Listy Biometryczne - Biometrical Letters
Vol. 37(2000), No. 2, 67-95

Equivalence theory for design augmentation and parsimonious model checking: response surfaces and yield density models

A.C. Atkinson[1], B. Bogacka[2] and S.S. Zocchi[3]

[1]Department of Statistics, The London School of Economics, London WC2A 2AE, UK,

[2]School of Mathematical Sciences, Queen Mary, University of London, London E1 4NS

[3]Departamento de Matemática e Estatística, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, SP, Brazil.

SUMMARY

We consider two applications of optimum design theory to three examples. In design augmentation, trials are added optimally to an existing non-optimal design. In model checking we design for departures from an assumed model, but in a way that involves few extra trials. For both problems we find D-optimum designs. Special cases of the equivalence theorem for D-optimality provide algorithms for design construction and insight into the structure of the designs. Simple examples of designs for second-order response surface models over regular and irregular regions are followed by an application of the optimum non-linear design to field experiments on plant density.

KEY WORDS: Bayesian design, D-optimality, irregular region, non-linear model, response surface.

1. Introduction

Many crops are grown in rows. The yield per plant, or per unit area, depends both on the distance between the rows and on the distance between the plants in the rows. In Section 6 we present a general nonlinear model for the dependence of plant yield on these two distances which contains seven parameters. A simple alternative model contains only three parameters. To estimate all parameters in the general model requires experiments at at least seven combinations of inter- and intra-plant distances. Data from such an experiment can be used to check whether the simple

model holds by testing whether four of the parameters are equal to zero. If they are, much of the experimental effort will have been devoted to model checking and relatively little to estimates of the three remaining parameters. We use developments of the methods of D-optimality to obtain parsimonious designs for checking the simple yield-density model which require experiments at only four combinations of distances and which are also efficient for estimating the parameters of the simple model.

If the design for model checking indicates that the simple model does not hold, we show how to augment the design in an optimal manner to get estimates of the parameters in the general model. Both design augmentation and model checking can be formulated as problems in which some prior information is available. The formulation for model checking is due to DuMouchel and Jones (1994). We extend the equivalence theorem for D-optimality to these cases in which specific forms of prior information are available. The extensions yield optimality conditions for which there is a ready physical interpretation. We use the results in calculating the optimum designs and in illustrating the change of structure of designs with the amount of prior information. There are three examples, two for linear models and one for the nonlinear yield-density model.

The theory is in Section 2, with the Equivalence Theorem conditions for D-optimality in Section 2.2. The algorithm for finding optimum designs is discussed in Section 2.3. The special cases for design augmentation, Section 3, and model checking, Section 4, follow. We exemplify the theory with designs of both kinds for second-order response surfaces over both regular and, in Section 5, irregular regression regions. Plots of the variances of prediction illustrate the properties of the designs. In Section 6 we apply the extension of the theory to nonlinear models and calculate D-optimum designs for checking the model on the relationship of yield and density. This model is a nonlinear form of a two-factor response surface model. Concluding remarks are in Section 7.

2. Theoretical background

2.1. Notation and definitions

We start with a short description of optimum design theory which is applicable to both linear and nonlinear models. Let the triple

$$\{\mathbf{y}, E(\mathbf{y}), D(\mathbf{y})\}$$

be the statistical model of the observations, where $\mathbf{y} \in \mathcal{R}^n$ is vector of n observable random variables, $E(\mathbf{y})$ its expected value and $D(\mathbf{y})$ its dispersion, or variance-covariance, matrix. In this paper bold lower case will be used to denote vectors and

bold upper case will denote matrices. The expected value of the jth observation is given by

$$E(y_j) = \eta(\mathbf{x}_j, \boldsymbol{\theta}), \quad j = 1, \ldots, n,$$

where $\mathbf{x}_j \in \mathcal{X}$ is an m-dimensional vector of experimental conditions, $\mathcal{X} \subset \mathcal{R}^m$ and $\boldsymbol{\theta} \in \Theta$ denotes a p-dimensional vector of parameters, $\Theta \subset \mathcal{R}^p$. The random errors of observation, say e_j, are assumed to be additive and normally, identically, independently distributed with zero mean and constant variance σ^2, that is $D(\mathbf{y}) = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. The model function η may be linear or nonlinear with respect to both $\boldsymbol{\theta}$ and $\mathbf{x}$.

An experimental design, denoted by ξ, is a set of conditions $\mathbf{x}_i$, $i = 1, \ldots, t$, together with the numbers of replications, r_i, to be taken at each $\mathbf{x}_i$. Formally, we write

$$\xi = \left\{ \begin{array}{c} \mathbf{x}_1, \ldots, \mathbf{x}_t \\ w_1, \ldots, w_t \end{array} \right\},$$

where $w_i = r_i/n$ is the proportion of the observations at $\mathbf{x}_i$ and $\sum_{i=1}^{t} r_i = n$. Such designs are called discrete or exact. The $\mathbf{x}_i$ are the points of support of ξ and w_i are the weights. The information matrix of an exact design ξ for the parameters $\boldsymbol{\theta}$ is

$$n\mathbf{M} = n \sum_{i=1}^{t} w_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i),$$

where

$$\mathbf{f}(\mathbf{x}_i) = \left(\frac{\partial \eta(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_1}, \ldots, \frac{\partial \eta(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_p} \right)^T$$

is the vector of partial derivatives of η with respect to the parameters θ_j, $j = 1, \ldots, p$.

In developing the theory we consider continuous designs, which have the same form, except that the w_i may take any value from $(0, 1\rangle$ such that $\sum_{i=1}^{t} w_i = 1$. The information matrix of a continuous design is correspondingly

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} \mathbf{f}(\mathbf{x})\mathbf{f}^T(\mathbf{x})\xi(dx).$$

In practice, when only n observations can be taken, an exact design will be required. Often the optimum continuous design is approximated by a design with the number of trials at $\mathbf{x}_i$ equal to the integer closest to nw_i.

If the function η is linear with respect to the parameters then the partial derivatives are functions of the experimental conditions only. For a nonlinear model, $\mathbf{M}$ also depends on the parameters and some initial values of the parameters are necessary in order to proceed with the design problem. In either case, the experimenter often has some prior knowledge about the parameters, or has some observations, perhaps from

previous experiments, which can be used to design further experiments in the most efficient manner.

Let the prior information be in the form of an information matrix $n_0\mathbf{M}_0$, where n_0 is the number of trials in the prior experiment. The new information from the experiment comes from an n-trial design ξ and information matrix $n\mathbf{M}(\xi)$. Combining prior and experimental information yields the information matrix

$$\widetilde{\mathbf{M}}(\xi) = n_0\mathbf{M}_0 + n\mathbf{M}(\xi),$$

which is the argument of the Bayesian D-optimality criterion. The criterion arises from a utility function based on the expected Shannon information of the posterior distribution of the parameter estimators, c.f. Chaloner and Verdinelli (1995). Maximizing the utility function reduces to maximizing the functional

$$\Phi[\widetilde{\mathbf{M}}(\xi)] = \int_{\Theta} \log\det\widetilde{\mathbf{M}}(\xi,\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta},$$

where $p(\boldsymbol{\theta})$ denotes the prior density function of $\boldsymbol{\theta}$. We consider only a point prior $\boldsymbol{\theta}^0$, so the criterion simplifies to

$$\Phi[\widetilde{\mathbf{M}}(\xi)] = \det\widetilde{\mathbf{M}}(\xi,\boldsymbol{\theta}^0).$$

This form of the criterion is used, for example by Pukelsheim (1993), in Bayesian optimum designs in linear models. For nonlinear models, the resulting designs are only locally optimum.

The D-efficiency of a design ξ relative to the optimum design is defined as

$$D_{\text{eff}} = \{|\widetilde{\mathbf{M}}(\xi)|/|\widetilde{\mathbf{M}}(\xi^*)|\}^{1/p}, \tag{1}$$

where raising the ratio of determinants to the power $1/p$ measures efficiency in terms of variance.

2.2. Equivalence Theorem condition

The Equivalence Theorem for D-optimum designs states that the designs are also G-optimum, that is they minimise the maximum variance of the predicted response over $\mathcal{X}$. This variance of prediction at any point of $\mathcal{X}$ is less than or equal to the number of parameters in the model and equality is obtained at the support points of the optimum design. The theorem has been extended to wider classes of optimality criteria including Bayesian designs. A thorough investigation of the theorem is given by Pukelsheim (1993). Other references include Whittle (1973), Silvey (1980), Pilz (1983), Chaloner (1984), Atkinson and Donev (1992), Fedorov and Hackl (1997). The theory is based on the convexity of the optimality criteria and holds only for

continuous designs. The condition on the variance of prediction given by the theorem is helpful in the search for optimum designs, a point sometimes overlooked in the literature, where the emphasis is rather on the use of the theorem to demonstrate the optimality of a design.

In our algorithm we use the normalized information matrix for a continuous design ξ

$$\mathbf{M}_\alpha(\xi) = (1-\alpha)\mathbf{M}_0 + \alpha\mathbf{M}(\xi), \tag{2}$$

where

$$\alpha = \frac{n}{n_0+n} \qquad \text{and} \qquad 1-\alpha = \frac{n_0}{n_0+n}.$$

Maximizing the functional $\Phi(\mathbf{M}_\alpha)$ for given α is equivalent to maximizing $\Phi(\widetilde{\mathbf{M}})$ for given n_0 and n.

Now, we can formally state the design problem: find a design ξ from a class of feasible designs, Ξ, maximizing a functional $\Phi : \mathcal{M}_\alpha \longrightarrow \mathcal{R}$, that is solve

$$\max_{\xi\in\Xi} \Phi\left\{\mathbf{M}_\alpha(\xi,\boldsymbol{\theta}^0)\right\}$$

subject to

$$\mathbf{M}_\alpha(\xi,\boldsymbol{\theta}^0) \in \mathcal{M}_\alpha = \left\{(1-\alpha)\mathbf{M}_0 + \alpha\mathbf{M}(\xi,\boldsymbol{\theta}^0), \quad \mathbf{M}(\xi,\boldsymbol{\theta}^0) \in \mathcal{M}(\Xi)\right\},$$

where $\mathcal{M}(\Xi)$ is the set of all information matrices for designs $\xi \in \Xi$.

By Theorem 11.6 and Lemma 6.16 of Pukelsheim (1993), we get the following result

THEOREM 1. *Assume that the prior information matrix* $\mathbf{M}_0$ *and a coefficient* $\alpha \in (0,1)$ *are given. Let the information matrix* $\mathbf{M}^* \in \mathcal{M}(\Xi)$ *be positive definite, that is, the information matrix* $\mathbf{M}^*_\alpha = (1-\alpha)\mathbf{M}_0 + \alpha\mathbf{M}^*$ *is feasible for estimating* $\boldsymbol{\theta}$. *Then* $\mathbf{M}^*$ *is D-optimal for* $\boldsymbol{\theta}$ *in* $\mathcal{M}(\Xi)$ *if and only if*

$$\mathbf{f}^T(\mathbf{x})\left(\mathbf{M}^*_\alpha\right)^{-1}\mathbf{f}(\mathbf{x}) \le tr\left\{\mathbf{M}^*\left(\mathbf{M}^*_\alpha\right)^{-1}\right\} \tag{3}$$

for all $\mathbf{x} \in \mathcal{X}$. *Equality obtains if for* $\mathbf{x}$ *we insert any support point* $\mathbf{x}'$ *of any design* ξ' *that is D-optimal for* $\boldsymbol{\theta}$ *in* $\mathcal{M}(\Xi)$.

As a consequence of this Theorem, we get the following inequality

$$d(\mathbf{x},\xi^*) \le p, \tag{4}$$

where p is the number of the model parameters (the dimension of the matrix $\mathbf{M}_\alpha$) and

$$d(\mathbf{x},\xi^*) = \alpha\mathbf{f}^T(\mathbf{x})(\mathbf{M}^*_\alpha)^{-1}\mathbf{f}(\mathbf{x}) + (1-\alpha)\mathrm{tr}\left\{\mathbf{M}_0(\mathbf{M}^*_\alpha)^{-1}\right\},$$

with the maxima again at the points of support of the optimum design.

A proof of the Theorem and the derivation of inequality (4) are given in the Appendix.

This result is very useful in searching for an optimum design when some prior information is given. In the next section we give the appropriate form of $d(\mathbf{x}, \xi)$ for design augmentation. In Section 4 we derive the special form when interest is in parsimonious model checking. But first we describe the use of the equivalence theorem in finding and checking optimum designs.

2.3. Computational procedure

To find the optimum design we need to find the t support points $\mathbf{x}_i$, together with their associated weights w_i. We also need to find the value of t, which is often, but not always, equal to the number of parameters p. Since the weights sum to one, the optimum design criterion Φ is a function of $2t-1$ variables. We can find the points at which the criterion attains its maximum either by a numerical optimization over all $2t-1$ variables at once, or by a two-phase procedure in which optimum weights are found for a given set of support points. In either case we use the equivalence theorem to confirm that the design is optimum, checking both that $d(\mathbf{x}, \xi) = p$ at the support points and also that the maximum value of $d(\mathbf{x}, \xi)$ over $\mathcal{X}$ is not greater than p.

In the two-phase algorithm we find the optimum weights for a given set of t support points, using information from the equivalence theorem to modify the support points and, perhaps, change their number. The search in $2t-1$ dimensions is thus reduced to a sequence of searches in $t-1$ dimensions. We use a grid of points in $\mathcal{X}$ both to provide a set of candidate points for the design and for checking the maximum of $d(\mathbf{x}, \xi)$. The optimum design weights are found for the set of support points forming a candidate design. We remove from the design any support points with zero weight. If the maximum of $d(\mathbf{x}, \xi)$ over the grid of $\mathcal{X}$ equals p the algorithm has converged and we have an optimum design. Otherwise there will be one or more points for which the value of $d(\mathbf{x}, \xi)$ is greater than p. We add the candidate point corresponding to the maximum value of $d(\mathbf{x}, \xi)$. In this way a new candidate design is found which may have a new value of t. The optimization of the weights is repeated and points added or deleted until convergence is obtained, which usually occurs in two or three iterations. Constraining the search to a grid of points both simplifies the search over $\mathcal{X}$ and results in experiments which are easier to perform than those where the $\mathbf{x}$ values can be any m-dimensional real vector within a given range.

For the numerical optimization we used the Fortran version of the conjugate gradients algorithm `frprmn` of Press et al. (1992, p. 416). If rounding of the weights of the continuous design does not give a satisfactory exact design, numerical methods for the construction of exact designs such as DETMAX (Mitchell 1974) have to be used (see Atkinson and Donev 1992, Chapter 15).

3. Design augmentation

3.1. Equivalence theory

We first consider the augmentation of a design in which there are n_0 prior observations with information matrix $n_0\mathbf{M}_0$. If the prior design is

$$\xi_0 = \left\{ \begin{array}{ccc} \mathbf{x}_1^0 & \dots & \mathbf{x}_q^0 \\ w_1^0 & \dots & w_q^0 \end{array} \right\} \tag{5}$$

then

$$\mathbf{M}_0 = \sum_{i=1}^{q} w_i^0 \mathbf{f}(\mathbf{x}_i^0)\mathbf{f}^T(\mathbf{x}_i^0). \tag{6}$$

D-optimum designs for design augmentation are described by Dykstra (1971b), who considers the sequential addition of one trial at a time. He later comments (Dykstra 1971a) that this method is equivalent to the sequential algorithm for the construction of D-optimum designs of Wynn (1970). Such designs are optimum as $n \to \infty$ but may not be optimum for small n. An example is noted in Section 3.2. Evans (1979) finds exact optimum designs for augmentation with specified n. Heiberger et al. (1993) also calculate exact optimum augmentation designs, but for a large and flexible family of criteria, which includes D-optimality. Since these papers describe exact designs, they do not use Equivalence theory to explore the properties of the designs.

We obtain a useful re-expression of the condition for the Equivalence Theorem by substituting $\mathbf{M}_0$ from (6) in (4)

$$d(\mathbf{x}, \xi^*) = \alpha \mathbf{f}^T(\mathbf{x})(\mathbf{M}_\alpha^*)^{-1}\mathbf{f}(\mathbf{x}) + (1-\alpha)\sum_{i=1}^{q} w_i^0 \mathbf{f}^T(\mathbf{x}_i^0)(\mathbf{M}_\alpha^*)^{-1}\mathbf{f}(\mathbf{x}_i^0) \leq p. \tag{7}$$

This condition for the optimum design has an informative statistical interpretation; it can be written as

$$d(\mathbf{x}, \xi^*) = (n/\sigma^2)\ \text{var}\ \{\hat{y}(\mathbf{x})\} + (n_0/\sigma^2)\sum_{i=1}^{q} w_i^0\ \text{var}\ \{\hat{y}(\mathbf{x}_i^0)\} \leq p.$$

The first variance term is the posterior variance at a point in $\mathcal{X}$ and the second a weighted sum of posterior variances at the points of the prior design. If the initial design is D-optimum, the standardized posterior variances in (7) are all equal to p, so that the optimum augmentation design is the D-optimum design in the absence of prior information, that is a replicate of the prior design. Usually this will not be the case. As we show, the augmentation design can be very different from the D-optimum design found in the absence of prior information.

3.2. Example 1. Second order response surface: regular design region

As a first example we consider designs for the second-order polynomial in two variables

$$E(y) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 + \theta_5 x_1^2 + \theta_6 x_2^2 \tag{8}$$

over the square design region $\mathcal{X} = \{-1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1\}$. The D-optimum design for this model and region, when no prior information is available ($\alpha = 1$), is supported on the points of the 3^2 factorial with weights as given in the penultimate column of Table 1. The good integer approximation to this design in the last column of the table is one replicate of the full factorial with one extra replicate of each corner point, making 13 trials in all.

Table 1. Response surface – regular design region: augmentation of the second-order design of Fig. 1 ($n_o = 6$). Continuous optimum designs and integer approximations $[nw_i]$ are given for each value of α, apart from $\alpha = 13/19$. The last two columns give the optimum design for a second-order response surface and an integer approximation.

x_1	x_2	$\alpha = 1/3$		$\alpha = 3/5$		$\alpha = 13/19$	$\alpha = 19/25$		$\alpha = 1$	
		w	$[3w]$	w	$[9w]$	w	w	$[19w]$	w	$[13w]$
−1	−1	–	–	0.0822	1	0.1074	0.1221	2	0.1458	2
0	−1	–	–	–	–	–	0.0005	0	0.0802	1
1	−1	0.2897	1	0.2270	2	0.2092	0.1934	4	0.1458	2
1	−0.1	0.0565	0	–	–	–	–	–	–	–
−1	0	–	–	0.0200	0	0.0463	0.0591	1	0.0802	1
0	0	–	–	–	–	0.0038	0.0364	0	0.0962	1
1	0	–	–	0.1142	1	0.1123	0.1036	2	0.0802	1
−1	1	0.2840	1	0.2226	2	0.2031	0.1866	4	0.1458	2
−0.1	1	0.0670	0	–	–	–	–	–	–	–
0	1	–	–	0.1089	1	0.1129	0.1102	2	0.0802	1
1	1	0.3027	1	0.2250	2	0.2049	0.1881	4	0.1458	2

Our six-trial starting design is far from this design. The six points, shown by circles in Fig. 1 are all in the lower left-hand quarter of the design region: no values of x_1 or of x_2 are greater than zero.

It is clear that any scheme for design augmentation will extend the design over the whole region. Since there are six points, the second-order model can be fitted. The contours of the variance function of the prediction are also plotted in Fig. 1. The maximum of 1,542 is at (1,1), the corner of the region most remote from the initial design. The augmentation procedure of Dykstra (1971b), equivalent to the sequential construction of the D-optimum design one trial at a time, would add a trial at this point. The next point to be added would be at (1, –0.9), not a point of the optimum second-order design. Such perpetuation of distortions introduced by the initial design is one of the drawbacks of the sequential approach. We instead

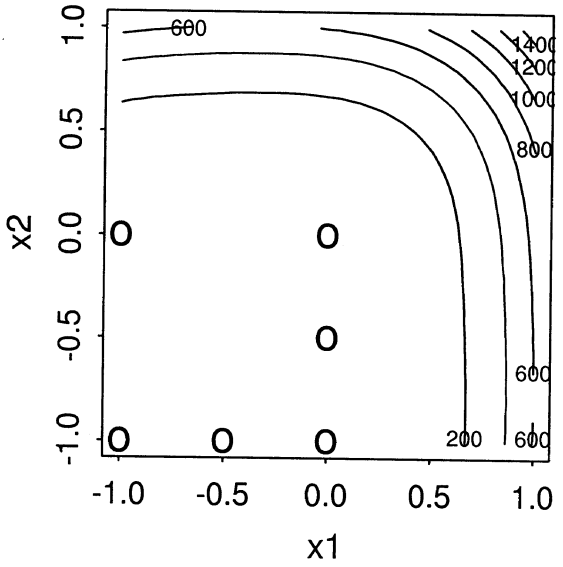

Figure 1. Regular design region: the six points of the second-order design which is to be augmented, with contours of the standardized variance $d(x, \xi)$.

find the optimum continuous design for a specified α and then calculate exact designs from approximations to our optimum continuous measures. In our response surface examples we search over grids of size 0.1 in x_1 and x_2.

Some augmentation designs are given in Table 1, together with integer approximations. For $\alpha = 1/3$ the optimum measure has five points of support. The contours of the variance function in Fig. 2 show maximum values of 6 at these five points, the corners most remote from the initial design, and at points near the centres of the remote sides of the region. The continuous design is indeed optimum. The weights on the three corners of the region at which there were no prior experiments account for 88% of the total, and are nearly equal. Since $\alpha = n/(n_0 + n)$, when $n_0 = 6$ and $\alpha = 1/3$ we require a design for $n = 3$. A good integer approximation to the three-trial design is to put one trial at each of the three corners of the region. This integer approximation is given in the fourth column of Table 1 with the number of replicates $r_i = [3w_i]$, the integer closest to $3w_i$. These three points are also plotted in Fig. 2. Since the three-trial integer approximation is not the optimum continuous design, the variance for the nine-trial design at the two points with small weights in column 3 of the table is greater than 6. In calculating this variance the form (7) was used, since we do have a prior design.

The other columns of the table give continuous designs for $\alpha = 3/5$, 13/19, and 19/25 and exact designs with $r_i = [nw_i]$. The continuous designs have respectively seven, eight and nine points of support, all at the points of the 3^2 factorial. For

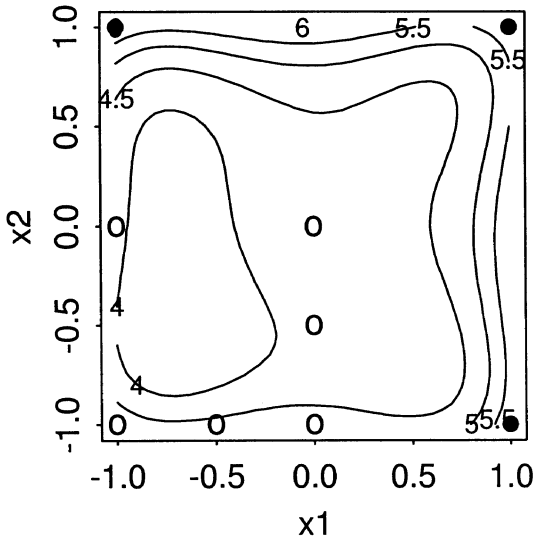

Figure 2. Regular design region: augmentation of the second-order design. Contours of the standardized variance $d(x, \xi^*)$ for the optimum design for $n = 3$ of Table 1.

$\alpha = 3/5$ a good integer approximation has six unequally replicated points of support whereas, for $\alpha = 19/25$, the discrete design has seven support points. The results for $\alpha = 13/19$ are included to emphasise that good integer approximations cannot be found for all α. Algorithms for exact designs may sometimes be necessary.

Also included in Table 1 is the design for $\alpha = 1$, the optimum design in the absence of an initial experiment. The design weights in Table 1 for the points of the 3^2 factorial show a smooth progression from $\alpha = 1/3$ to $\alpha = 1$. The weights for the three corner points included when $\alpha = 1/3$ decrease steadily to the final values while the other weights increase with α once their support points have been included in the design.

4. Parsimonious model checking

4.1. Formulation of prior information

Suppose that the terms in the linear model (or a linear approximation of a non-linear model) can be partitioned into two groups

$$E(y) = \boldsymbol{\theta}_r^T \mathbf{f}_r(\mathbf{x}) + \boldsymbol{\theta}_s^T \mathbf{f}_s(\mathbf{x}), \tag{9}$$

where $\boldsymbol{\theta}_r$ is a vector of r primary parameters and $\boldsymbol{\theta}_s$ is a vector of s secondary

parameters. Accordingly, $\mathbf{f}_r(\mathbf{x})$ and $\mathbf{f}_s(\mathbf{x})$ denote vectors of functions of experimental conditions. The terms in $\mathbf{f}_r(\mathbf{x})$ are those that it is believed are required in the model, for example first-order terms in a polynomial model. However as well as designing the experiment to estimate $\boldsymbol{\theta}_r$ it is also required to check that none of the terms of $\mathbf{f}_s(\mathbf{x})$ are required. These will typically be higher-order polynomial terms. The parsimonious design problem is to find a design which allows such checking without requiring estimation of all $r + s$ parameters.

DuMouchel and Jones (1994) suggest finding D-optimum designs for the model (9) when there is no prior information about the primary parameters, but there is some about the secondary ones. The effect of prior information on the secondary parameters is to concentrate experimental effort on the primary ones. This procedure is a special case of Bayesian D-optimality.

The absence of prior information about the primary parameters $\boldsymbol{\theta}_r$ is represented by using a diffuse prior distribution. Let

$$\boldsymbol{\theta}_r \sim \mathcal{N}_r(\boldsymbol{\theta}_r^0, \gamma^2 \mathbf{I}_r),$$

where $\gamma^2 \longrightarrow \infty$. However, we assume that there is some prior information for the secondary parameters, which have distribution

$$\boldsymbol{\theta}_s \sim \mathcal{N}_s(\mathbf{0}_s, \tau^2 \mathbf{I}_s),$$

where τ^2 is a small positive value. Then the prior distribution of all the $p = r + s$ parameters is

$$\boldsymbol{\theta} \sim \mathcal{N}_p\left(\begin{bmatrix} \boldsymbol{\theta}_r^0 \\ \mathbf{0}_s \end{bmatrix}, \begin{bmatrix} \gamma^2 \mathbf{I}_r & \mathbf{0}_{r\times s} \\ \mathbf{0}_{s\times r} & \tau^2 \mathbf{I}_s \end{bmatrix}\right)$$

For designing experiments we require the prior information matrix for $\boldsymbol{\theta}$, that is the inverse of the dispersion matrix. So when $\gamma^2 \longrightarrow \infty$

$$[D(\boldsymbol{\theta})]^{-1} \longrightarrow \frac{1}{\tau^2}\mathbf{K},$$

where

$$\mathbf{K} = \begin{pmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times s} \\ \mathbf{0}_{s\times r} & \mathbf{I}_s \end{pmatrix}.$$

Hence the posterior information matrix (Lee, 1997 or Pukelsheim, 1993) for $\boldsymbol{\theta}$, given ξ, is

$$\widetilde{\mathbf{M}} = n_0\mathbf{K} + n\mathbf{M},$$

where $n_0 = \sigma^2/\tau^2$. As we did for design augmentation, we calculate optimum designs using the normalized form of the information matrix $\mathbf{M}_\alpha = (1-\alpha)\mathbf{K} + \alpha\mathbf{M}$. But now α can be expressed in terms of n, τ^2 and σ^2. We assume that the variance of the

observations, σ^2, is constant, so increasing values of τ^2, which mean less precise prior information about the secondary parameters, leads to larger values of α. As $\alpha \to 1$ the design tends to the D-optimum design when the model with all $r+s$ parameters is of interest. Conversely, decreasing α implies more prior knowledge about the secondary terms. As $\alpha \to 0$ the design tends to the D-optimum design for the model with just r parameters. The equivalence theorem of §2.2 holds with $p = r+s$ and $\mathbf{M}_0 = \mathbf{K}$.

We can again find a simple expression for the function $d(\mathbf{x}, \xi^*)$ in (4). Let

$$(\mathbf{M}_\alpha^*)^{-1} = \{(m_\alpha^*)^{ij}\}_{i,j=1,\ldots,r+s},$$

when

$$\mathrm{tr}\mathbf{K}(\mathbf{M}_\alpha^*)^{-1} = \sum_{j=r+1}^{r+s} (m_\alpha^*)^{jj}$$

and

$$d(\mathbf{x}, \xi^*) = \alpha \mathbf{f}^T(\mathbf{x}) (\mathbf{M}_\alpha^*)^{-1} \mathbf{f}(\mathbf{x}) + (1-\alpha) \sum_{j=r+1}^{r+s} (m_\alpha^*)^{jj} \le r+s, \tag{10}$$

which is easily calculated.

This concludes our theoretical development of the relationship between the procedure of DuMouchel and Jones (1994) and equivalence theory for Bayesian D-optimality. However some comments on the prior specification are appropriate. The first is that since information is available on only one group of terms, the secondary terms should be orthogonal to the primary ones, in order to keep the information separate. This orthogonality is achieved by regressing each secondary term on the primary terms. The regression is over a list of candidate points in $\mathcal{X}$. The residuals of the secondary terms are then scaled and used in the construction of the design. The details are in §6.2. Since the design uses only a few of the candidate points, it might be more logical to perform weighted regression over the support points of the design, using the design weights. The residuals then have to be updated for each design. Atkinson and Zocchi (2000) find little difference between the designs obtained from such a procedure and those from regressing over the candidate points, which is the procedure followed here.

DuMouchel and Jones also suggest two scalings. The first is that of the primary terms. However Atkinson and Zocchi (2000) prove that the D-optimum designs are invariant to this scaling. The second scaling is of the residuals of the secondary terms after regression on the primary terms, so that they all have the same range, or the same variance. If the scales of the residuals are very different, interpretation of τ^2 as a common variance is strained. In our response surface examples the range of the variables is similar and we do not make this adjustment, calculating designs using the unscaled residuals of the $\mathbf{f}_s(\mathbf{x})$.

4.2. Example 1 continued. Second-order response surface: regular design region

The simplest problem in checking the model is to treat the three second-order terms (the two quadratic terms and the interaction) as secondary. Designs for this problem have a very simple structure. For small α, that is large prior information, the 2^2 factorial is optimum – since the prior is overwhelming, there is no need for experimental evidence to check the model. For larger α, the optimum design is the 2^2 factorial with some weight on a centre point. For example, if $\alpha = 2/3$ the optimum design has weight 0.2343 on each factorial point and the remaining 0.0629 on the centre point. The contours of the variance function for this design are plotted in Fig. 3. This has a maximum of six at the five design points. But the plot also shows local maxima at the centres of the sides of the region. These are the four remaining points of the 3^2 factorial, at which $d(\mathbf{x}, \xi^*) = 5.843$. As α increases slightly to $8/11 = 0.7273$ this design is no longer optimum, since the variances at the centres of the sides are now 6.025: these points should be included in the design. In calculating this variance, the form (10) was used.

These designs have one expected feature – that the points of the 2^2 factorial plus centre point can be optimum for checking the first order model against the second. But it is surprising that such a small weight goes on the centre point. For $\alpha = 0.7188$

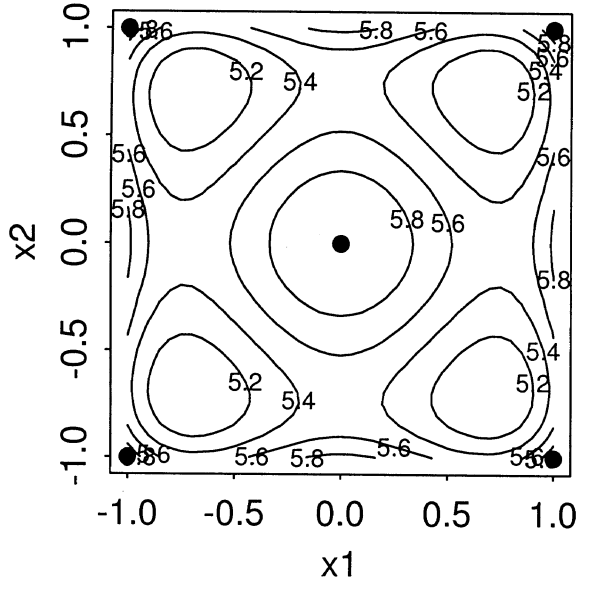

Figure 3. Regular design region: design of Section 4.2 for checking the first-order model for $\alpha = 2/3$ together with contours of the standardized variance $d(x, \xi^*)$

the ratio of factorial weights to that for the centre point is $0.2276/0.0894 = 2.546$. For slightly higher values of α the five-point design is not optimal. This is very different from the customary advice on this problem where the factorial might be augmented by two or three centre points, giving a ratio of design weights of 0.5 or less. However such advice may also incorporate the desire to obtain a rough estimate of the error variance.

5. Example 2. Second-order response surface: constrained design region

5.1. Background

Often, in chemical or biological experiments, high levels of all factors can lead to conditions which are so severe that inorganic molecules decompose or plants wither. Several examples are mentioned in Chapter 16 of Atkinson and Donev (1992). Avoidance of such conditions leads to a constrained design region. Since some of the structure of the preceding designs depended on the symmetries of the design region, we now consider a modified version of the problem in which constraints make the design region less regular. With x_1 and x_2 both scaled so that $-1 \leq x_i \leq 1$, $i = 1, 2$, we add the linear constraints that

$$2x_1 + x_2 \leq 1$$

$$x_1 + x_2 \geq 1$$

$$x_2 - x_1 \leq 1.5.$$

The resulting irregularly hexagonal design region $\mathcal{X}$ is shown in Fig. 4.

We are again concerned with designs for checking models and for the augmentation of existing designs. To get a feel for the variety of designs which might be encountered we calculate the D-optimum designs for the first- and second-order models when no prior information is available. The resulting designs are given in Table 2 and Fig. 4. The first-order design has three points of support with weights 1/3. To check the optimality of the design, the variance of the prediction was calculated over the hexagonal region. Not only was the value 3 at the design points, it was also 3 at (0,–1), but this point had a weight of zero and so was not included in the optimum design.

The optimum design for the second-order model has eight points of support – Table 2 and Fig. 4(b). In order to give some visual impression of the design weights, the diameter of the dots in Fig. 4 (and also in Fig. 5) is proportional to $w_i^{0.8}$. The weight on (–0.5, 1) is a negligible 0.0061, and so is not visible in the figure. If this

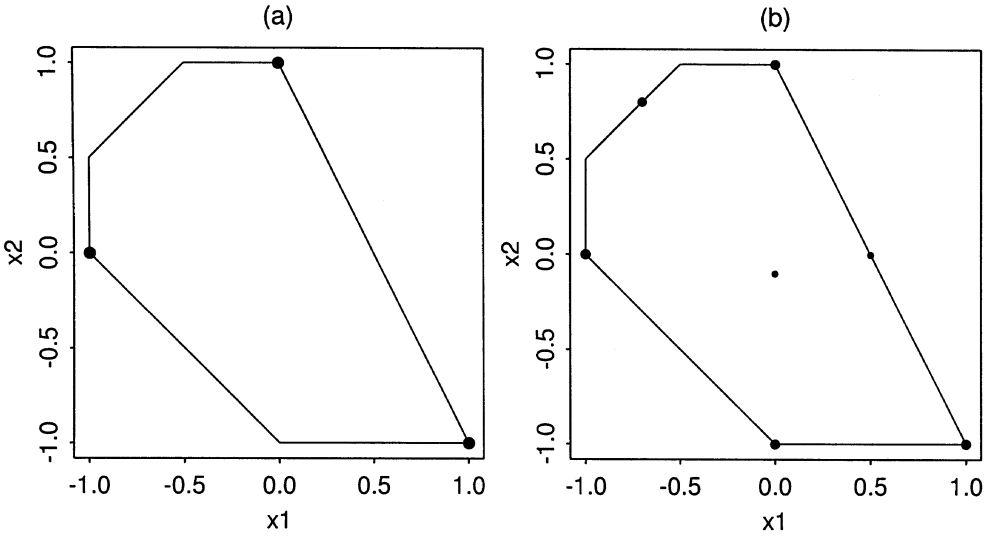

Figure 4. Constrained design region: optimum designs of Table 2: (a) First-order ($\alpha = 0$) and (b) Second-order ($\alpha = 1$). Dot diameter $\propto w_i^{0.8}$.

Table 2. Response surface – constrained design region: optimum first-order and second-order designs.

		First	Second-Order Designs	
x_1	x_2	Order	8 points	Exact: $n = 19$
		w	w	$[19w]$
0	−1	–	0.1595	3
1	−1	1/3	0.1626	3
0	−0.1	–	0.1010	2
−1	0	1/3	0.1648	3
0.5	0	–	0.1010	2
−0.7	0.8	–	0.1479	3
−0.5	1	–	0.0061	0
0	1	1/3	0.1571	3
D–efficiencies			100%	99.95%

point is dropped, the seven-point design for the six parameter model has weights in the approximate ratio 3:2 and can be well approximated by the 19 trial integer design of Table 2. The D-efficiency (1) of this exact design, relative to the eight-point design, is 99.95%.

5.2. Model checking

We now consider designs for checking first-order design which require fewer than seven different sets of design conditions. When α is small, so that much of the information is coming from the prior, the first-order design is augmented by extra trials at the point (0, −1), which we have already noticed as having a high variance for the first-order design. Designs with these four points of support are optimum up to $\alpha = 0.318$. The weights are given in Table 3. For this value of α, the largest value of $d(\mathbf{x}, \xi^*)$ at a point not in the design is 5.9923 at $(-1, 0.5)$. As α increases, this point is included in the design which now has five points. These design points are optimum, with weights changing with α, until $\alpha = 0.531$ when the maximum, apart from points in the design, $d\{(0.5, 0), \xi^*\} = 5.9871$. Unlike (−1, 0.5), (0.5, 0) is one of the points of support of the optimum second-order design. Four model-checking designs are shown in Table 3 and in Fig. 5. Each one arises from a value of α at which another support point almost needs to be added to the design measure. For each design the point about to be entered is given in the table, along with the value of $d(\mathbf{x}, \xi^*)$, which is just less than six. The ability to determine which points will be included in the design if α increases modestly is a further useful application of the equivalence theorem. There is also a steady change in the design weights as α increases. Those for the three points of the first-order design decrease from their initial values of 1/3, whereas that for (−1, 0.5) increases and then decreases. In the limit for large α this point will not be present in the design.

It is interesting to continue the process of finding optimum designs for model checking for larger values of α. However, the design that is of most practical importance

Table 3. Response surface – constrained design region: parsimonious designs for checking first-order model, showing the number of support points of the design increasing with α. Designs are given for values of α such that further increase will augment the design by $\mathbf{x}_{\text{next}}$.

x_1	x_2	$\alpha = 0.318$ w	$[10w]$	$\alpha = 0.531$ w	$\alpha = 0.647$ w	$\alpha = 0.949$ w
0	−1	0.1105	1	0.1691	0.1849	0.1893
1	−1	0.2871	3	0.2625	0.2344	0.1932
−1	0	0.2871	3	0.1893	0.1522	0.1403
0.5	0	–	–	–	0.0659	0.1660
−1	0.5	–	–	0.0919	0.1223	0.0786
−0.5	1	–	–	–	–	0.0989
0	1	0.3153	3	0.2871	0.2468	0.1337
Next to enter		(−1, 0.5)		(0.5, 0)	(−0.5, 1)	(0, −0.1)
$d(\mathbf{x}_{\text{next}}, \xi^*)$		5.992		5.987	5.970	5.991

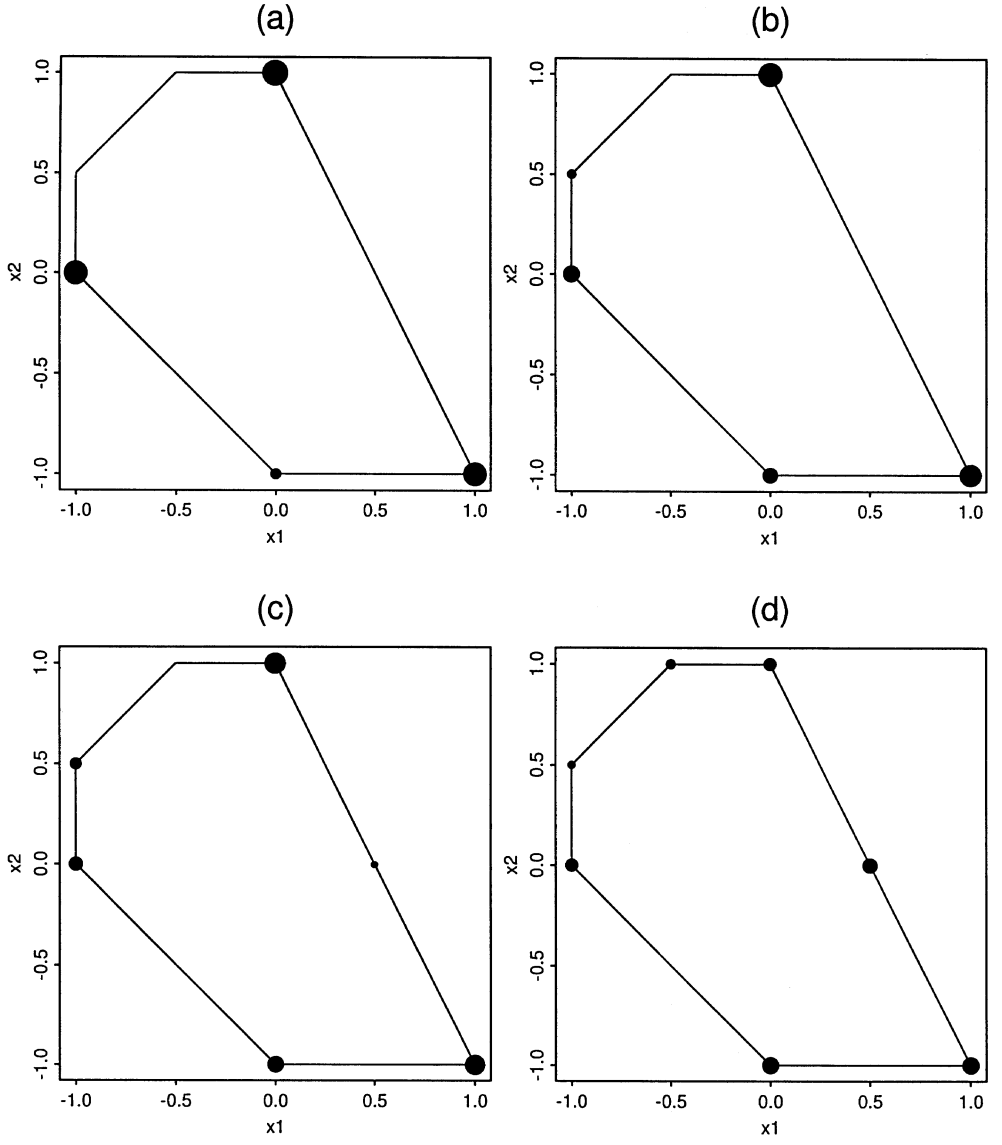

Figure 5. Constrained design region. Designs for model checking of Table 3: (a) $\alpha = 0.318$, (b) $\alpha = 0.531$, (c) $\alpha = 0.647$ and (d) $\alpha = 0.949$. Dot diameter $\propto w_i^{0.8}$.

is the four-point design. An intriguing feature of this design is that all four points are on the edge of the design region, a design which is unlikely to be found by unaided intuition. The designs we have found for large α are optimum for the model-checking criterion, but are no more parsimonious than the seven-point design for the second-order model. Continuing the process of design construction will of course finally lead to the optimum design for the second-order model.

5.3. *Design augmentation*

We now use the theory of Section 3 for design augmentation starting from a design for model checking, a design strategy which would be necessary if the simple model were found not to hold. A good approximation to the design for model checking with four support points is given in Tables 3 and 4. This ten-point design has three trials at each of the points of the first-order design of Table 2 and one at the fourth point. Suppose that, as a result of this experiment, it seems that a second-order model is needed and so the design is to be augmented. Suppose also that a further five trials are required. If we take as the support points those of the seven-point optimum design of Table 2, we find that trials are required at only four support points, those that are not part of the first-order design. The resulting design is in Table 4 and Fig. 6. That this design is indeed optimum is checked by calculating $d(\mathbf{x}, \xi^*)$, which is 6 at the design points and less elsewhere. The five trials can be approximated by weights of 1,1,1 and 2 as shown in the table, a design with a D-efficiency of 97.224% relative to the eight-point design for the second-order model of Table 2.

If, instead, nine trials are to be added, the weights are as shown in Table 4. Now the weights on the three points of the first-order design are close to zero, with a maximum

Table 4. Response surface – constrained design region: augmentation of parsimonious design for checking first-order model.

x_1	x_2	Check 1st order	$\alpha = 0.333$			$\alpha = 0.474$		
			w	$[5w]$	Total	w	$[9w]$	Total
0	−1	1	0.1757	1	2	0.2255	2	3
1	−1	3	–	–	3	0.0098	0	3
0	−0.1	–	0.2309	1	1	0.2121	2	2
−1	0	3	–	–	3	0.0146	0	3
0.5	0	–	0.2335	1	1	0.2142	2	2
−0.7	0.8	–	0.3599	2	2	0.3227	3	3
−0.5	1	–	–	–	–	–	–	–
0	1	3	–	–	3	0.0011	0	3
D-efficiencies					97.224%			99.95%

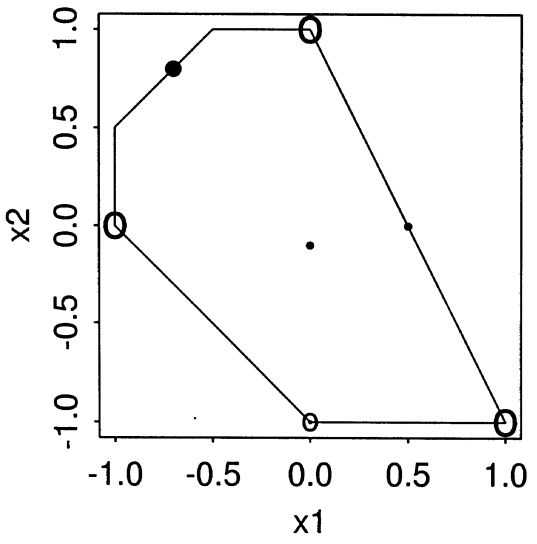

Figure 6. Exact augmentation design of Table 4 for $n = 5$: o original points, • augmentation.

of 1.46%. An integer approximation to the augmentation design has either two or three trials at each of the other points of support. The combination of this design with the initial design gives the approximation to the second-order design of Table 2. A final example of the usefulness of the equivalence theorem is that the continuous design found in the table is not quite D-optimum: $d\{(-0.5, 1), \xi^*\} = 6.0292$, a further reminder that the optimum design of Table 2 has eight, not seven, points of support.

6. Example 3. A nonlinear model for crop yield and plant density

6.1. Background

Models quantifying the relationship between crop yield and plant density are of great importance in agriculture. The review of Wiley and Heath (1969) lists the principal models. Mead (1970) examines the theoretical properties of a family of inverse power response curves and illustrates methods for fitting these models using two examples. A comparison of the models in Wiley and Heath (1969) is given by Mead (1979). Gillis and Ratkovsky (1978) and Ratkovsky (1989) investigate yield-density models using simulation techniques. Seber and Wild (1989, Section 7.6) present a compact review of the subject.

Plants compete for resources. The yield per plant therefore tends to decrease as plant density increases although the yield per unit area often continues to increase until appreciably higher plant densities are reached. The yield then either reaches an

asymptote or slowly decreases. Because of this behaviour nonlinear models are often used.

As an example we use data on the yields of soya beans grown in regularly spaced rows – there is a regular distance both between the rows and between plants within the rows. Although the simplest models assume that it is only the area per plant that matters, rather than the shape of the area for each plant, we want a general model which allows for different effects of the two distances, which are of different magnitudes. The equivalent of the general second-order polynomial response surface model of the earlier sections is the general seven-parameter model

$$E(y) = \left(\theta_1 + \theta_2 \frac{1}{x_1} + \theta_3 \frac{1}{x_2} + \theta_4 \frac{1}{x_1 x_2} + \theta_5 \frac{1}{x_1^2} + \theta_6 \frac{1}{x_2^2}\right)^{-\frac{1}{\theta_7}} \tag{11}$$

where:

$E(y)$ is the expected yield per plant or biologically definable part of the plant;

$\theta_1, \ldots, \theta_7$ are positive parameters and $0 < \theta_7 \leq 1$;

x_1 is the intra-row spacing, i.e. the spacing between plants within a row and

x_2 is the inter-row spacing, i.e. the spacing between rows.

The area per plant is therefore $x_1 x_2$ with $1/x_1 x_2$ the density, i.e. the number of plants per unit area.

This general model is a second-order polynomial with terms in $1/x_1$ and $1/x_2$. Although the parameters enter linearly, the model is made nonlinear by the presence of the power θ_7. Instead of (11) we work with the related model for expected yield per unit area

$$E(y^*) = \frac{1}{x_1 x_2}\left(\theta_1 + \theta_2 \frac{1}{x_1} + \theta_3 \frac{1}{x_2} + \theta_4 \frac{1}{x_1 x_2} + \theta_5 \frac{1}{x_1^2} + \theta_6 \frac{1}{x_2^2}\right)^{-\frac{1}{\theta_7}}, \tag{12}$$

obtained by dividing both sides of (11) by $x_1 x_2$.

Several simpler models have been proposed, which are special cases of (12). The three parameter model of Shinozaki and Kira (1956) for expected yield per unit area

$$E(y^*) = \frac{1}{x_1 x_2}\left(\theta_1 + \theta_4 \frac{1}{x_1 x_2}\right)^{-\frac{1}{\theta_7}} \tag{13}$$

is obtained by putting $\theta_2 = \theta_3 = \theta_5 = \theta_6 = 0$ and so depends only on the area per plant, ignoring the shape of that area. A four-parameter model due to Berry (1967) allows for the shape of the area

$$E(y^*) = \frac{1}{x_1 x_2}\left(\theta_1 + \theta_2 \frac{1}{x_1} + \theta_3 \frac{1}{x_2} + \theta_4 \frac{1}{x_1 x_2}\right)^{-\frac{1}{\theta_7}} \tag{14}$$

and can be obtained from (12) by putting $\theta_5 = \theta_6 = 0$. A further model due to Berry is the special case of (14) with $\theta_2 = \theta_3 = 0$, in which x_1 and x_2 are treated symmetrically.

To estimate the parameters in the general model (12) requires experiments at at least seven combinations of x_1 and x_2. As we implied in the introduction, such a design will usually be inefficient if the experimental purpose is to check the three-parameter model (13). We now find an optimum model-checking design which requires trials at only four treatment combinations, rather than seven.

6.2. Parsimonious Model Checking

Since the models we are using are nonlinear, we now need prior estimates of the parameters to be able to design the experiment. For this purpose we use data from Lin and Morse (1975) on the effect of spacing on the yield of soya beans. The data are in Table 5. The four levels of the inter row-spacing factor (0.18, 0.36, 0.54 and 0.72 m) and four levels of the intra-row spacing factor (0.03, 0.06, 0.09 and 0.12 m) were used to study the optimum spacing for maximum yield per unit area. The maximum observed yield is near the centre of the region.

Table 5. Yield-Density relationship: the mean grain yield, in g/m^2, for sixteen spacing treatments of the soya bean variety Altona (from Lin and Morse, 1975).

Inter-row spacing (m)	Intra-row spacing (m)			
	0.03	0.06	0.09	0.12
0.18	260.0	344.7	279.9	309.2
0.36	305.3	358.3	312.2	267.8
0.54	283.9	342.0	269.0	253.9
0.72	221.8	287.9	230.9	196.9

When we tried to fit the simple model (13) to these data, we found that the fit was improved, as judged by residual plots, if we took logarithms of both sides, so that the primary model for expected yield becomes

$$\eta_0(\mathbf{x}, \boldsymbol{\theta}) = E(\log y^*) = -\frac{1}{\theta_7} \log \left(\theta_1 + \theta_4 \frac{1}{x_1 x_2} \right) - \log(x_1 x_2).$$

Since yield cannot be negative, a model such as this, which gives a lognormal distribution for y^*, is more plausible than one with a normal distribution of errors on the untransformed scale. The gamma models suggested by McCullagh and Nelder (1989, p. 291) have a similar justification. An interesting feature of the logged model is that the models for yield per plant and yield per unit area only differ by the subtraction of an offset. The parameter estimates for the logged model were $\hat{\theta}_1 = 0.07469$, $\hat{\theta}_4 = 0.003751$ and $\hat{\theta}_7 = 0.7363$.

The extended seven-parameter logged $\eta(\mathbf{x}, \boldsymbol{\theta})$ model is likewise found by taking the logarithm of (12). The two groups of parameters in Section 4.1 are thus the primary parameters

$$\boldsymbol{\theta}_3 = \begin{bmatrix} \theta_1 & \theta_4 & \theta_7 \end{bmatrix}^T,$$

with prior values

$$\boldsymbol{\theta}_3^0 = \begin{bmatrix} \hat{\theta}_1 & \hat{\theta}_4 & \hat{\theta}_7 \end{bmatrix}^T,$$

and the secondary parameters

$$\boldsymbol{\theta}_4 = \begin{bmatrix} \theta_2 & \theta_3 & \theta_5 & \theta_6 \end{bmatrix}^T,$$

with priors $\boldsymbol{\theta}_4^0$ taken as their expected value zero.

The elements of $\mathbf{f}_3(\mathbf{x})$ and $\mathbf{f}_4(\mathbf{x})$ are the derivatives of $\eta(\mathbf{x}, \boldsymbol{\theta})$ with respect to $\theta_j, j = 1, \ldots, 7$, evaluated at the prior

$$[\boldsymbol{\theta}^0]^T = \begin{bmatrix} (\boldsymbol{\theta}_3^0)^T & (\boldsymbol{\theta}_4^0)^T \end{bmatrix},$$

namely

$$\begin{aligned} f_1(\mathbf{x}) &= -\left\{\hat{\theta}_7 \left(\hat{\theta}_1 + \hat{\theta}_4 \frac{1}{x_1 x_2}\right)\right\}^{-1} \\ f_j(\mathbf{x}) &= z_j f_1(\mathbf{x}) \qquad\qquad j = 2, \ldots, 6. \qquad (15) \\ f_7(\mathbf{x}) &= \hat{\theta}_7^{-2} \log\left(\hat{\theta}_1 + \hat{\theta}_4 \frac{1}{x_1 x_2}\right). \end{aligned}$$

In (15) z_j is the coefficient of θ_j in (12).

The vectors of the derivatives defined in (15) form the rows of the design matrices $\mathbf{F}_r(\xi)$ and $\mathbf{F}_s(\xi)$ of primary and secondary terms. For design region we let $\mathcal{X} = \{(x_1, x_2) : 0.15 \leq x_1 \leq 0.8 \text{ and } 0.03 \leq x_2 \leq 0.2\}$, slightly larger than that used by Lin and Morse (1975). The prior of Section 4 for model checking makes sense if the secondary terms are orthogonal to the primary ones and if the columns of $\mathbf{F}_s(\xi)$ are so scaled that the coefficients $\boldsymbol{\theta}_s$ have a common prior variance τ^2. In the response surface examples of Sections 4 and 5 these conditions were satisfied by the scaling of the variables from -1 to 1. We satisfy these conditions for our nonlinear model by using scaled residuals from regression to provide the necessary orthogonality. For the regression we use a design measure ξ_c which is uniform over the 14×14 set of points:

$$\{0.15, 0.2, \ldots, 0.8\} \times \{0.03, 0.04307, 0.05615, \ldots, 0.2\},$$

the results not being sensitive to the number of points in $\mathcal{X}$ which are used. The steps of the **scaling procedure** are:

1. Perform the regression of the extra terms on the primary terms computing $\mathbf{B} = \{\mathbf{F}_r^T(\xi_c)\mathbf{F}_r(\xi_c)\}^{-1}\{\mathbf{F}_r^T(\xi_c)\mathbf{F}_s(\xi_c)\}$ and the residual matrix $\mathbf{R} = \mathbf{F}_s(\xi_c) - \mathbf{F}_r(\xi_c)\mathbf{B}$;
2. Calculate the range of each column of $\mathbf{R}$, i.e., compute $\text{range}(\mathbf{r}_j) = \max(\mathbf{r}_j) - \min(\mathbf{r}_j)$ $(j = 1, \ldots, s)$ where $\mathbf{r}_j$ is the j-th column of $\mathbf{R}$;
3. Compute $\mathbf{W}_s = \text{diag}\{\text{range}^{-1}(\mathbf{r}_1), \ldots, \text{range}^{-1}(\mathbf{r}_s)\}$;
4. Then the scaled residuals used in constructing the information matrix of the design measure ξ are

$$\mathbf{F}_s^*(\xi) = \{\mathbf{F}_s(\xi) - \mathbf{F}_r(\xi)\mathbf{B}\}\mathbf{W}_s.$$

In Tables 6 and 7 and Fig. 7 we give designs for three different values of α. These were found by numerical maximisation of the design criterion, the equivalence theorem being used to check the optimality of our designs. The first, for α in (2) equal to zero, is the family of D-optimum designs for estimating the primary parameters. Since the primary model depends on x_1 and x_2 only through the product x_1x_2, the optimum design specifies three values of x_1x_2, without, for one design point, specifying individual values for x_1 or x_2. Specifically,

$$\begin{aligned} x_1x_2 &= 0.0045 \quad \text{for } (x_1, x_2) = (0.15, 0.03), \\ x_1x_2 &= 0.16 \quad \text{for } (x_1, x_2) = (0.8, 0.2), \\ \text{and } x_1x_2 &= 0.02205 \quad \text{for any } (x_1, x_2) \text{ giving this value.} \end{aligned}$$

The optimum design is therefore not unique. It puts equal weights (1/3) at low, high and intermediate values of x_1x_2. The graphical representation of the design thus consists of two points at edges corners of the design region and a line of possible third values, all of which give the same value of the optimality criterion, regardless of the shape of the experimental plot. The specific value for the experiment, if the model were known to be true, could be chosen with respect to an auxiliary criterion.

Table 6. Yield-Density relationship: design for $\alpha = 1$: D-optimum design for the seven-parameter model.

x_1	x_2	w	$[7w]$
0.15	0.03	0.143	1
0.15	0.0895	0.041	0
0.15	0.2	0.134	1
0.266	0.0633	0.115	1
0.284	0.2	0.138	1
0.421	0.03	0.003	0
0.8	0.03	0.142	1
0.8	0.0613	0.142	1
0.8	0.2	0.142	1
Efficiency			97.0%

Table 7. Yield-Density relationship: design for $\alpha = 0.25$, a parsimonious D-optimum design for model checking.

Area/plant	x_1	x_2	w	$[6w]$
0.0045	0.15	0.03	0.324	2
0.0237	0.15	0.158	0.176	1
0.0240	0.8	0.03	0.176	1
0.1600	0.8	0.2	0.324	2
Efficiency				99.8%

For α equal to one we obtain the D-optimum design for the full model with all seven terms. This design, like the design for the second-order response surface on a rectangular design region, has nine support points. But now, because the model is nonlinear and the range of the two factors is not the same, the design has, as the second panel of Fig. 7 shows, an approximate symmetry about one-diagonal of the design region. It also has very uneven weights on the support points which are plotted with the dots having diameter proportional to $w_i^{0.8}$. As a result two design points are hard to see. In Table 6 we give a seven point approximation to the design, which has an efficiency of 97.0%.

Finally we use the method of this paper for model checking to obtain a design with only four support points. We tried several values of α but here give only the results for $\alpha = 0.25$. This design again is not symmetrical, nor does it have equal weight on the four points. However a good approximation can be found. The resulting design, plotted in the third panel of Fig. 7, is close to the D-optimum design for $\alpha = 0$ in the first panel: its two middle points give areas close to the former 0.02205 and the other points are at the same corners of the design region as before. This parsimonious design is therefore highly efficient both for checking the model and for estimating the parameters in the model if it holds.

A design with four support points cannot pick up all departures from the three parameter model which are possible when the general model has seven parameters. The situation is similar to that of using trials at the centre of a two-level factorial to check that quadratic terms are not needed. If the coefficients of the quadratic terms are not zero, but sum to zero, the quadratic terms will not be detected. Protection against such an unlikely happening for the nonlinear model could be obtained by using larger values of α to generate designs with more support points. Such designs would reflect the increased emphasis on model checking implied by larger α.

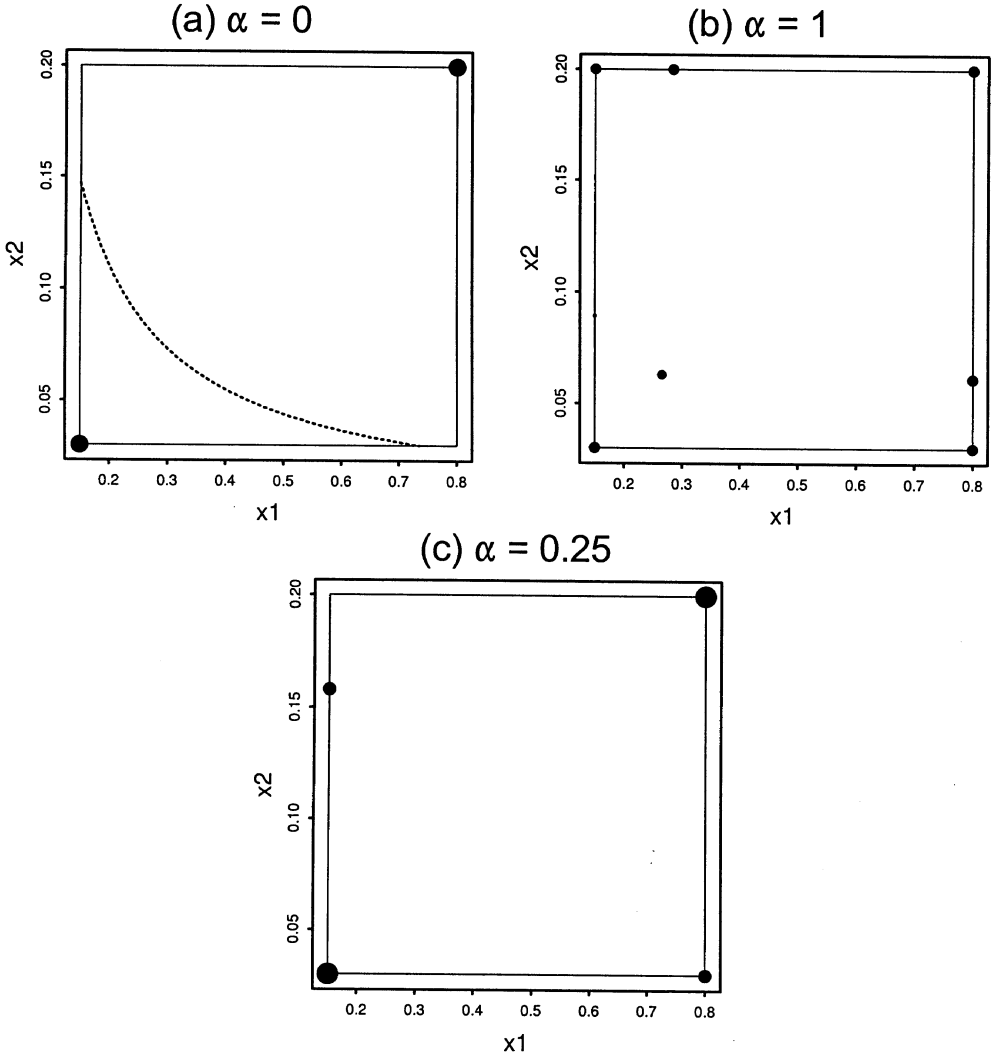

Figure 7. Yield-density relationship. Designs for model checking: (a) $\alpha = 0$: D-optimum design for primary terms. This design depends only on the value of x_1x_2 and the third design point can be anywhere on the dotted line; (b) $\alpha = 1$: D-optimum design for the seven-parameter model with dot diameter $\propto w_i^{0.8}$. Since the design weights are far from equal, all nine design points may not be evident; (c) $\alpha = 0.25$: optimum design for checking the three parameter model. The design points are close to those in panel (a) so that the exact design of Table 7 is also highly efficient for estimation in the three-parameter model.

7. Discussion

Both continuous and exact designs can be found by direct numerical maximization of the design criterion, the procedure we followed for the yield-density model. This is the method used by the authors cited here both for design augmentation and for model checking. All authors were concerned solely with linear models. We have enriched this procedure by using the equivalence theorem to construct designs, to check the optimality of proposed designs and to understand changes in the structure of the designs as the proportion of prior information changes.

8. Appendix

In Theorem 11.6 of Pukelsheim (1993) the maximization is over a set $\mathcal{M}$ which is a subset of non-negative definite matrices and contains at least one feasible matrix for $\boldsymbol{\theta}$. Theorem 1 of Section 2.2 restricts the consideration to a set $\mathcal{M}(\Xi)$ of all information matrices for designs $\xi \in \Xi$. We therefore need to show that maximization over $\mathcal{M}(\Xi)$ also yields the maximum over $\mathcal{M}$. The proof of the theorem is straightforward. The result comes from Theorem 11.6 and Lemma 6.16 of Pukelsheim (1993).

Proof. By Lemma 6.16 we get a unique matrix $\mathbf{N} = \left\{(\mathbf{M}_\alpha^*)^{-1}\right\}/p$ which simplifies the two conditions of Theorem 11.6, i.e., the *polarity equation*

$$\Phi(\mathbf{M}_\alpha^*)\, inf_{\mathcal{M}} \frac{tr\mathbf{M}_\alpha\mathbf{N}}{\Phi(\mathbf{M}_\alpha)} = tr\,(\mathbf{M}_\alpha^*\mathbf{N}) = 1$$

and the *normality inequality*

$$tr(\mathbf{AN}) \leq tr(\mathbf{M}^*\mathbf{N}) \qquad \text{for all } \mathbf{A} \in \mathcal{M}$$

to the following

$$tr\left\{\mathbf{A}\,(\mathbf{M}_\alpha^*)^{-1}\right\} \leq tr\left\{\mathbf{M}^*\,(\mathbf{M}_\alpha^*)^{-1}\right\} \qquad \text{for all } \mathbf{A} \in \mathcal{M}. \tag{A1}$$

The set $\mathcal{M}(\Xi)$ contains matrices of the form $\mathbf{A} = \mathbf{f}(\mathbf{x})\mathbf{f}^T(\mathbf{x})$, so (A1) implies the inequality (3).

Conversely, any matrix $\mathbf{A} \in \mathcal{M}(\Xi)$ has a representation

$$\mathbf{A} = \sum_{i=1}^{t} w_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i),$$

and we get

$$\begin{aligned} tr\left\{\mathbf{A}\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} &= \sum_{i=1}^t tr\left\{w_i \mathbf{f}(\mathbf{x}_i)\mathbf{f}^T(\mathbf{x}_i)\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} \\ &= \sum_{i=1}^t w_i \mathbf{f}^T(\mathbf{x}_i)\left(\mathbf{M}_\alpha^*\right)^{-1}\mathbf{f}(x_i), \\ \text{which by (3) is} \quad &\leq \sum_{i=1}^t w_i tr\left\{\mathbf{M}^*\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} \\ &= tr\left\{\mathbf{M}^*\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} \end{aligned}$$

as $\sum_{i=1}^t w_i = 1$.

Now, to show equality in (3) let $\xi' \in \Xi$, with support points $\mathbf{x}_i'$, $i = 1, \ldots, t$, be any Bayesian D-optimum design for $\boldsymbol{\theta}$, so that its information matrix satisfies the condition $tr(\mathbf{M}_\alpha'\mathbf{N}) = 1$. Let us assume that for some $\mathbf{x}_i'$ we have strong inequality

$$\mathbf{f}^T(\mathbf{x}_i')\mathbf{N}\mathbf{f}(\mathbf{x}_i') < tr(\mathbf{M}^*\mathbf{N}). \tag{A2}$$

Then

$$\begin{aligned} 1 &= tr(\mathbf{M}_\alpha'\mathbf{N}) \\ &= tr\left[\{(1-\alpha)\mathbf{M}_0 + \alpha\mathbf{M}'\}\mathbf{N}\right] \\ &= (1-\alpha)tr(\mathbf{M}_0\mathbf{N}) + \alpha tr(\mathbf{M}'\mathbf{N}) \\ &= (1-\alpha)tr(\mathbf{M}_0\mathbf{N}) + \alpha\sum_{i=1}^t w_i \mathbf{f}^T(\mathbf{x}_i')\mathbf{N}\mathbf{f}(x_i') \\ &< (1-\alpha)tr(\mathbf{M}_0\mathbf{N}) + \alpha tr(\mathbf{M}^*\mathbf{N}) \\ &= tr(\mathbf{M}_\alpha^*\mathbf{N}) = 1, \end{aligned}$$

which is a contradiction. ■

Condition (3) leads directly to inequality (4). Let us write

$$\begin{aligned} p &= tr\left\{\mathbf{M}_\alpha^*\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} \\ &= tr\left[\{(1-\alpha)\mathbf{M}_0 + \alpha\mathbf{M}^*\}\left(\mathbf{M}_\alpha^*\right)^{-1}\right] \\ &= (1-\alpha)tr\left\{\mathbf{M}_0\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} + \alpha tr\left\{\mathbf{M}^*\left(\mathbf{M}_\alpha^*\right)^{-1}\right\}. \end{aligned}$$

Hence

$$tr\left\{\mathbf{M}^*\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} + \frac{1-\alpha}{\alpha} tr\left\{\mathbf{M}_0\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} = \frac{p}{\alpha}$$

and by (3) we get the following inequality

$$\mathbf{f}^T(\mathbf{x})\left(\mathbf{M}_\alpha^*\right)^{-1}\mathbf{f}(\mathbf{x}) + \frac{1-\alpha}{\alpha} tr\left\{\mathbf{M}_0\left(\mathbf{M}_\alpha^*\right)^{-1}\right\} \leq \frac{p}{\alpha}$$

which gives (4).

References

Atkinson, A.C., Donev, A.N. (1992). *Optimum Experimental Designs.* Oxford University Press, Oxford.

Atkinson, A.C., Zocchi, S.S. (2000). Parsimonious designs for detecting departures from nonlinear and generalized linear models. (submitted).

Berry, G. (1967). A mathematical model relating plant yield with arrangement for regularly spaced crops. *Biometrics* **23**, 505–515.

Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *Annals of Statistics* **12**, 283–300.

Chaloner, K., Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science* **10**, 273–304.

DuMouchel, W., Jones, B. (1994). A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics* **36**, 37–47.

Dykstra, O. (1971a). Addendum to "The augmentation of experimental data to maximize $|X^TX|$". *Technometrics* **13**, 927.

Dykstra, O. (1971b). The augmentation of experimental data to maximize $|X^TX|$. *Technometrics* **13**, 682–688.

Evans, J.W. (1979). The computer augmentation of experimental designs to maximize $|X^TX|$. *Technometrics* **21**, 321–330.

Fedorov, V.V., Hackl, P. (1997). *Model-Oriented Design of Experiments.* Lecture Notes in Statistics 125. New York: Springer.

Gillis, P.R., Ratkovsky, D.A. (1978). The behavior of estimators of the parameters of various yield-density relationships. *Biometrics* **34**, 191–198.

Heiberger, R.M., Bhaumik, D.K., Holland, B. (1993). Optimal data augmentation strategies for additive models. *Journal of the American Statistical Association* **88**, 926–938.

Lee, P.M. (1997). *Bayesian Statistics.* London: Arnold.

Lin, C.S., Morse, P.M. (1975). A compact design for spacing experiments. *Biometrics* **31**, 661–671.

McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models (2nd edition).* London: Chapman and Hall.

Mead, R. (1970). Plant density and crop field. *Applied Statistics* **19**, 64–81.

Mead, R. (1979). Competition experiments. *Biometrics* **35**, 41–54.

Mitchell, T.J. (1974). An algorithm for the construction of 'D-optimum' experimental designs. *Technometrics* **16**, 203–210.

Pilz, J. (1983). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Leipzig: Teubner.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery B.P. (1992). *Numerical Recipes in Fortran (2nd edition).* Cambridge, England: Cambridge University Press.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* New York: Wiley.

Ratkovsky, D.A. (1989). *Nonlinear Regression Modeling.* New York: Marcel Dekker.

Seber, G. A.F., Wild, C.J. (1989). *Nonlinear Regression.* New York: Wiley.

Shinozaki, K., Kira, T. (1956). Intraspecific competition among higher plants. VII. Logistic theory of the C–D effect. *J. Inst. Polytech Osaka City Univ. D7*, 35–72.

Silvey, S.D. (1980). *Optimum Design.* London: Chapman and Hall.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *Journal of the Royal Statistical Society, Series B* **35**, 123–130.

Wiley, R.W., Heath, S.B. (1969). The quantitative relationships between plant population and crop field. *Adv. Agronomy* **21**, 281–321.

Wynn, H.P. (1970). The sequential generation of D-optimal experimental designs. *Annals of Mathematical Statistics* **41**, 1055–1064.

Received 27 March 2000; revised 15 October 2000

Warunek równoważości zastosowany do optymalnego powiększania układu eksperymentalnego i do efektywnego sprawdzania modelu: powierzchnie wielomianowe i model plonu

STRESZCZENIE

Rozważamy zastosowanie dwóch problemów teorii układów eksperymentalnych. W celu powiększenia planu eksperymentu dodajemy punkty planu w sposób optymalny pod względem estymacji parametrów modelu. Dla sprawdzenia modelu znajdujemy optymalny plan o niewielkiej liczbie punktów. W obu przypadkach znajdujemy układy D-optymalne. Szczególne przypadki warunku równoważności dla D-optymalności układu stanowią podstawę algorytmu konstrukcji układów jak również dają wgląd w ich strukturę. Przykłady planów dla powierzchni wielomianowych drugiego rzędu nad obszarem regularnym i nieregularnym poprzedzają przykład zastosowania układu optymalnego do nieliniowego modelu plonu w doświadczeniu polowym.

SŁOWA KLUCZOWE: Układ Bayesowski, D-optymalność, obszar nieregularny, model nieliniowy, powierzchnia wielomianowa.