
Journal of Global Optimization **18:** 385–398, 2000.

Global Optimization Problems in Optimal Design of Experiments in Regression Models

E.P.J. BOER and E.M.T. HENDRIX
Sub-department of Mathematics, Wageningen University, Dreyenlaan 4, 6703 HA Wageningen, The Netherlands (e-mail: eric.boer@wts.wk.wau.nl)

Abstract. In this paper we show that optimal design of experiments, a specific topic in statistics, constitutes a challenging application field for global optimization. This paper shows how various structures in optimal design of experiments problems determine the structure of corresponding challenging global optimization problems. Three different kinds of experimental designs are discussed: discrete designs, exact designs and replicationfree designs. Finding optimal designs for these three concepts involves different optimization problems.

Key words: Optimal design of experiments, Nonconvex structure, Mixed integer/continuous optimization, Parameter estimation

1. Introduction

In many fields of sciences, experiments are done in order to estimate parameters of regression models. Optimal experimental designs can be used to maximize the precision of the least squares estimator, given the total number of observations. The theory of optimal experimental design has been explained (among others) in the monographs of Fedorov (1972), Silvey (1980) and Pukelsheim (1993). Atkinson and Donev (1992), Atkinson (1996) and Müller (1998) show the usefulness of optimal experimental designs in a more practical setting. Given the total number of observations, the optimal design is determined by the design space (experimental region), the regression model and the optimality criterion. Searching for these optimal designs yields challenging optimization problems (Zhigljavsky, 1991), which has resulted in a large number of publications (among others: Welch, 1982; Gaffke and Mathar, 1992; Jones and Wang, 1999). In this paper it is shown how general and more specific properties of experimental design problems result in properties of optimization problems for three different kinds of experimental designs. Important properties of optimal experimental designs are discussed and it is indicated how these properties can be helpful by solving the optimization problems for finding the optimal design.

This paper considers optimal experimental design in the context of regression models. Let

$$Y_i = \eta(x_i, \theta) + E_i, \quad x_i \in X \subset \mathbb{R}^k \tag{1}$$

be a (statistical) regression model with a regression function η and i.i.d. zero-mean error terms E_i. The unknown θ is a parameter vector with m elements, $\theta^T = (\theta_1, \ldots, \theta_m) \in \Omega \subset \mathbb{R}^m$. Further we assume that η is a twice differentiable continuous function.

2. Theory of optimal design of experiments

A concept of an experimental design in regression analysis, frequently used in literature, is that of a so-called *discrete design*. A discrete design ϵ is written as:

$$\epsilon = \begin{pmatrix} x_1 & x_2 & \cdots & x_r \\ p_1 & p_2 & \cdots & p_r \end{pmatrix} \tag{2}$$

where p_i indicates measurement weight at support point x_i, $i = 1, 2, \ldots, r$. The weights sum to unity: $\sum_{i=1}^{r} p_i = 1$, $p_i \geqslant 0$. The support points are chosen from the design space X; $x_i \in X$. The design space X may have dimension $\geqslant 1$, which means that also spatial problems could be considered (Müller, 1998). From an optimization point of view, we would like to find, for given number of support points r, the best values for p_i and x_i (in a sense to be specified). Notice however, that in some situations this number r is not known beforehand.

A more practical definition of an experimental design is that of a normalized exact design. In this design, for all p_i holds that $p_i N$ is integer, where N is the maximum number of observations allowed in the experiment. An *exact design* (not normalized) $\epsilon(N)$ is usually written as follows:

$$\epsilon(N) = \begin{pmatrix} x_1 & x_2 & \cdots & x_r \\ n_1 & n_2 & \cdots & n_r \end{pmatrix} \tag{3}$$

where n_i is the number of replications at each support point, $\sum_{i=1}^{r} n_i = N$. An exact design becomes discrete by using $p_i = n_i/N$. It is worth noticing, that searching for an exact design (in practice all designs are exact) results in a mixed continuous/integer optimization problem. These problems are in general hard to solve.

In spatial problems, but also in other problems, observing in a point of the design space is often restricted to a certain number of replications. If the number of replications is restricted to one (*replicationfree design*), the observations have a minimal distance between each other. In this case, the design space X is often (see e.g. Fedorov, 1989) considered as (a grid of) Q candidate points or possible measurement points (observations). The design problem becomes a combinatorial problem of selecting N observations from Q candidate points. The solution of the problem will give an exact design with only one replication at each support point ($N = r$).

Optimality of a design depends on the function $\eta(x, \theta)$ with parameter vector θ, under consideration. In many cases, research focuses on models with are linear in the parameters; then $\eta(x, 0)$ can be written as $\theta^T f(x)$. Moreover performance

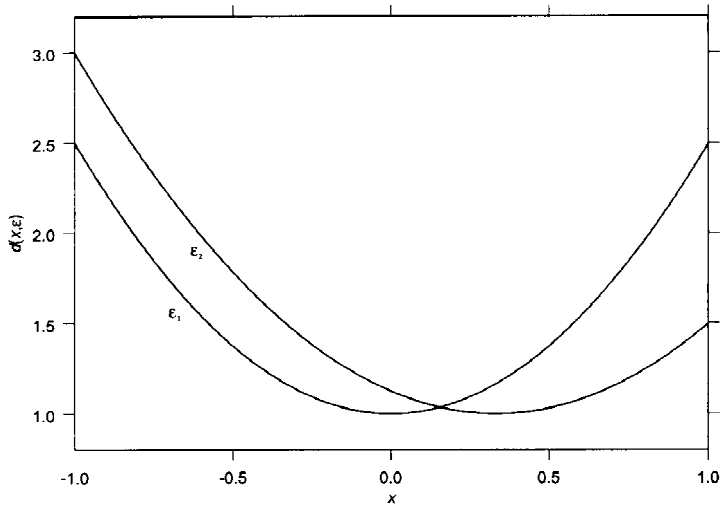

Figure 1. Standardized variance functions for the two normalized exact designs ϵ_1 and ϵ_2 for $N = 3$.

depends on a specific criterion which is a function of the so-called information matrix. Experiments which contain a lot of information enlarge the precision of the estimation of the parameters of the model. If a linear model with the usual regression assumptions of independent errors and constant variance is studied, the information matrix for a discrete design is given by:

$$M[\epsilon] = \sum_{i=1}^{r} p_i f(x_i) f^T(x_i) \tag{4}$$

The inverse of the information matrix (M^{-1}, for exact designs the variance-covariance matrix of $\hat{\theta}$) is helpful to represent the variance of the predictor $\eta(x, \hat{\theta})$ on the design space X by means of the standardized variance function. The standardized variance function (under the usual statistical assumptions) is defined as follows:

$$d(x, \epsilon) = \text{var}[\eta(x, \hat{\theta})] = f^T(x) M^{-1} f(x) \tag{5}$$

This standardized variance function makes the design problem easier to understand from a graphical point of view. Note that for linear models the standardized variance function is independent of θ, because f does not depend on the parameter vector θ. Figure 1 shows the standardized variance functions of two designs ($N = 3$) for a simple linear model, $\eta(x, \theta) = \theta_1 + \theta_2 x$, $x \in [-1, 1]$. The designs are chosen as follows:

$$\epsilon_1 = \begin{pmatrix} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}, \qquad \epsilon_2 = \begin{pmatrix} -1 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

The two designs are almost the same, only the measurement at $x = 0$ in design ϵ_1 is moved to the right end of the interval for design ϵ_2. Figure 1 shows the result of this movement, as the standardized variance at the right end of the interval is lowered. The maxima of $d(x, \epsilon)$ are found, for any design, at the extreme points of the design space. In Section 2.2 we will come back to this.

The theory of optimal experimental designs can be extended to nonlinear models by considering the Taylor series expansion (Atkinson and Donev, 1992). In this case, for $f^T(x)$ the vector of partial derivatives is used

$$f^T(x) = \left[\frac{\partial \eta(x,\theta)}{\partial \theta_1} \quad \frac{\partial \eta(x,\theta)}{\partial \theta_2} \quad \cdots \quad \frac{\partial \eta(x,\theta)}{\partial \theta_m} \right] \tag{6}$$

Optimal designs for nonlinear models are called *locally optimal designs* because $f(x)$ depends on values of θ, so local with respect to parameter values of θ. This is confusing given another interpretation of the terminology of locally optimal solutions in global optimization. The information matrix in the nonlinear case is denoted by $M[\theta, \epsilon]$.

2.1. CRITERIA

Many criteria for optimal designs are functions of the information matrix, say $\phi(M[\epsilon])$. The most popular criterion is *D-optimality*, which minimizes the generalized variance of the parameter estimates. This corresponds to minimizing the value of the determinant of the variance-covariance matrix $M^{-1}[\varepsilon]$. When interest is focused on estimation of a subset of elements of θ, the criterion is written as D_s. A design which minimizes the maximum of the standardized variance function over the design region X is called a *G-optimal* design. The criteria are determined by minimizing the following functions:

$$\text{D-optimality: } \det(M^{-1}[\epsilon]) \tag{7}$$

$$\text{D}_s\text{-optimality: } \det(M^{11}[\epsilon]) \tag{8}$$

$$\text{G-optimality: } \max_{x \in X} d(x, \epsilon) \tag{9}$$

where $M^{11}[\epsilon]$ is the $s \times s$ submatrix of $M^{-1}[\epsilon]$ with rows and columns corresponding to the s selected elements of θ. The exact designs ($N = 3$) obtained from ϵ_1 and ϵ_2 presented in Figure 1 are G-optimal and D-optimal designs respectively.

2.2. PROPERTIES OF OPTIMAL DESIGN PROBLEMS

The optimization problem for a discrete design can be considered as choosing the best x_i and p_i. A problem in using general purpose optimization methods is that the number of support points is not known beforehand (Jones and Wang, 1999). The optimal design problem becomes even more difficult when an exact design

is needed, which results into a mixed continuous/integer optimization problem. Boer et al. (2000) show that this problem can not be solved very easily, because of local optima. Some important properties (theorems) from the optimal experimental design theory can assist in solving the optimization problems. We will present them without going very much into detail giving the reader a flavour of the existing theory. The properties will be ordered from general properties of models and criteria to more specific cases.

- The criterion function $\phi(M[\epsilon])$ has certain properties that capture the idea of whether the information in matrix M is large or small. If a design ϵ^* is at least as good as ϵ, the information in matrix $M[\epsilon^*]$ is considered larger than that in $M[\epsilon]$ in a certain ordering. A reasonable criterion of the information matrix is that the value of a criterion function is non-decreasing (monotonic) when measurements are removed from an existing design. An application of this property in a Branch-and-Bound algorithm for optimal replicationfree designs is discussed in Section 3.3.
- Although, the number of support points r of a design is variable in the optimization problem considered here, certain bounds can be given. These bounds can be derived by looking at some basic properties of the information matrix M (Fedorov, 1972):
  1. For any design $\epsilon : M[\epsilon]$ is positive-semidefinite.
  2. If $r < m$ then $\det(M[\epsilon]) = 0$, i.e. $M^{-1}[\epsilon]$ does not exist.
  3. For any compact design space X, the set $\{M[\epsilon]; \epsilon \text{ is discrete}\}$ is convex.

  From property 2 it is clear that the number of support points should at least be equal to the number of parameters ($r \geqslant m$). Property 3 leads together with Carathéodory's Theorem (see Silvey, appendix 2, 1980) to an upper bound for the minimum number of support points. This upper bound is equal to $\frac{1}{2}m(m+1)+1$. For D-optimality this can be strengthened to $\frac{1}{2}m(m+1)$. Thus, for certain criteria $\phi(M)$ there exists an optimal design with at least m and at most $\frac{1}{2}m(m+1)+1$ support points. These bounds for the number of support points are especially useful for general purpose optimization (Section 3.4).
- The most celebrated theorem in optimal design of experiments is undoubtedly the Equivalence Theorem of Kiefer and Wolfowitz (1960). This theorem states that the following characterizations of an optimal discrete design ϵ^* are equivalent.

$$(i) \quad \text{design } \epsilon^* \text{ is D-optimal} \tag{10}$$

$$(ii) \quad \epsilon^* \text{ minimizes } \max_x d(x, \epsilon) \text{ or } \epsilon^* \text{ is G-optimal} \tag{11}$$

$$(iii) \quad \max_x d(x, \epsilon^*) = m \tag{12}$$

  For a discussion and proof see (among others) Silvey (1980). This theorem was first derived by Kiefer and Wolfowitz (1960) for linear models, but White (1973) showed that it can be extended to nonlinear models. Note that this

theorem holds for discrete designs, not for all exact designs. Figure 1 shows an example of (normalized) exact designs where the D- and G-optimal designs are different.

This theorem gives the opportunity to calculate D-optimal discrete designs by means of properties of G-optimal (discrete) designs. For G-optimality the maximum of the standardized variance matrix is minimized. It is known – see (*iii*) of the Equivalence theorem – that as long as this maximum is larger than m, the design is not G-optimal and thus not D-optimal. By putting (additional) weight at point x^* where the maximum of $d(x, \epsilon)$ is reached, the standardized variance at point x^* can be lowered (see Figure 1). This concept is used in the development of an algorithm (Fedorov, 1972), which will be elaborated further in Section 3.2.

– Optimal *exact* designs are often difficult to calculate because the number of replications at each support point should be integer. In the special case of $r = m$ and D-optimality, the number of replications should be chosen as equal as possible. Rasch (1990) shows this by rewriting the information matrix $M[\theta, \epsilon]$ in the following matrix notation:

$$M[\theta, \epsilon] = G^T(\theta, \epsilon(N)) \mathcal{N} G(\theta, \epsilon(N)) \tag{13}$$

where

$$G^T(\theta, \epsilon(N)) = [f(x_1) \quad f(x_2) \quad \cdots \quad f(x_r)]$$

and

$$\mathcal{N} = \text{diag}(n_1, n_2, \ldots, n_r)$$

Minimizing $\det(M^{-1}[\epsilon])$ means maximizing $\det(M[\epsilon])$. Now

$$|G^T(\theta, \epsilon(N)) \mathcal{N} G(\theta, \epsilon(N))| = |G^T(\theta, \epsilon(N))||\mathcal{N}||G(\theta, \epsilon(N))| \tag{14}$$

G is independent of n_i and $|\mathcal{N}| = \prod n_i$ is maximized when the values of n_i are as equal as possible.

– Figure 1 illustrates the standardized variance functions for two designs for the easy case of simple linear regression. The maxima of the standardized variance functions can be found in the extreme points of the design space, due to the convexity of these functions. As long as $f(x)$ is linear in x, the standardized variance function $d(x, \epsilon) = f^T(x) M^{-1} f(x)$ is a convex quadratic function. If $f(x)$ is nonlinear in x, the resulting standardized variance function is not quadratic. Consider the following quasi-linear (linear in parameters) model:

$$\eta(x, \theta) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2 \tag{15}$$

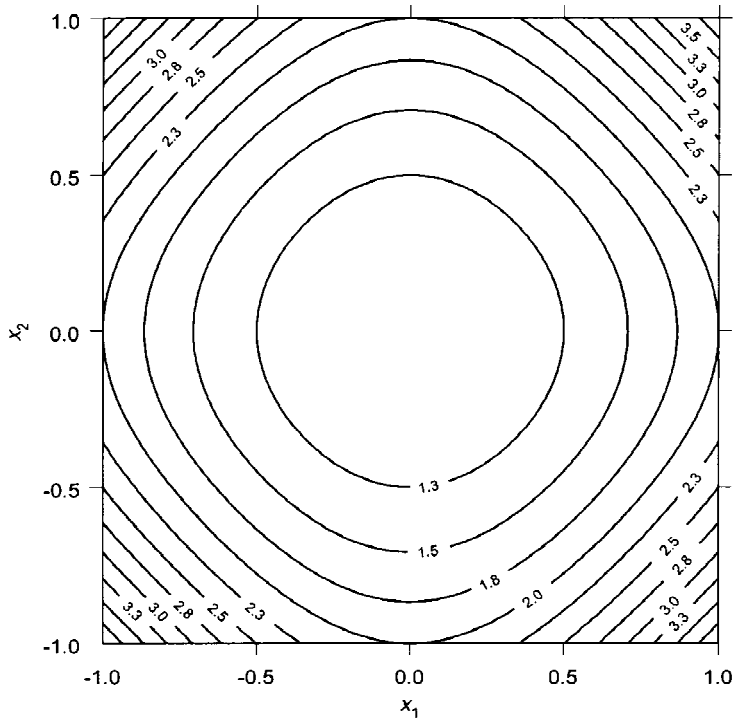

Figure 2. Contour map of the standardized variance function of $\eta(x, \theta) = \theta_1 + \theta_2 x_1 + \theta_3 x_2 + \theta_4 x_1 x_2$ for design ϵ_3 on a unit square.

where x_1 and x_2 can be chosen from $X = [-1, 1]^2 \subset \mathbb{R}^2$. A D-optimal design for this function is equal to:

$$\epsilon_3 = \begin{pmatrix} 1 & 1 & -1 & -1 \\ -1 & 1 & 1 & -1 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{pmatrix} \tag{16}$$

where the first row contains the coordinates of x_1 and the second row the coordinates of x_2. It can be shown that this design is D-optimal by calculating the standardized variance function for design ϵ_3

$$d(x, \epsilon_3) = 1 + x_1^2 + x_2^2 + x_1^2 x_2^2 \tag{17}$$

A plot of this standardized variance function on a unit square is given in Figure 2. Note that this design is indeed D-optimal, because the maximum on the unit square is equal to the number of parameters (see equation (12); Müller, 1998).

3. Searching the optimal design

In this section different ways of finding optimal designs are discussed. The properties of the design problem for a certain model and criterion can be extended to a complete analytical solution for a specific design problem (Section 3.1). However, most problems are too complex to find an analytical solution. Therefore, Fedorov introduced an algorithm to find the optimal solution (Section 3.2). In Section 3.3 a combinatorial optimization algorithm is outlined, for the special case of a design space consisting of a finite set of candidate points. Finally, it is illustrated how general purpose optimization algorithms will perform for these kind of problems (Section 3.4).

3.1. ANALYTICAL RESULTS

Many examples of analytical derivations of optimal designs can be found in literature (e.g. Fedorov, 1972; Rasch, 1990; Vila, 1991). An illustrative example of an analytical solution of a design problem is given by Boer et al. (2000). An exact D-optimal two-point design for the Michaelis-Menten function ($r = m = 2$) can be found by minimizing

$$\phi(M^{-1}[\epsilon]) = K(x_1, x_2, n_1, n_2) = \frac{(1+\beta x_1)^4(1+\beta x_2)^4}{\alpha^2(x_1 - x_2)^2 x_1^2 x_2^2 n_1 n_2} \tag{18}$$

It is obvious that n_1 and n_2 should be chosen as equal as possible, as already has been shown by the third theoretical property in Section 2.2. The choice of x_1 and x_2 is more complicated. For $x \in [0, x_u]$, $x_u > 0$ it can be derived that the following exact design is D-optimal

$$\begin{pmatrix} \frac{x_u}{2 + x_u\beta} & x_u \\ n_1 & n_2 \end{pmatrix} \tag{19}$$

with $n_1 + n_2 = N$. For $N = 2n$, n_1 equals n_2 and for $N = 2n + 1$ choose $n_1 = n, n_2 = n + 1$ or $n_1 = n + 1, n_2 = n$ (Ermakov and Zhigljasky, 1987).

3.2. SPECIAL ALGORITHMS FOR OPTIMAL DESIGNS

Because an analytical solution can not be found for every design problem, some specific algorithms have been constructed to find the optimal solution. Although more algorithms are available, we restrict ourselves to the V-(Fedorov, 1972) algorithm, which can be described as follows:

1. Given start design ϵ_0, stopping criteria, $s = 0$, r_0 the number of support points of ϵ_0.

2. Determine:

$$M[\epsilon_s] = \sum_{i=1}^{r_s} p_{is} f(x_{is}) f^T(x_{is}).$$

3. Calculate $D[\epsilon_s] = M^{-1}[\epsilon_s]$.
4. Now $d(x, \epsilon_s) = f^T(x) D[\epsilon_s] f(x)$.
   Determine:

$$\delta_s = \max_x d(x, \epsilon_s) - m.$$

$$\text{a point } x_s^* \in \arg\max_x d(x, \epsilon_s).$$

5. The step-size: $\alpha_s = \delta_s/(\delta_s + (m-1))m$.
6. ϵ_{s+1} is calculated by:
   a)
   Recalculate all the weights ($i = 1, 2, \ldots, r_s$) of ϵ_s in the following way:

$$p_{i(s+1)} = p_{is}(1 - \alpha_s).$$

   b)
   Add x_s^* to design ϵ_s with weight α_s, update r_{s+1}. If $x_s^* \in \epsilon_s$, update ϵ_{s+1}.
7. Check stopping criteria, $s := s + 1$ and go to 2.

This algorithm is mainly based on the properties of the Equivalence Theorem. It is known that a D-optimal (discrete) design minimizes the maximum of the standardized variance function. This algorithm puts (additional) weight on the value of a point x^* (step 6) where the standardized variance function reaches its maximum (step 4), as long as that maximum is larger than the number of parameters considered. Note that step 4 implies a global optimization problem. Jones and Wang (1999) mention some pros and cons of this algorithm. The main advantage of this algorithm is that the number of support points does not have to be fixed beforehand. Further, it is important that the algorithm ensures convergence to the optimal design under some conditions. The main disadvantage is that the algorithm may be very slow for some problems (Atkinson and Donev, 1992). This is mainly caused by the fact that after introduction, a (possibly non-optimal) support point does not disappear, probably only its weight decreases. The following example is given as an illustration of the V-algorithm.

Consider the following model

$$\eta(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 \tag{20}$$

with $X = [0, 1]$. Figure 3 shows the standardized variance function of the following (not optimal) design, which is used as start design for the V-algorithm.

$$\epsilon_0 = \begin{pmatrix} 0 & 0.2 & 1 \\ 0.3333 & 0.3333 & 0.3333 \end{pmatrix}$$

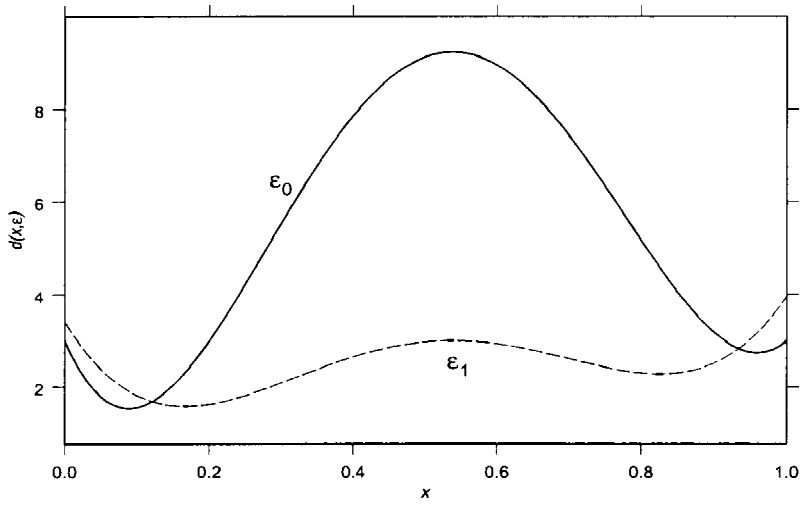

Figure 3. Illustration of the V-algorithm for model (20) with start design ϵ_0.

After one iteration of the V-algorithm the design becomes as follows:

$$\epsilon_1 = \begin{pmatrix} 0 & 0.2 & 1 & 0.5378 \\ 0.2492 & 0.2492 & 0.2492 & 0.2523 \end{pmatrix}$$

The standardized variance function of design ϵ_1 is graphically represented in Figure 3. We restrict ourselves to one iteration. The final result of the algorithm converges to the D-optimal design, which is:

$$\epsilon^* = \begin{pmatrix} 0 & 0.5 & 1 \\ 0.3333 & 0.3333 & 0.3333 \end{pmatrix}$$

3.3. COMBINATORIAL OPTIMIZATION OF THE OPTIMAL DESIGN

The optimal design problem becomes a combinatorial optimization problem, when the design space is restricted to a finite discrete set of Q candidate (design) points, $B_Q = \{x_1, x_2, \ldots, x_Q\}$. Rasch et al. (1997) show some algorithms for selecting N design points out of B_Q.

The calculation time of full enumeration of this problem was reduced considerably by applying a Branch-and-Bound algorithm. This fast Branch-and-Bound algorithm is based on the fact that the criterion function $\phi(M[\epsilon])$ is monotonic (see Section 2.2). The drawback of this procedure is that the number of candidate points is restricted to about 30. Figure 4 gives an impression of this combinatorial optimization problem for one of the examples used by Rasch et al. (1997).

Müller and Pázman (1998) constructed an algorithm to find optimal designs with more candidate points. The algorithm makes use of a corresponding information matrix, which approximates the information matrix for exact designs. Prom-

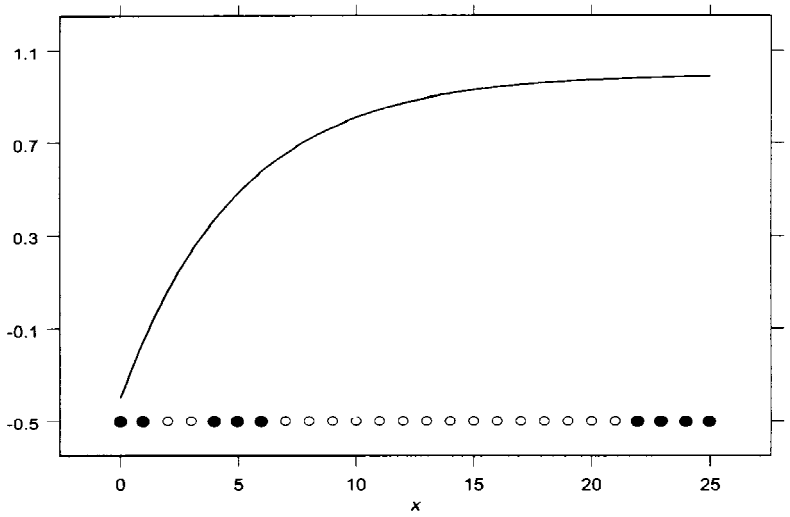

Figure 4. A plot of the function $1 - 1.4e^{-0.2x}$ with the corresponding 9-point replicationfree D-optimal design selected from 26 candidate points.

ising results are shown for a spatial example of Fedorov (Fedorov, 1989) of a 20×20 point grid.

3.4. GENERAL PURPOSE OPTIMIZATION

The difficulty in finding optimal designs with general purpose optimization procedures is that the number of support points is not known beforehand. We saw already that there can be given certain bounds for the number of support points in Section 2.2. However, it would be preferable when a general purpose optimization procedure does not depend on the number of support points. Boer et al. (2000) illustrate with the Michaelis-Menten function, that the mixed continuous/integer programming problem can be rewritten into a fully continuous nonlinear programming problem, formulated as follows:

$$\begin{aligned} &\min\{K(x'_1, x'_2, \ldots, x'_N)\} \\ &\quad \text{under the condition:} \\ x_l \leqslant x'_1 \leqslant x'_2 \leqslant \cdots \leqslant x'_N \leqslant x_u \end{aligned} \tag{21}$$

where K is equal to a certain criterion, x_l is the lower bound and x_u the upper bound of the one dimensional design space. In this case, x'_i are (single) measurement points in the design space.

In the paper of Boer et al. (2000) it is shown that a (sub)-D-optimal design with 6 and 4 replications at the two support points is a local minimum of the continuous optimization problem. Figure 5 illustrates this by changing the value of variable x'_6 of the D-optimal design ($x'_1, \ldots, x'_5 = 28.32$, and $x'_6, \ldots, x'_{10} = 1440$) from the

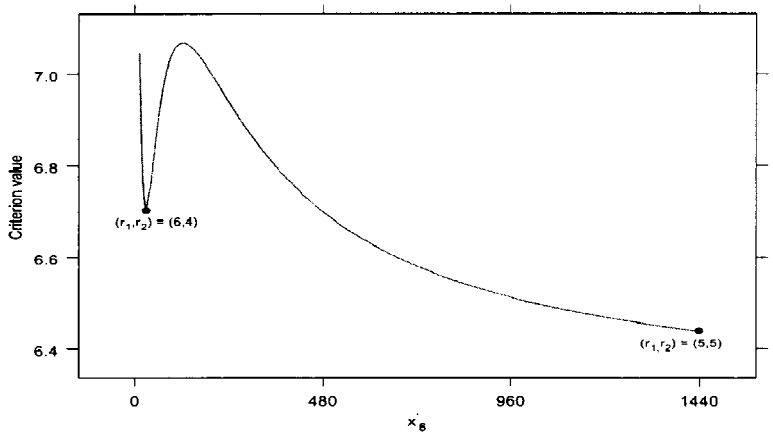

Figure 5. Non-convexity of the continuous NLP formulation of the D-optimal design problem

lower to the upper bound of the design space. In this way many local minima may appear.

Jones and Wang (1999) argue that general optimization procedures are more efficient than special algorithms like the V-algorithm. They use global optimization methods, because the criterion considered has several local optima. They discuss two well-known stochastic global optimization methods: multi-start local search and simulated annealing. For the last it is suggested to stop the annealing procedure at a certain point and then continue the search by an effective local search procedure.

4. Conclusions

This paper shows how the structure of the design space, model and criterion in optimal design of experiments problems determines the structure of corresponding challenging global optimization problems. Three different kinds of experimental designs are discussed: discrete designs, exact designs and replicationfree designs. Finding the optimal designs for these three concepts involves different optimization problems.

Discrete design problems are most easy to solve. There are many examples of a complete analytical derivation of the optimal design, without using optimization methods. However, if an analytical solution is not available, optimization methods are needed. Fedorov (1972) proposes a specific algorithm which ensures convergence to the optimal discrete design, but may be very slow for some problems. General purpose optimization does often not work adequately, because the number of support points is often not known beforehand and local minima may occur (Jones and Wang, 1999).

Exact design problems are hard to solve, because finding the optimal design implies solving a mixed continuous/integer optimization problem. Boer et al. (2000)

show that a fully continuous formulation of the problem results in many local minima. An other interesting approach to find exact designs is to construct exact design from the optimal discrete design with a certain rounding method (Pukelsheim and Rieder, 1992; Gaffke and Heiligers, 1995). It can be shown that the criterion values of these exact designs have a limited loss of efficiency compared to the criterion values of optimal exact designs.

If the design space is restricted to a set of candidate points, combinatorial optimization can be applied to find the optimal solution. Rasch et al. (1997) show a Branch-and-Bound algorithm (full enumeration) for this, based on the fact that every reasonable design criterion is monotonic. A same kind of algorithm, in this case for maximum entropy sampling, can be found in Ko et al. (1995). For both articles, full enumeration is only applicable when the number of candidate points is restricted. Larger problems have to be solved with search algorithms (Fedorov, 1989; Müller, 1998).

Up to now no specific global optimization algorithms have been developed in the field of optimal experimental designs. In our opinion, optimal design of experiments constitutes a challenging application field for global optimization.

Acknowledgements

We would like to thank A.C. van Eijnsbergen, D.A.M.K. Rasch and two referees for reading and giving helpful comments on earlier versions of the manuscript. We are grateful for the useful remarks of our recently deceased colleague J.M. ten Vregelaar.

References

1. Atkinson, A.C. (1996), The usefulness of optimum experimental designs, *Journal of the Royal Statistical Society* B 58(1): 59–76.
2. Atkinson, A.C. and Donev, A.N. (1992), *Optimum experimental designs*. Oxford University Press, Oxford.
3. Boer, E.P.J., Rasch, D.A.M.K. and Hendrix, E.M.T. (2000), Locally optimal designs in non-linear regression: A case study of the Michaelis-Menten function, in Balakrishnan, N., Ermakov, S.M. and Melas, V.B. (eds.), *Advances in Stochastic Simulation Methods*, Birkhauser, Boston.
4. Ermakov, S.M. and Zhiglijavsky, A.A. (1987), *Matematitscheskaja teorija optimalnich experimentov.* Nauka, Moskva.
5. Fedorov, V.V. (1972), *Theory of optimal experiments*. Academic Press, New York.
6. Fedorov, V.V. (1989), Optimal design with bounded density: Optimization algorithms of the exchange type, *Journal of Statistical Planning and Inference* 22: 1–13.
7. Gaffke, N. and Mathar, R. (1992), On a class of algorithms from experimental design theory, *Optimization* 24: 91–126.
8. Gaffke, N. and Heiligers, B. (1995), Algorithms for optimal design with application to multiple polynomial regression, *Metrika* 42: 173–190.
9. Jones, B. and Wang, J. (1999), Constructing optimal designs for fitting pharmacokinetic models, *Computational Statistics* 9: 209–218.

10. Kiefer, J.C. and Wolfowitz, J. (1960), The equivalence of two extremum problems, *Canadian Journal of Mathematics* 12: 363–366.
11. Ko, C., Lee, J. and Queyranne, M. (1995), An exact algorithm for maximum entropy sampling, *Operations Research* 43(4): 684–691.
12. Müller, W.G. (1998), *Collecting spatial Data – Optimum design of experiments for random fields*. Physica-Verlag, Heidelberg.
13. Müller, W.G. and Pázmann, A. (1998), Design measures and approximate information matrices for experiments without replications, *Journal of Statistical Planning and Inference* 71: 349–362.
14. Pukelsheim, F. (1993), *Optimal design of experiments*. Wiley, New York.
15. Pukelsheim, F. and Rieder, S. (1992), Efficient rounding of approximate designs, *Biometrika* 79(4): 763–770.
16. Rasch, D.A.M.K. (1990), Optimum experimental design in nonlinear regression, *Commun. Statist.-Theory Meth.* 19(12): 4789–4806.
17. Rasch, D.A.M.K., Hendrix, E.M.T. and Boer E.P.J. (1997), Replication-free optimal design in regresion analysis, *Computational Statistics* 12: 19–52.
18. Silvey, S.D. (1980), *Optimal design*. Chapman and Hall, London.
19. Vila, J.P. (1991), Local optimality of replications from a minimal D-optimal design in regression: A sufficient and quasi-necessary condition, *Journal of Statistical Planning and Inference* 29: 261–277.
20. Welch, W.J. (1982), Branch-and-Bound search for experimental designs based on *D* optimality and other criteria, *Technometrics* 24(1): 41–48.
21. White, L.V. (1973), An extension to the general equivalence theorem for nonlinear models, *Biometrika* 60: 345–348.
22. Zhigljavsky, A.A. (1991), *Theory of Global Random Search.* Kluwer Academic Publishers, Dordrecht.