag{6}$$

where $c_i = \left(1+\left(n\eta_i/b_i\right)\right)^{-1}$, $\mathbf{X}_{ij}$ is the part of $\mathbf{X}$ assigned to the jth level of the ith blocking factor, $a_i = n/b_i$ and

$$d = \frac{n\sum_{i=1}^{B_R}\sum_{\substack{j=1\\ j\neq i}}^{B_R}\left(\eta_i\eta_j/b_ib_j\right)c_i}{1+n\sum_{i=1}^{B_R}\eta_i/b_i}. \tag{7}$$

This expression generalizes a similar expression for a single random-blocking variable in Khuri (1992) and follows directly from Lemma 1 (see below). The lemma is also useful for the proofs of the theorems in the next section and can be used as a basis for fast update formulas in a design construction algorithm.

Lemma 1. *If all block effects in the model are random and additive, and k experimental runs are performed at each combination of levels of the blocking variables, then*

$$\mathbf{V} = \left(\mathbf{I}_n + \sum_{i=1}^{B_R}\eta_i\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}}\right)\sigma_\varepsilon^2 \tag{8}$$

and its inverse is given by

$$\mathbf{V}^{-1} = \left(\mathbf{I}_n - \sum_{i=1}^{B_R}\eta_i c_i\mathbf{Z}_i\mathbf{Z}_i^{\mathrm{T}} + d\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\sigma_\varepsilon^{-2}, \tag{9}$$

where d is given by (7).

A proof of Lemma 1 is obtained by multiplying (8) and (9) and observing that this yields the identity matrix.

3. Orthogonality

It is well known that when the experimental design for the explanatory variables is orthogonal to that for the blocking variables, the estimation and the interpretation of the results are simplified. In this section, conditions for orthogonal blocking of experiments involving quantitative variables generated by both fixed and random blocking variables are derived.

3.1. Conditions for orthogonality

If we denote $\mathbf{X} = \begin{bmatrix}\mathbf{1}_n & \tilde{\mathbf{X}}\end{bmatrix}$ and $\boldsymbol{\beta}^{\mathrm{T}} = \begin{bmatrix}\beta_0 & \tilde{\boldsymbol{\beta}}^{\mathrm{T}}\end{bmatrix}$, model (1) can be rewritten as

$$\mathbf{y} = \lambda_0\mathbf{1}_n + \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{W}}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where

$$\tilde{\mathbf{W}} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\mathbf{W}$$

Table 1
An orthogonally blocked design for a quadratic model in one explanatory variable

Blocking variable 2	Blocking variable 1			
	Level 1		Level 2	
	x	x^2	x	x^2
Level 1	−1	1		
	−1	1		
			−1	1
	0	0		
			0	0
	0	0		
			1	1
	1	1		
	1	1		
Level 2	−1	1		
	−1	1		
			−1	1
	0	0		
			0	0
	0	0		
			1	1
	1	1		
	1	1		

and

$$\lambda_0 = \beta_0 + \frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\mathbf{W}\boldsymbol{\alpha}.$$

By definition, a design is orthogonally blocked if the columns of $\mathbf{X}$ are orthogonal to those of $\tilde{\mathbf{W}}$, that is if

$$\mathbf{X}^{\mathrm{T}}\tilde{\mathbf{W}} = \mathbf{X}^{\mathrm{T}}\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\mathbf{W} = \mathbf{0}_{p\times(f+r)}.$$

Here, $\mathbf{0}_{p\times(f+r)}$ is a $p \times (f+r)$ matrix of zeros. This condition holds if

$$\frac{1}{n_{ij}}\mathbf{X}_{ij}^{\mathrm{T}}\mathbf{1}_{n_{ij}} = \frac{1}{n}\mathbf{X}^{\mathrm{T}}\mathbf{1}_n, \quad i = 1, 2, \ldots, B; \quad j = 1, 2, \ldots, b_i, \tag{10}$$

where $\mathbf{X}_{ij}$ is the part of $\mathbf{X}$ corresponding to the jth level of the ith blocking variable and n_{ij} is the number of observations at that level. Note that $\sum_{j=1}^{b_i} n_{ij} = n$ for every i. As a result, the condition for orthogonal blocking states that the average level of the regressors should be the same at each level of each blocking variable. The conditions defined by (10) extend Box and Hunter's (1957) and Khuri's (1992) definitions of orthogonal blocking in the case of one blocking variable. Using tedious matrix algebra it can be shown that if an experiment is orthogonally blocked and the block sizes are equal, the estimates of $\boldsymbol{\beta}$ obtained by ignoring the blocks, by treating them as fixed or by treating them as random are equivalent. This extends similar results in Khuri (1992) and Goos and Vandebroek (2004) for blocked response surface designs generated by a single blocking variable.

Example 3. Suppose that an experimental design for a quadratic model in one explanatory variable with 18 observations divided in 4 blocks is needed. The blocks are formed by the two levels of two blocking variables. Eq. (10) allows us to verify that the estimators for the model parameters of interest are orthogonal to those for the block effects for the design listed in Table 1. This design has two blocks of size 6 and two blocks of size 3. Replicating the smaller blocks would also result in an orthogonal design. These designs are fully orthogonal regardless of whether the blocks are fixed or random. It could be easily shown that both designs are also D-optimum.

3.2. Experiments with random blocks

Lemma 1 allows us to prove the following useful result for blocking an experimental design with random blocks.

Theorem 1. *When the effects of all blocking variables are assumed random and independent from each other, and the number of runs at all combinations of levels of the blocking variables is equal:*

(i) *for a given design, with extended design matrix* $\mathbf{X}$ *and a given variance–covariance matrix* $\mathbf{V}$, *the assignment of the observations to the blocks that produces an orthogonally blocked design is best with respect to any generalized optimality criterion based on the information matrix of the design;*
(ii) *a design which maximizes* $\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)$ *and is orthogonally blocked maximizes also* $\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right)$ *for a given* $\mathbf{V}$.

The proof of the theorem is provided in Appendix A. Regarding part (ii) of the theorem, it should be pointed out that if the design which maximizes $\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)$ cannot be orthogonally blocked, the determinant of the resulting information matrix, $\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right)$, can be larger or smaller than that of an orthogonally blocked design obtained from a design that does not maximize $\det(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ depending on the variance components.

3.3. Experiments with fixed blocks

We obtain stronger results than those in the previous section for the case when the blocks are fixed because, as opposed to Theorem 1, the following result, a proof of which is given in Appendix B, is also valid for unequal numbers of runs at the combinations of the levels of the blocking variables.

Theorem 2. *When the effects of all blocking variables are assumed fixed and additive,*

(i) *for a given design, with extended design matrix* $\mathbf{X}$ *and a given block structure, the assignment of the observations to the blocks that produces an orthogonally blocked design is best with respect to the D-optimality criterion* $\det\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}\right)$;
(ii) *a design which maximizes* $\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)$ *and is orthogonally blocked maximizes also* $\det\left(\mathbf{F}^{\mathrm{T}}\mathbf{F}\right)$ *for a given block structure.*

Theorems 1 and 2 extend the results of Goos and Vandebroek (2001), who show that orthogonal blocking is an optimal assignment strategy if there is a single random-blocking variable and the block sizes are equal, and if the block effects are fixed. Extending Theorems 1 and 2 to the case where there are both fixed and random blocks and where there are interactions between blocking variables is complicated but our empirical experience suggests that similar results also hold in these cases.

Note that Example 3 illustrates Theorem 2. Also, arranging the points of a 2-level complete or fractional factorial design in blocks so that $\mathbf{X}_{ij}^{\mathrm{T}}\mathbf{1}_{a_i}=\mathbf{0}_p$, where $a_i=n/b_i$, produces a D-optimum design for first-order models that can include interactions between the explanatory variables in both cases where the blocks are fixed or random. The condition $a_i=n/b_i$ is only needed when the block effects are treated random.

4. Construction of block designs

Finding an experimental design with the required block structure that is D-optimum is usually not easy. Combinatorial results could help but only in some particular cases. In general, however, block experimental designs can be obtained using a computer search. As already noted, the available algorithms for experiments involving quantitative variables can only be used to produce designs where a single blocking variable generates the blocks. The following algorithm can be used to construct designs with complicated block structures. We start by describing the algorithm's input and output. Next, the structure of the algorithm is outlined and an example is discussed.

4.1. Input and output

The input to the algorithm is essentially a description of the design problem and includes the specification of the polynomial regression model in the experimental variables, the number of blocking variables, their nature (random or fixed), the number of levels of each of them, and the number of observations for each combination of these levels. The number of observations at each combination of the levels of the blocking variables is allowed to be heterogeneous.

In addition to these parameters, the user has to specify an η_i-value for each blocking variable that is treated as random. This could be problematic as the magnitude of this parameter is unknown prior to the experiment. Fortunately, the designs generated by the algorithm do not vary much for practical η_i-values. This is illustrated in Example 4.

A set of candidate test combinations of the experimental variables has to be provided as well. By default, the algorithm uses a factorial design as a candidate set. The number of levels used for each variable depends on the maximum order with which these variables appear in the regression model. For example, two levels are used for variables whose linear effects are of interest and three levels are used for variables whose quadratic effects are to be estimated. A different candidate set can be specified if the levels of the variables are subject to constraints or if the experimental region is spherical.

Finally, the user has to specify the number of times the algorithm is run. The purpose of running the algorithm several times is to decrease the probability of ending up in a local optimum. The algorithm's output consists of a D-optimum design having the required block structure.

4.2. Structure of the algorithm

Step 1: A non-singular, feasible starting design is generated. Part of this design is created by randomly selecting points from the candidate set and assigning them randomly to the blocks of the experiment. The starting design is then completed by sequentially allotting the candidate with the largest prediction variance to a randomly chosen block, provided the number of candidates already assigned to that block does not equal the maximum number.

Step 2: The algorithm then attempts to improve the starting design with respect to the criterion of optimality (5) in two ways:

(a) All possible exchanges of design points and candidate points are evaluated. The exchange that generates the largest increase is stored.
(b) All possible swaps of two points from different blocks are evaluated. The swap that generates the largest increase in the D-criterion value is stored.

Step 3: Next, the best exchange is performed by selecting the better of the two stored changes.

Steps 2 and 3 are repeated until no more beneficial exchanges can be found.

Step 4: The resulting design is compared to the best one found so far and stored if it is better.

Steps 1–3 are repeated the required number of times. The larger the number of times this is done the larger the probability that the global optimum will be found.

4.3. Computational results

In this section, we illustrate a number of important aspects with an example. The example clearly demonstrates that the structure of the optimum design depends on the nature of the blocking variables. Also, it shows that the optimum designs are not sensitive to the η_i-values specified by the user, provided these values are not too close to zero. This is usually the case for experiments where blocking is considered. Finally, the example illustrates that using the D-optimality criterion often leads to finding an optimum design that is orthogonally blocked. The larger the η_i-values, the more likely it is that the design produced by the algorithm will be orthogonally blocked. However it should be stressed that there are many experimental situations in which no orthogonally blocked designs can be found. This is especially the case for experiments with small numbers of observations in the blocks and for models involving higher order terms.

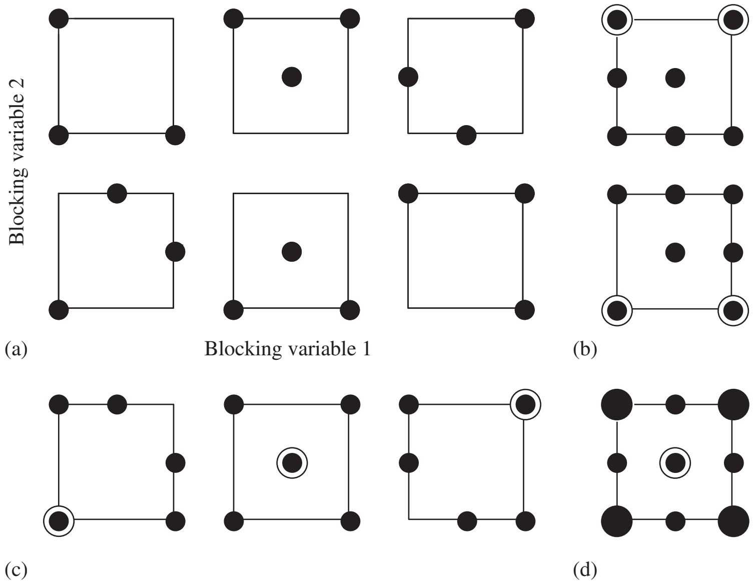

Fig. 1. D-optimum design for a full quadratic model in two explanatory variables in the presence of two blocking variables when the values of η_1 and η_2 fall in the region I shown in Fig. 5. A small bullet represents a single design point, a circled bullet represents a duplicated design point, and a large bullet represents a triplicated design point: (a) Design I; (b) Horizontal projection of Design I; (c) Vertical projection of Design I; (d) Overall projection of Design I.

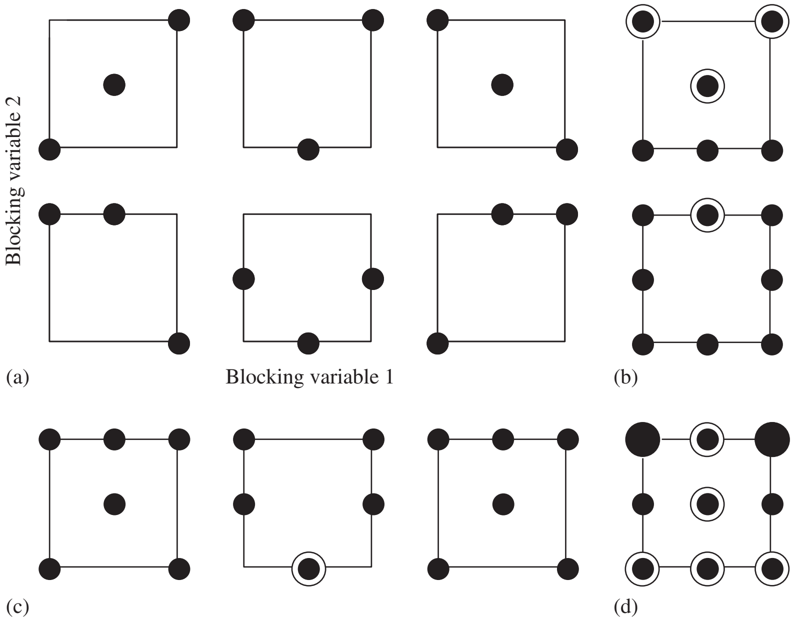

Fig. 2. D-optimum design for a full quadratic model in two explanatory variables in the presence of two blocking variables when the values of η_1 and η_2 fall in the region II shown in Fig. 5. A small bullet represents a single design point, a circled bullet represents a duplicated design point, and a large bullet represents a triplicated design point: (a) Design II; (b) Horizontal projection of Design II; (c) Vertical projection of Design II; (d) Overall projection of Design II.

Example 4. Suppose that a second-order polynomial model in two explanatory variables can explain the response, and that there are two random-blocking variables, acting at 2 and 3 levels, respectively. The experimenter can carry out three observations for each combination of levels of the blocking variables, so that the entire experiment comprises

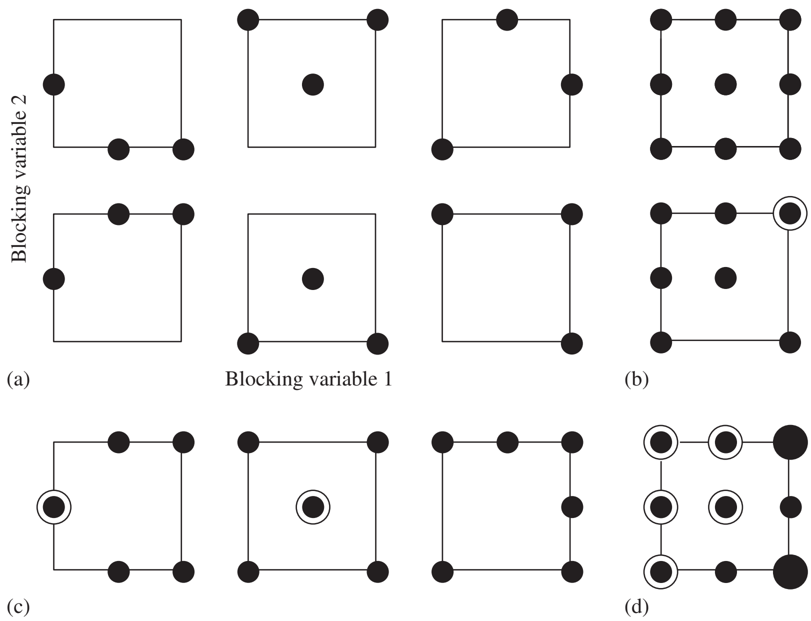

Fig. 3. *D*-optimum design for a full quadratic model in two explanatory variables in the presence of two blocking variables when the values of η_1 and η_2 fall in the region III shown in Fig. 5. A small bullet represents a single design point, a circled bullet represents a duplicated design point, and a large bullet represents a triplicated design point: (a) Design III; (b) Horizontal projection of Design III; (c) Vertical projection of Design III; (d) Overall projection of Design III.

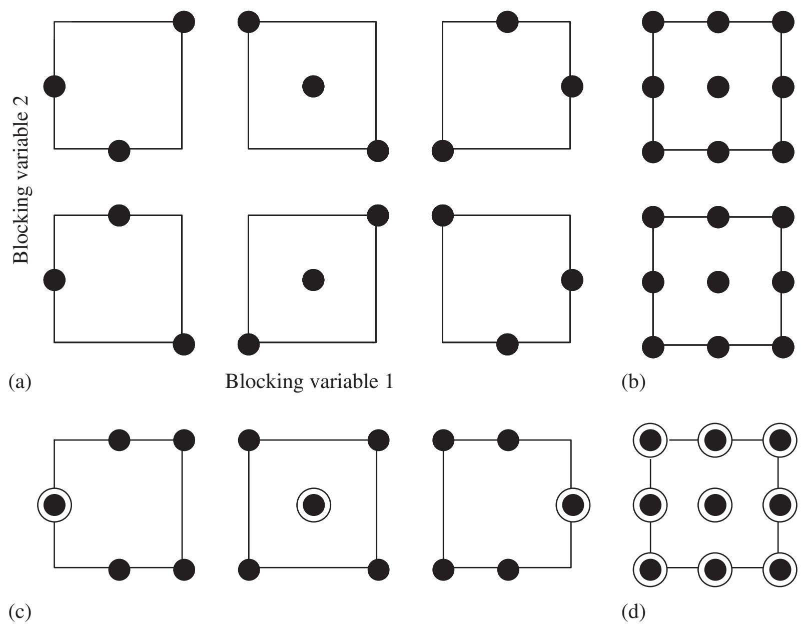

Fig. 4. *D*-optimum design for a full quadratic model in two explanatory variables in the presence of two blocking variables when the values of η_1 and η_2 fall in the region IV shown in Fig. 5. A small bullet represents a single design point, a circled bullet represents a duplicated design point, and a large bullet represents a triplicated design point: (a) Design IV; (b) Horizontal projection of Design IV; (c) Vertical projection of Design IV; (d) Overall projection of Design IV.

$n = 18$ observations. Fig. 1 (a), Fig. 2(a), Fig. 3(a) and Fig. 4(a) show designs that are best with respect to the *D*-optimality criterion for different values of η_1 and η_2. Fig. 5 gives the values of η_1 and η_2 for which each of these designs is *D*-optimum. Note that the case when η_1 and η_2 are very large is equivalent to that when both blocking variables generate

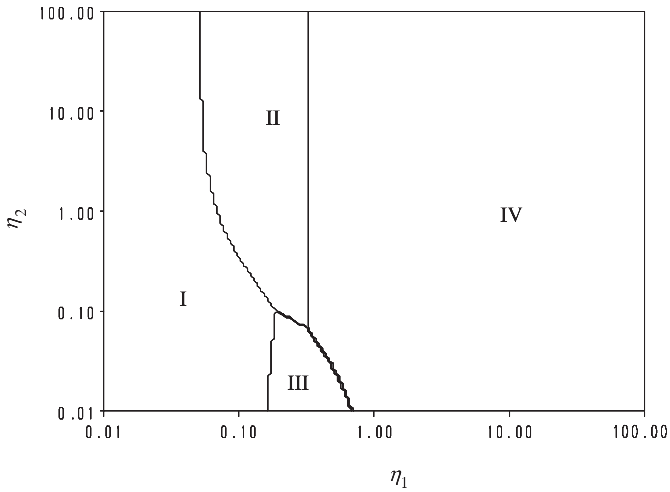

Fig. 5. Regions of optimality of Designs I, II, III and IV.

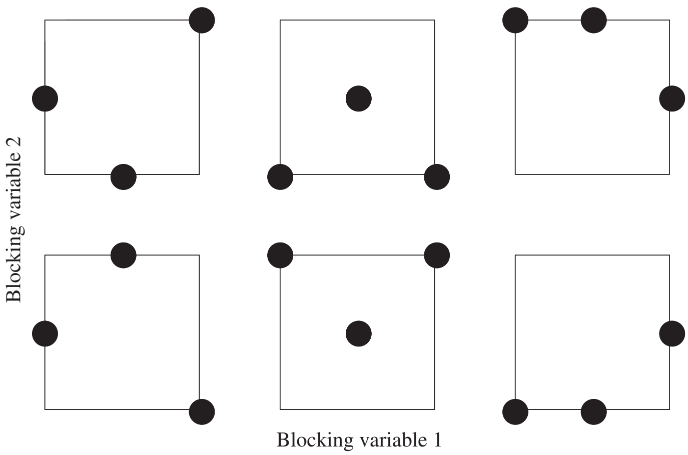

Fig. 6. *D*-optimum design for a full quadratic model in two explanatory variables in the presence of two blocking variables when the values of η_1 and η_2 fall in the region IV shown in Fig. 5.

fixed effects. On the other hand, when η_i is small there is no efficiency benefit from blocking with the *i*th blocking variable.

It can easily be verified that Design IV satisfies Eq. (10) and is orthogonal, whereas Design I–III are not orthogonal. The projections with respect to each of the blocking variables and with respect to both blocking variables are shown in parts (b), (c) and (d) of each of the Figs. 1–4. It is interesting to note that the designs and the projections are different, and that only the overall projection of Design I is identical to the 18-point *D*-optimum design when the design has not been blocked, i.e. when $\eta_1 = \eta_2 = 0$. This illustrates the trade-off the algorithm makes between the choice of the design points and the orthogonality of their arrangement in blocks.

The resulting designs in this example are fairly robust against inaccurate specification of these values, especially when they are relatively large. For example, Fig. 5 shows that Design IV is *D*-optimum for $\eta_1 > 19/60$ and for almost any η_2.

Finally, it is interesting to note that a *D*-optimal design may not be unique. For example, the design shown in Fig. 6 is also *D*-optimum and orthogonal for all values of η_1 and η_2 for which Design IV is *D*-optimum. A secondary criterion can be used to choose amongst designs that perform equally well with respect to the main criterion of optimality. As one of the referees suggested, the design in Fig. 4 may be preferred over that in Fig. 6 as each level of each of the quantitative variables appears at least once in each block, thus ensuring some form of balance within blocks. Other criteria may also be considered. For example, one may take into account the cost of the experiment when each of these designs is used, or the maximum or average prediction variances.

5. Discussion

The previous section demonstrates that when several variables define the block structure of an experiment, the algorithmic construction leads to experimental designs with excellent statistical properties. Although the focus in this paper is on quantitative explanatory variables, the algorithm can easily handle problems involving also one or more qualitative variables. Experiments with mixtures or with other constraints can also be tackled (see, for example, Goos and Donev (2006)).

Block designs with more blocking variables and different numbers of levels can be generated in a similar way. However, the probability of finding just a local optimum increases with the number of levels of the blocking variables and with the complexity of the regression model under investigation. In such cases the algorithm has to be run a sufficiently large number of times. The latter would usually be easy to do as the modern computers are fast and relatively cheap.

Using the D-criterion of optimality for blocking experimental designs introduces a number of desirable features to the studies where they are used. However, it also brings in complexity in the interpretation of the results. This problem is considerably reduced when the designs are orthogonally blocked, i.e. when the effects of the experimental variables contained within $\boldsymbol{\beta}$ are estimated independently from the block effects. Although there are many cases when the D-optimum designs are orthogonal, there are a lot more when an orthogonal design for the required block structure does not exist. Fortunately, many designs that are D-optimum are nearly orthogonal. This is particularly useful in cases where the block structure is complex. There are also situations in which the D-optimum design is orthogonal for a subset of the parameters of interest. For example, Design II in Fig. 2 is orthogonally blocked for a regression model comprising only the terms x_1, x_1^2 and x_1x_2, but not for models containing terms for x_2 or x_2^2.

Acknowledgements

Most of the research that led to this paper was carried out while Alexander Donev was a lecturer at the Department of Probability and Statistics of the University of Sheffield and while Peter Goos was associated to the Department of Applied Economics of the Katholieke Universiteit Leuven as a Postdoctoral Researcher of the Fund for Scientific Research, Flanders (Belgium). The authors are grateful to Alan Collins and Stefaan Roels for providing the industrial examples and to Randy Tobias for pointing out the optimality of the design in Fig. 3.

Appendix A. Proof of Theorem 1

Substituting the conditions for orthogonal blocking defined by Eq. (10) in (6) and noting that, in the case of equal numbers of runs at each combination of levels of the blocking variables, $n/n_{ij} = b_i$, we obtain

$$\begin{aligned}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X} &= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - \sum_{i=1}^{B_R}\eta_i c_i \sum_{j=1}^{b_i}\frac{1}{b_i^2}\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X} + d\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right)\sigma_\varepsilon^{-2}\\ &= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - \sum_{i=1}^{B_R}\frac{\eta_i c_i}{b_i}\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X} + d\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right)\sigma_\varepsilon^{-2}\\ &= \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - g\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right)\sigma_\varepsilon^{-2}, \qquad \text{(A.1)}\end{aligned}$$

where

$$g=\sum_{i=1}^{B_R}\frac{\eta_i c_i}{b_i}-d.$$

If a design is not orthogonally blocked and has equal block sizes,

$$\frac{1}{b_i}\mathbf{X}^{\mathrm{T}}\mathbf{1}_n+\mathbf{o}_{ij}=\mathbf{X}_{ij}^{\mathrm{T}}\mathbf{1}_{a_i},\quad i=1,2,\ldots,B_R;\quad j=1,2,\ldots,b_i,$$

where $a_i=n/b_i$ and $\mathbf{o}_{ij}$ is a vector with at least one non-zero element. Then the information matrix (A.1) can be written as

$$\begin{aligned}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}&=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}-\sum_{i=1}^{B_R}\eta_i c_i\sum_{j=1}^{b_i}\left(\frac{1}{b_i}\mathbf{X}^{\mathrm{T}}\mathbf{1}_n+\mathbf{o}_{ij}\right)\left(\frac{1}{b_i}\mathbf{1}_n^{\mathrm{T}}\mathbf{X}+\mathbf{o}_{ij}^{\mathrm{T}}\right)+d\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right)\sigma_\varepsilon^{-2}\\&=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}-g\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}-\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\sum_{i=1}^{B_R}\frac{\eta_i c_i}{b_i}\sum_{j=1}^{b_i}\mathbf{o}_{ij}^{\mathrm{T}}-\left(\sum_{i=1}^{B_R}\frac{\eta_i c_i}{b_i}\sum_{i=1}^{b_i}\mathbf{o}_{ij}\right)\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right.\\&\quad\left.-\sum_{i=1}^{B_R}\eta_i c_i\sum_{j=1}^{b}\mathbf{o}_{ij}\mathbf{o}_{ij}^{\mathrm{T}}\right)\sigma_\varepsilon^{-2}.\end{aligned}$$

Since $\sum_{j=1}^{b_i}\mathbf{o}_{ij}=\mathbf{0}_p$ for every possible arrangement (orthogonal or not) of a given design, this can be simplified to

$$\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}=\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}-g\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}-\sum_{j=1}^{B_R}\eta_i c_i\sum_{j=1}^{b_i}\mathbf{o}_{ij}\mathbf{o}_{ij}^{\mathrm{T}}\right)\sigma_\varepsilon^{-2}.\tag{A.2}$$

This shows that the difference between the information matrices of an orthogonally blocked design and that of a non-orthogonally blocked design is a non-negative definite matrix. Hence, orthogonal blocking, if it exists, of a given extended design matrix $\mathbf{X}$ ensures that the resulting design is not just D-optimum but also optimum with respect to any generalized criterion based on the information matrix of the design regardless of the values of η_i. This concludes the proof of part (i) of the theorem.

Assume that a design with an extended design matrix $\mathbf{X}$ maximizes $\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)$. If $\mathbf{X}$ can be arranged in blocks so that the resulting experiment is orthogonally blocked, the information matrix is given by (A.1). The determinant of this matrix is

$$\begin{aligned}\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X}\right)&=\sigma_\varepsilon^{-2p}\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}-g\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\right)\\&=\sigma_\varepsilon^{-2p}\left(1-g\mathbf{1}_n^{\mathrm{T}}\mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\right)\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\\&=\sigma_\varepsilon^{-2p}\left(1-g\mathbf{u}_1^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{1}_n\right)\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\\&=\sigma_\varepsilon^{-2p}(1-gn)\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right),\end{aligned}\tag{A.3}$$

by observing that $\mathbf{u}_1^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}=\mathbf{1}_n^{\mathrm{T}}$, where $\mathbf{u}_1^{\mathrm{T}}=\left[1\ \ \mathbf{0}_{p-1}^{\mathrm{T}}\right]$. We now show that any other design with extended design matrix $\mathbf{A}$ for which $\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)<\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)$ will produce a smaller determinant $\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{A}\right)$ than that given by (A.3). Note that irrespective of how the design points of $\mathbf{A}$ are assigned to the blocks, the corresponding information matrix can be written as

$$\mathbf{A}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{A}=\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}-g\mathbf{A}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{A}-\sum_{i=1}^{B_R}\eta_i c_i\sum_{j=1}^{b_i}\mathbf{o}_{ij}\mathbf{o}_{ij}^{\mathrm{T}}\right)\sigma_\varepsilon^{-2},\tag{A.4}$$

where

$$\frac{1}{b_i}\mathbf{A}^{\mathrm{T}}\mathbf{1}_n + \mathbf{o}_{ij} = \mathbf{A}_{ij}^{\mathrm{T}}\mathbf{1}_{a_i}, \quad i = 1, 2, \ldots, B_R; \quad j = 1, 2, \ldots, b_i.$$

From (A.3) and (A.4) it follows that

$$\begin{aligned}\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} - g\mathbf{A}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{A}\right) &= (1 - gn)\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right) \\ &< (1 - gn)\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)\end{aligned}$$

and that

$$\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{A}\right) < \sigma_{\varepsilon}^{-2p}\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} - g\mathbf{A}^{\mathrm{T}}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\mathbf{A}\right),$$

so that

$$\det\left(\mathbf{A}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{A}\right) < (1 - gn)\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right).$$

This completes the proof of the theorem.

Appendix B. Proof of Theorem 2

Assume the block effects in the model are all fixed. The information matrix is given by

$$\sigma_{\varepsilon}^{-2}\mathbf{F}^{\mathrm{T}}\mathbf{F} = \sigma_{\varepsilon}^{-2}\begin{bmatrix}\mathbf{X}^{\mathrm{T}}\mathbf{X} & \mathbf{X}^{\mathrm{T}}\mathbf{C} \\ \mathbf{C}^{\mathrm{T}}\mathbf{X} & \mathbf{C}^{\mathrm{T}}\mathbf{C}\end{bmatrix}$$

and its determinant is given by

$$\det\left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - \mathbf{X}^{\mathrm{T}}\mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{X}\right)\sigma_{\varepsilon}^{-2(p+f)}.$$

Since the blocking arrangement is pre-specified, the first factor in this expression is a constant and the D-optimum design is obtained by maximizing

$$\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - \mathbf{X}^{\mathrm{T}}\mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{C}^{-1}\right)\mathbf{C}^{\mathrm{T}}\mathbf{X}\right).$$

For a given design matrix $\mathbf{X}$, this determinant is maximum if the design is blocked orthogonally. The easiest way to see this is by using so-called effects type coding instead of using dummy coding for parametrizing the block effects. In that case, the elements of $\mathbf{C}$ corresponding to observations at the last level of a blocking variable are no longer equal to 0 but to -1. An orthogonal arrangement of $\mathbf{X}$ then makes that $\mathbf{X}^{\mathrm{T}}\mathbf{C} = \mathbf{0}_{p\times f}$ and that

$$\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X} - \mathbf{X}^{\mathrm{T}}\mathbf{C}\left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)^{-1}\mathbf{C}^{\mathrm{T}}\mathbf{X}\right) = \det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right).$$

This proves part (i) of the theorem. Obviously, if $\mathbf{X}$ is chosen to maximise $\det\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)$, then the design is also D-optimum. This proves part (ii) of the theorem.

References

Ankenman, B.E., Avilés, A.I., Pinheiro, J.C., 2003. Optimal designs for mixed effect models with two random nested factors. Statist. Sin. 13, 385–401.

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

Atkinson, A.C., Donev, A.N., 1992. Optimum Experimental Design. Clarendon Press, Oxford.

Box, G.E.P., Hunter, J.S., 1957. Multi-factor experimental designs for exploring response surfaces. Ann. Math. Statist. 28, 195–241.

Cook, R.D., Nachtsheim, C.J., 1989. Computer-aided blocking of factorial and response-surface designs. Technometrics 31, 339–346.

Ganju, J., 2000. On choosing between fixed and random block effects in some no-interaction models. J. Statist. Plann. Inference 90, 323–334.

Ganju, J., Lucas, J.M., 2000. Analysis of unbalanced data from an experiment with random block effects and unequally spaced factors. Amer. Statist. 54, 5–11.

Gilmour, S.G., Trinca, L.A., 2000. Some practical advice on polynomial regression analysis from blocked response surface designs. Comm. Statist.: Theory Methods 29, 2157–2180.

Gilmour, S.G., Trinca, L.A., 2003. Row-column response surface designs. J. Qual. Technol. 35, 184–193.

Goos, P., 2002. The Optimal Design of Blocked and Split-plot Experiments. Springer, New York.

Goos, P., Donev, A.N., 2006. The *D*-optimal design of blocked experiments with mixture components. J. Qual. Technol., to appear.

Goos, P., Tack, L., Vandebroek, M., 2005. The optimal design of blocked experiments in industry. In: Berger, M., Wong, W.K. (Eds.), Applied Optimal Design. Wiley, New York, pp. 247–279.

Goos, P., Vandebroek, M., 2001. *D*-optimal response surface designs in the presence of random block effects. Comput. Statist. Data Anal. 37, 433–453.

Goos, P., Vandebroek, M., 2004. Estimating the Intercept in an Orthogonally Blocked Experiment When the Block Effects Are Random. Comm. Statist.: Theory Methods 33, 873–890.

Khuri, A.I., 1992. Response surface models with random block effects. Technometrics 34, 26–37.

Khuri, A.I., 1994. Effect of blocking on the estimation of a response surface. J. Appl. Statist. 21, 305–316.

Miller, A.J., Nguyen, N.K., 1994. AS 295—A Fedorov exchange algorithm for *D*-optimal design. Appl. Statist. 43, 669–678.

Trinca, L.A., Gilmour, S.G., 2000. An algorithm for arranging response surface designs in small blocks. Comput. Statist. Data Anal. 33, 25–43.