Biometrical Journal **42** (2000) 3, 349–362

Online Calculation of Efficient Designs for Multi-Factor Models

HOLGER DETTE

Faculty and Institute for Mathematics
Ruhr-University Bochum
Germany

INGO ROEDER

Institute for Medical Informatics, Statistics und Epidemiology
University Leipzig
Germany

Summary

A class of "incomplete" multivariate polynomial regression models on the q-cube is considered. A simple algorithm for the determination of D-optimal product designs is developed and illustrated by specific examples. The methods are implemented on an IBM compatible PC under MS-DOS and provide an effective method for the numerical solution of an optimal design problem, which was unsolved in nearly all cases of practical interest.

Key words: Multivariate polynomial regression; Missing interactions; Product designs.

1. Introduction

The construction of optimal experimental designs is of fundamental importance in response surface analysis, especially when the functional form of the relationship between the response and explanatory variables is not completely known. Assume that there is a one dimensional response, say Y, which depends on a q-dimensional explanatory variable $x = (x_1, \ldots, x_q)$. It is common practice to approximate this response surface relationship $z = y(x_1, \ldots, x_q)$ by a multivariate polynomial of degree m, say

$$f(x) = \alpha_0 + \sum_{i=1}^{q} \alpha_i x_i + \sum_{1 \le i_1 \le i_2 \le q} \alpha_{i_1, i_2} x_{i_1} x_{i_2} + \cdots + \sum_{1 \le i_1 \le \ldots \le i_m \le q} \alpha_{i_1, \ldots, i_m} \prod_{j=1}^{m} x_{i_j}\,, \qquad (1)$$

and then fit the standard polynomial model using least squares methods. The polynomial expression (1) can be thought as a Taylor's series expansion of the "true" underlying function truncated after terms of mth order. In practice an increasing degree m usually improves the approximation of the unknown regression function y by the polynomial model f and the smaller the region of approximation needs to be made the better is this approximation. Usually, lower order polynomials are used for these approximation in order to avoid models with a large number of parameters.

Various properties of the resulting fitted models have been investigated, especially those properties which are influenced by the choice of an experimental design. The most popular optimality criterion for the choice of a design is D-optimality which minimizes the volume of the confidence ellipsoid for the unknown parameters of the regression. Properties of designs for the multivariate polynomial regression (1) and product type multivariate models have been studied intensively in the literature [see e.g. KONO, 1962; FARELL, KIEFER, and WALBRAN, 1967; LIM and STUDDEN, 1988: RAFAJLOWICZ and MYSZKA, 1988, 1992; and WONG, 1994]. Most authors assume that all regression functions

$$\prod_{j=1}^{q} x_j^{h_j} \quad \text{with} \quad \sum_{j=1}^{q} h_j \leq m \tag{2}$$

are present in the model (1). However, in many cases of practical interest it can be justified that some of the multiple monomials do not appear in the regression. For example, it is possible that some of the interactions or some of the highest terms can be neglected in the response surface relationship (1) [see UPPERMANN (1993)]. DETTE and RÖDER (1996) studied "incomplete" multivariate polynomial regression models and provided a characterization of D-optimal product designs for a large class of models of this type. It is the purpose of the present paper to develop an efficient algorithm, which is based on these results and allows an "online" calulation of D-optimal product designs on the q-dimensional cube.

Section 2 introduces and illustrates the multivariate "incomplete" polynomial regression model while Section 3 provides some of the theoretical background which is needed in our algorithm for the calculation of the D-optimal product designs. Section 4 describes the algorithm and its implementation on an IBM compatible PC under MS-DOS. Finally, the impact of our method in applied regression analysis is illustrated in a concrete example in Section 5.

2. "Incomplete" Multivariate Polynomial Regression

Our paper is motivated by the observation that there are many cases where the experimenter already knows that not necessarily all monomials of the form (2) appear in the multivariate polynomial regression model (1). In order to fix ideas we start with a small example.

Example 2.1: *Consider a second order response surface relationship in three variables, i.e.*

$$f(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 x_2 + \alpha_5 x_1 x_3 + \alpha_6 x_2 x_3 + \alpha_7 x_1^2 + \alpha_8 x_2^2 + \alpha_9 x_3^2 \qquad (3)$$

In practice an experimenter would always try to reduce the number of parameters in this model. Usually this is done by performing a sequence of F-tests in order to check which of the parameters are needed in the model. In contrast to this approach we consider here the case where it is already clear that the above model is "incomplete" before any experiments have been conducted. For example, by physical considerations, it could be argued that there is no interaction between x_1 and x_3 and x_2 and x_3. Morover, assume that the response function is increasing with x_2 and x_3 and therefore can not be quadratic in these variables. In this case the model (3) reduces to the "incomplete" second order regression

$$f(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 x_2 + \alpha_5 x_1^2 \,. \qquad (4)$$

In order to construct D-optimal designs for general "incomplete" models of the form (4) we need a reformulation of the response surface model (1) which allows to omit any of the regression functions in (2). To this end let for $h_1, \ldots, h_q \in \{0, \ldots, m\}$ with $\sum_{j=1}^{q} h_j \le m$

$$\mathcal{I}_{h_1,\ldots,h_q} = \begin{cases} 1 & \text{if } \prod_{j=1}^{q} x_j^{h_j} \text{ appears in the} \\ & \text{multivariate polynomial regression} \\ 0 & \text{else} \end{cases} \qquad (5)$$

denote given numbers with values 0 or 1 and define a linear model by

$$g(x) = \sum_{\substack{h_1,\ldots,h_q \in \{0,\ldots,m\} \\ \sum_{j=1}^{q} h_j \le m}} \mathcal{I}_{h_1,\ldots,h_q} \left(\alpha_{h_1,\ldots,h_q} \prod_{j=1}^{q} x_j^{h_j} \right) . \qquad (6)$$

The multivariate polynomial regression (6) has

$$N_{q,m,\mathcal{I}} := \sum_{\substack{h_1,\ldots,h_q \in \{0,\ldots,m\} \\ \sum_{j=1}^{q} h_j \le m}} \mathcal{I}_{h_1,\ldots,h_q}$$

parameters, and the indicator functions $\mathcal{I}_{h_1,\ldots,h_q}$ specify the monomials defined by (2) which appear in the "incomplete" model (6).

Note that the "complete" multivariate regression of degree m is obtained by putting all indicator functions $\mathcal{I}_{h_1,\ldots,h_q} \equiv 1$. As a further illustration recall the "incomplete" second order model (4) of Example 2.1 which emerges from (6) by the special choice $m = 2$, $q = 3$ and

$$\mathcal{I}_{0,0,0} = \mathcal{I}_{1,0,0} = \mathcal{I}_{0,1,0} = \mathcal{I}_{0,0,1} = \mathcal{I}_{1,1,0} = \mathcal{I}_{2,0,0} = 1\,,$$
$$\mathcal{I}_{1,0,1} = \mathcal{I}_{0,1,1} = \mathcal{I}_{0,2,0} = \mathcal{I}_{0,0,2} = 0\,.$$

Throughout this paper we will restrict ourselves to "incomplete" multivariate models which satisfy the following *basic assumption*

$$\text{if} \quad \mathcal{I}_{h_1,\ldots,h_q} = 1\,, \quad \text{then} \quad \mathcal{I}_{h'_1,\ldots,h'_q} = 1 \quad \text{for every set} \quad (h'_1,\ldots,h'_q) \quad \text{such that} \quad h'_j \le h_j \quad \text{and} \quad h_j \quad \text{and} \quad h'_j \quad \text{have the same parity for} \quad j = 1,\ldots,q\,, \tag{7}$$

which turns out to be crucial for our approach. Roughly speaking, it means that if the experimenter deletes a regression function, say x_j^{2i} (x_j^{2i-1}) all terms which contain even (odd) powers of x_j larger or equal than $2i$ ($2i - 1$) have also to be deleted. From a practical point of view this seems to be reasonable and therefore the basic assumption is satisfied for many response surface models used in applied regression analysis. The following examples also demonstrate that it is easy to check in a given regression.

Example 2.2: *The basic assumption (7) is satisfied in the second order response relationship (4) of Example 2.1. Similarly, in the third order "incomplete" model*

$$\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1^2 + \alpha_4 x_1^3 + \alpha_5 x_1^2 x_2 \tag{8}$$

assumption (7) can readily be verified by noting that the quantities $\mathcal{I}_{h_1,h_2}$ in our general "incomplete" model (6) are given by

$$\mathcal{I}_{0,0} = \mathcal{I}_{1,0} = \mathcal{I}_{0,1} = \mathcal{I}_{2,0} = \mathcal{I}_{3,0} = \mathcal{I}_{2,1} = 1\,,$$
$$\mathcal{I}_{0,2} = \mathcal{I}_{1,2} = \mathcal{I}_{0,3} = \mathcal{I}_{1,1} = 0\,.$$

A rather large class of models satisfying (7) is the class of odd or even (multivariate) polynomials. An example where our basic assumption (7) is not fulfilled is the third order one dimensional regression

$$g(x) = \alpha_1 + \alpha_2 x_1^2 + \alpha_3 x_1^3\,.$$

If we would require this model to satisfy (7) we would have to add the term x_1 in the regression.

3. *D*-optimal Designs for "Incomplete" Multivariate Polynomials

Assume that the experimenter observes n independent, normally distributed responses, say $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ with common variance $\sigma^2 > 0$ and mean $I\!E[Y_j] = g(x^{(j)})$, $j = 1, \ldots, n$, where $g(x)$ is the "incomplete" mth degree polynomial defined in (6) and the explanatory variable varies in the q-dimensional cube.

For the sake of simplicity we restrict ourselves at this point to the unit cube $[-1, 1]^q$. The case of a general symmetric cube does not cause any further difficulties and is briefly discussed at the end of this section. If $\boldsymbol{X}^T\boldsymbol{X}$ denotes the design matrix for this model, then the inverse of the volume of the confidence ellipsoid for the vector of parameters is proportional to

$$|\boldsymbol{X}^T\boldsymbol{X}| \tag{9}$$

and consequently a "good" design will make this determinant as large as possible. Designs which maximize $|\boldsymbol{X}^T\boldsymbol{X}|$ are called *D-optimal.*

This paper deals with approximate designs [see KIEFER (1974)] which means that a design is treated as a probability measure on the cube $[-1, 1]^q$ with finite support. Thus an approximate design η requires the obserations to be taken at the support points of the probability measure, say $x^{(1)}, \ldots, x^{(k)}$, and in proportion to the masses $\eta_1, \ldots, \eta_k$ at the corresponding support points. In practice, approximate designs have to be implemented by an appropriate rounding procedure [see e.g. KIEFER (1974)]. For a design η on $[-1, 1]^q$ we define

$$\boldsymbol{M}(\eta) = \int_{[-1,1]^q} g(x)\, g(x)^T \, d\eta\,(x) = \sum_{l=1}^{k} \eta_l g(x^{(l)})\, g(x^{(l)})^T \tag{10}$$

as the *information matrix* of η where the vector of regression functions $g(x)$ in the model (6) consists of the $N_{q,m,\mathcal{I}}$ monomials $\prod_{j=1}^{q} x_j^{h_j}$ satisfying $\sum_{j=1}^{q} h_j \leq m$ and $\mathcal{I}_{h_1,\ldots,h_q} = 1$. The matrix $\boldsymbol{M}(\eta)$ is the continuous analogue of the design matrix $\boldsymbol{X}^T\boldsymbol{X}$, the quantity corresponding to the determinant in (9) is given by

$$|\boldsymbol{M}(\eta)|$$

and an *approximate D-optimal design* maximizes this determinant.

Example 3.1: *As an illustration for the above terminology consider the model (4) of Example 2.1 and a design which takes n_j observation at the points $x^{(j)} = (x_1^{(j)}, x_2^{(j)}, x_3^{(j)})$, $j = 1, \ldots, k$, $\sum_{l=1}^{k} n_l = n$. If η_n denotes the design with masses n_j/n at the points $x^{(j)}$, $j = 1, \ldots, k$, then it is easy to see that the design matrix in the "incomplete" second order regression (4) can be represented as*

$$\boldsymbol{X}^T\boldsymbol{X} = n \cdot \boldsymbol{M}(\eta_n) \tag{11}$$

where the information matrix $\boldsymbol{M}(\eta)$ is defined by (10) and the vector of regression functions is given by $h(x) = (1, x_1, x_2, x_3, x_1x_2, x_1^2)^T$. Note that the relation (11) is exact whenever the masses of η_n are multiples of $1/n$. For example, if $k = 5$, $\eta_n(x^{(j)}) = 1/5$, $j = 1, \ldots, 5$ and 100 observations can be taken, the experimenter observes 20 times at each point. However, if the number of observations is 101, a rounding procedure has to be applied. This could produce a design with 21 observations at $x^{(1)}$ and 20 observations at each of the remaing four points. In this *case the equality (11) would only hold approximately, which explains the notation "approximate" design.*

D-optimal designs for the "complete" polynomial regression (1) [or equivalently for model (6) with $\mathcal{I}_{h_1,\ldots,h_q} = 1$ for all $h_1, \ldots, h_q$] have been determined numerically by FARELL, et al. (1967) $(m = 3, q = 2)$, LIM and STUDDEN (1988) $(m = 3, 4, 5, q = 2; m = q = 3)$ or GAFFKE and HEILIGERS (1995). Even in these models the numerical effort is considerable and usually the optimization problems are reduced to lower dimensional maximization problems by using standard invariance arguments. For "incomplete" multivariate regression models these techniques can not be applied any longer and the numerical difficulties increase rapidly. For this reason DETTE and RÖDER (1996) proposed to determine optimal designs in the subclass of all product designs on the q-dimensional cube. They demonstrated that the D-optimal product designs provide very efficient solutions for the design problem in the "complete" and "incomplete" multivariate polynomial regression model. In the cases where the D-optimal designs in the class of all designs are known, the loss of efficiency by using D-optimal product designs is usually between $1\% - 3\%$ [see LIM and STUDDEN (1988) or DETTE and RÖDER (1996)].

To be precise let $\eta = \xi_1 \times \cdots \times \xi_q$ denote a product design on the q-cube $[-1, 1]^q$ (which means that ξ_j is a probability measure with finite support on the interval $[-1, 1]$, $j = 1, \ldots, q$), and define Ξ_q as the set of all product designs on $[-1, 1]^q$. A D-optimal product design for the "incomplete" polynomial regression (6) is a solution of the problem

$$\textit{maximize} \quad |\boldsymbol{M}(\eta)| \quad \textit{with respect to} \quad \eta \in \Xi_q\,. \tag{12}$$

In order to describe the D-optimal product design for the "incomplete" multivariate regression explicitly we have to characterize the components of the optimal product measure. These can easily be described in terms of their canonical moments. For a precise definition of canonical moments we refer to the work of STUDDEN (1980, 1982a, b) and to the monograph of DETTE and STUDDEN (1997). For the sake of simplicity we present here a rather heuristic description of these quantities which will be sufficient in order to explain our algorithm.

To this end let ξ denote a probability measure on the interval $[-1, 1]$ with moments $c_i = \int_{-1}^{1} x^i \, d\xi(x)$, $i = 1, 2, \ldots$ It is well known that ξ is determined by its moments. The sequence $(c_1, c_2, \ldots)$ can be mapped in a one to one manner onto a sequence $(p_1, p_2, \ldots)$ whose elements vary in the interval $[0, 1]$ and are called *canonical moments* of the measure ξ. Because the design ξ is determined by $(c_1, c_2, \ldots)$ it follows that it is also determined by its sequence of canonical moments, symbolically

$$\xi \leftrightarrow c_1, c_2, \ldots \leftrightarrow p_1, p_2, \ldots\,.$$

If i is the first index for which $p_j \in \{0, 1\}$, then the sequence of canonical moments terminates at p_i and the design ξ is supported at a finite number of points.

The support points of the corresponding measure can be obtained by determining the roots of a certain polynomial. The corresponding masses can be computed by evaluating a second polynomial at these roots (see DETTE and RÖDER (1996), Theorem A.1 for more details). Both polynomials can be calculated recursively. In this sense a design on the interval $[-1, 1]$ corresponding to a terminating sequence of canonical moments has finite support and can quickly be identified by evaluating two polynomials.

Similarly, a product design $\eta = \xi_1 \times \ldots \times \xi_q \in \Xi_q$ is determined by q sequences of canonical moments $\boldsymbol{p}_j = (p_1^{(j)}, p_2^{(j)}, \ldots)$, $j = 1, \ldots, q$ where the jth sequence corresponds to the jth factor of the product design η, i.e.

$$\eta \in \Xi_q \leftrightarrow \xi_1 \times \ldots \times \xi_q \leftrightarrow \boldsymbol{p}_1, \ldots, \boldsymbol{p}_q .$$

It turns out that the D-optimal product design can easily be described by the canonical moments of its factors and that all canonical moment sequences $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_q$ of the D-optimal product measure terminate with $p_{i_j}^{(j)} = 1$ for some $i_j \in N$, $j = 1, \ldots, q$. In order to describe this design explicitly define

$$m_j := \max \{h_j \mid \mathcal{I}_{h_1, \ldots, h_q} = 1\} \tag{13}$$

as the largest exponent of the variable x_j in the "incomplete" regression (6) ($j = 1, \ldots, q$) and

$$\mathcal{I}(i, j) := \sum_{\substack{\sum_{l=1}^q h_l \leq m \\ h_j = i}} \mathcal{I}_{h_1, \ldots, h_q} \tag{14}$$

as the number of terms in the model (6) which are of degree smaller than or equal to m and which contain the factor factor x_j^i ($i = 1, \ldots, m$, $j = 1, \ldots, q$). The following result has been proved by DETTE and RÖDER (1996) and provides a complete solution of the D-optimal product design problem for an "incomplete" multivariate polynomial from a theoretical point of view. It also turns out to be the basic ingredient for our algorithm.

Theorem 3.2: *The D-optimal product design for an "incomplete" polynomial regression (6) satisfying the basic assumption (7) is given by* $\eta^* = \xi_1^* \times \cdots \times \xi_q^*$. *Here, for* $j = 1, \ldots, q$, *the jth factor* ξ_j^* *is a design which is uniquely determined by its canonical moments* $\boldsymbol{p}_j = (p_1^{(j)}, p_2^{(j)}, \ldots)$, *where*

$$p_{2l-1}^{(j)} = \tfrac{1}{2}, \qquad l = 1, \ldots, m_j ,$$

$$p_{2l}^{(j)} = \frac{\sum_{i=l}^{m_j} \mathcal{I}(i, j)}{\sum_{i=l}^{m_j} \mathcal{I}(i, j) + \sum_{i=l+1}^{m_j} \mathcal{I}(i, j)}, \qquad l = 1, \ldots, m_j - 1, \tag{15}$$

$$p_{2m_j}^{(j)} = 1 ,$$

and m_j *and* $\mathcal{I}(i, j)$ *are defined in (13) and (14).*

In the case of a general symmetric cube as design space, say $\mathcal{X} = \bigotimes_{j=1}^{q}[-b_j, b_j]$, a similar anlysis applies. Using well known invariance properties of the D-optimality criterion [see FEDOROV (1972)] it turns out that the factors of the D-optimal product design $\eta^{**} = \xi_1^{**} \times \cdots \times \xi_q^{**}$ on $\mathcal{X}$ can be obtained as follows. In a first step the design problem on the cube $[-1, 1]^q$ is solved. In a second step every factor ξ_j^* of the D-optimal product design $\eta^* = \xi_1^{**} \times \cdots \times \xi_q^*$ on $[-1, 1]^q$ is transformed linearly onto the correponding interval $[-b_j, b_j]$ giving the factor ξ_j^{**}, $j = 1, \ldots, q$. We have implemented this general design space in our program which will be described in the next section.

4. The Algorithm and its Implementation

4.1 *Algorithm*

The program can be devided into two general parts, the graphical representation of the model combined with the output of the results and the calculation of the D-optimal designs including the selection of a special model structure. There is also an option of storing the results in a file, which could be considered as a third part. The general structure of the program is shown in a flowchart in Figure 1. For a more detailed description of the algorithm we restrict ourselves to the "second" part.

Before determining the optimal designs, a concrete "incomplete" polynomial regression model has to be specified by the experimenter. For this reason the algorithm realizes in a first step the assignment of the indicator functions (see (5)), which determine the model (6) completely. This part of the algorithm controls also the conformity of the chosen model with assumption (7) by deleting or adding regression functions such that the final model satisfies (7). If the model has been specified correctly and the boundaries of the q-dimensional cube have been fixed the calculation of the D-optimal product designs is performed. This is done by applying the procedure described in the previous section. In a first step the algorithm calculates the q sequences of canonical moments $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_q$ defined in Theorem 3.2 by equation (15). As pointed out in Section 3 these sequences correspond to the factors $\xi_1^*, \ldots, \xi_q^*$ of the D-optimal product design. Note that in contrast to many other algorithms in optimal design theory these calculations are not performed iteratively and only require

$$\binom{q + m_j - 1}{q} - 1$$

additions and $m_j - 1$ divisions for each of the j factors ($j = 1, \ldots q$). This follows by a straightforward calculation from (13), (14) and (15). The knowledge of the canonical moments of the univariate designs enables us now to specify support points and masses of the corresponding measures explicitly by evaluating two polynomials which can easily be obtained by a recursive procedure [see DETTE

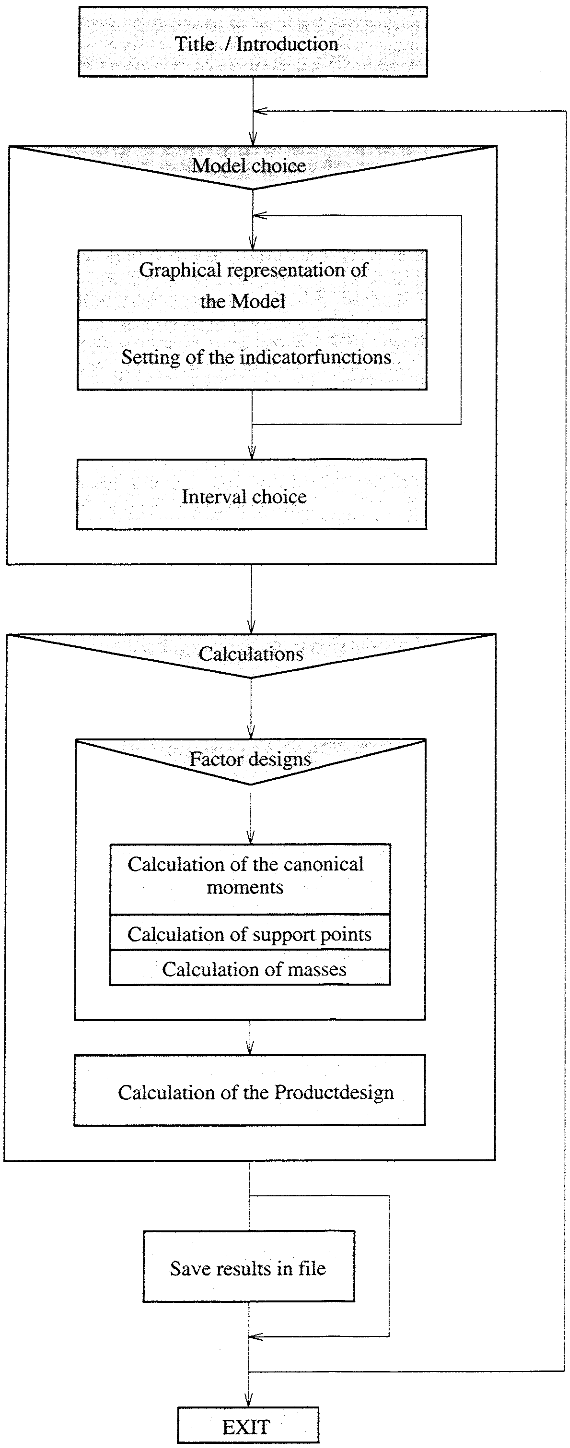

Fig. 1. Simple flowchart of the program OPTIMAL.EXE

and RÖDER (1996)]. If the largest degree of the variable x_j is m_j, the algorithm determines numerically $m_j - 1$ zeros of a symmetric polynomial of degree $m_j - 1$, $j = 1, \ldots, q$. For lower degree explicit formulas for the zeros are also implemented in the algorithm. In a final step a simple multiplication of the probablity masses of the factor designs provides the masses of the D-optimal product design at the appropriate points in the q-dimensional design space.

4.2 *Implementation and application*

The program is written in Turbo Pascal 6.0 under the operating system MS-DOS 6.20. To run our software without problems an IBM compatible PC is required which is operating under (MS-)DOS, has a VGA/EGA graphic device and a mouse installed. All needed units, graphic drivers and fonts are already implemented in the object-code which is available from the authors via ftp[1]. The program is designed for easy and quick use so the handling should be self-explaining. A brief description will be given here.

To start the program simply run OPTIMAL.EXE from the DOS-prompt. The user has now the opportunity to choose a short online introduction/description or to skip this and go to the next screen which provides an environment for the selection of the "incomplete" model. Here a complete polynomial regression model in four variables up to degree four is shown, which is the maximal model that can be dealt with by the present program. The restriction on the number of variables and the degree is not caused by any numerical difficulties. It is introduced due to practical and ergonometric reasons in the presentation of the "complete" model on a computer sceen. In general there is no restriction to the dimension of the cube and the degree of the polynomial. The algorithm can easily be extended to these cases.

The model shows the regession functions of the "complete" model (not the corresponding coefficients) which are arranged in increasing order of degree. That means there are five "sections". The first line contains the intercept, the second line the linear components etc. The last four lines contain the regression functions of order four. By using the mouse it is now possible to adjust the model to the concrete needs of the experimenter. Clicking the left[2] mouse button into a chosen component will remove it from the model. Any canceled component can be got back by simply clicking the right mouse button on it. Using left and right mouse button at the same time removes all components of the corresponding degree. At the bottom of this screen there are two extra buttons which produce an empty or a complete model.

As mentioned in the description of the algorithm the program will control the conformity with the basic assumption (7) automatically. So any misspecification will be corrected immediately by adding or deleting all necessary components in the model.

1 +ftp://ftp.imise.uni-leipzig.de/pub/ingo/program/
2 depending on the configuration of the computer

After the specification of the model the user can decide if the results should be saved in a file (which will be written into the current working directory). The structure of the these files will be illustrated in the example of the following section. In the next step it is requested to choose the regression region. The default (pressing ⟨*ESC*⟩) will result in the cube $[-1, 1]^q$. For more general cubes $\mathcal{X} = \bigotimes_{j=1}^{q} [-b_j, b_j]$ all boundaries of the intervals have to be specified separately.

The following screen shows the factors of the *D*-optimal product design. After that the complete *D*-optimal product design for the chosen model is presented. The last screen (most models produce product designs over several screens) offers the possibility to quit the program (by pressing ⟨*x*⟩) or to select a new model.

5. Example

Consider the "incomplete" second order response surface relationship (3) of Example 2.1 on the three dimensional cube $[0, 1]^3$

$$\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_1 x_2 + \alpha_5 x_1^2 . \tag{16}$$

To select this model in the program the user has to cancel all components of the "complete" polynomial regression except the intercept, x_1, x_2, x_3, $x_1 x_2$ and x_1^2. Alternatively one could also first use the option "empty model" and then add the corresponding components. The resulting model is displayed in Figure 2. In the next step the design space $[0, 1]^3$ has to be specified by entering the boundaries of each interval. The following screen then presents the factors of the *D*-optimal product design, which is illustrated in Figure 3. The final screen (for models with a large number of parameters there can be several pages) will show the *D*-optimal product design in the specified model (see Figure 4).

If the user chooses the file-save option, the output corresponding to Example 2.1 will have the following structure:

components of the chosen model:
(0, 0, 0, 0) (1, 0, 0, 0) (0, 1, 0, 0) (0, 0, 1, 0) (2, 0, 0, 0) (1, 1, 0, 0)

factors of the D-optimal product design:
x1 [0, 1]: (0/37.50) (0.5/25.00) (1/37.50)
x2 [0, 1]: (0/50.00) (1/50.00)
x3 [0, 1]: (0/50.00) (1/50.00)

D-optimal product design:
(0, 0, 0, –/9.38) (0, 0, 1, –/9.38) (0, 1, 0, –/9.38)
(0, 1, 1, –/9.38) (0.5, 0, 0, –/6.25) (0.5, 0, 1, –/6.25)
(0.5, 1, 0, –/6.25) (0.5, 1, 1, –/6.25) (1, 0, 0, –/9.38)
(1, 0, 1, –/9.38) (1, 1, 0, –/9.38) (1, 1, 1, –/9.38)

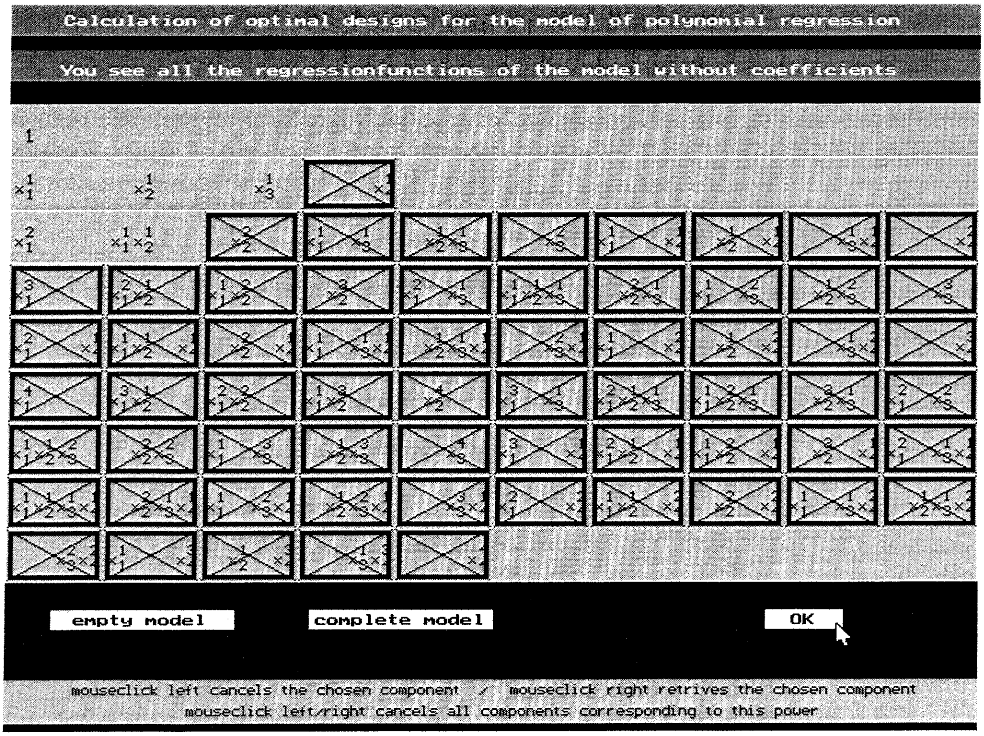

Fig. 2. Screenshot: Selection of the model for Example 2.1

x_1	support-points :	0.00	0.50	1.00
	masses :	37.500	25.000	37.500
x_2	support-points :	0.00	1.00	
	masses :	50.000	50.000	
x_3	support-points :	0.00	1.00	
	masses :	50.000	50.000	
x_4	support-points :	not in the model		
	masses :			

Fig. 3. Screenshot: Factors of the *D*-optimal product design for Example 2.1

x_3	x_2 \ x_1	0.00	0.50	1.00
0.00	0.00	9.375	6.250	9.375
	1.00	9.375	6.250	9.375
1.00	0.00	9.375	6.250	9.375
	1.00	9.375	6.250	9.375

Fig. 4. Screenshot: *D*-optimal product design for Example 2.1

In order to explain the first two lines recall that the model is determined by the indicator functions in (5). Each position inside the brackets represents the power of the corresponding variable. For example (0, 0, 0, 0) and (1, 1, 0, 0) represent the regression functions $1 = x_1^0 x_2^0 x_3^0 x_4^0$ and $x_1 x_2 = x_1^1 x_2^1 x_3^0 x_4^0$, respectively. In the following lines the factors of the *D*-optimal product design are given in the form:

variable [*regression region*] : (*support point* / *probability mass in % at this point*) .

For example, in the "incomplete" second order model (3) the first factor of the *D*-optimal product design has mass 37.50% at the point 0. Finally, the product designs are represented as follows:

(*components of the support point* / *percent of observations to be taken at this point*) .

For example, in the "incomplete" second order model (3) the *D*-optimal product design advises the experimenter to take 9.38% of the observations at the origin.

It might also be of interest to compare the design for the "incomplete" model with the optimal product design for the "complete" second order response surface relationship (3) which is given in Figure 5. We observe that this design advises

x_3	x_2 \ x_1	0.00	0.50	1.00
0.00	0.00	6.400	3.200	6.400
	0.50	3.200	1.600	3.200
	1.00	6.400	3.200	6.400
0.50	0.00	3.200	1.600	3.200
	0.50	1.600	0.800	1.600
	1.00	3.200	1.600	3.200
1.00	0.00	6.400	3.200	6.400
	0.50	3.200	1.600	3.200
	1.00	6.400	3.200	6.400

Fig. 5. Screenshot: *D*-optimal product design for the "complete" second order model (3)

the experimenter to take observations at 27 different points. The optimal design for the "incomplete" model only needs 12 points, because the parameters corresponding to the quadratic terms of variable x_2 and x_3 do not have to be estimated.

References

GAFFKE, N. and HEILIGERS, B., 1995: Computing Optimal Approximate Invariant Designs for Cubic Regression on Multidimesional Balls and Cubes. *J. Statistical Planning and Inference* **47**, 347–376

DETTE, H. and RÖDER, I., 1996: Optimal Product Designs for Multivariate Regression with missing Terms. *Scand. J. Statist.* **23**, 195–208.

DETTE, H. and STUDDEN, W. J., 1998: *The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis.* Wiley, New York.

FEDOROV, V. V., 1972: *Theory of Optimal Experiments.* Academic Press, New York.

FARRELL, H. R., KIEFER, J. and WALBRAN, J., 1967: Optimum Multivariate Designs. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.*, Univ. California Press 1, 113–138.

KIEFER, J., 1974: General Equivalence Theory for Optimum Designs. *The Annals of Statistics* **2**, 849–879.

KONO, K., 1962: Designs for Quadratic Regression on the k-Cube. *Mem. Faqc. Sci. Kyushu Univ. Ser.* A **16**, 114–122.

LIM, Y. B. and STUDDEN, W. J., 1988: Efficient D_s-Optimal Designs for Multivariate Polynomial Regression on the q-Cube. *The Annals of Statistics* **16**, 1225–1240.

RAFAJLOWICZ, E. and MYSZKA, W., 1988: Optimum Experimental Design for a Regression on a Hypercube – Generalization of Hoel's Result. *Ann. Inst. Statist. Math.* **40**, 821–828.

RAFAJLOWICZ, E. and MYSZKA, W., 1992: When Product Type Experimental Design is Optimal? Brief Survey and New Results. *Metrika* **39**, 321–333.

STUDDEN W. J., 1980: D_s-Optimal Designs for Polynomial Regression Using Continued Fractions. *The Annals of Statistics* **8**, 1132–1141.

STUDDEN, W. J., 1982a: Some Robust-Type D-Optimal Designs in Polynomial Regression. *Journal of the American Statistical Association* **77**, 916–921.

STUDDEN, W. J., 1982b: Optimal Designs for Weighted Polynomial Regression Using Canonical Moments. In: S. S. Gupta and J. O. Berger (eds.): *Statistical Decision Theory and Related Topics III 2*. Academic Press, New York., 335–350.

UPPERMANN, P. M., 1993: *Designs With a Small Number of Runs for Factorial Experiments.* Dissertation, Technische Universiteit Eindhoven.

WONG, W. K., 1994: G-optimal Designs for Multi-factor Experiments with heteroscedastic Errors. *Journal of Statistical Planning and Inference* **40**, 127–133.

Prof. Dr. HOLGER DETTE
Fakultät und Institut für Mathematik
Ruhr-Universität Bochum Universitätsstr. 150, Gebäude NA
44780 Bochum
Germany
Email: holger.dette@ruhr-uni-bochum.de

Received, March 1999
Revised, November 1999
Accepted, January 2000

Dipl.-Math. INGO ROEDER
Institut für Medizinische Informatik, Statistik und Epidemiologie
Universität Leipzig
Liebigstr. 27
04103 Leipzig
Germany
Email: Roeder@imise.uni-leipzig.de

UNIVERSITÄTSBIBLIOT
HANNOVER UND
TECHNISCHE
INFORMATIONSBIBLIO

Neue Preise für Dienstleistungen der UB/TIB ab 1. September 2000

- aufgrund des neuen Gesamtvertrages „Direktkopienversand" zwischen Bund/Lä
und der VG Wort -

Verehrte Kundschaft der UB/TIB,
Sehr geehrte Damen und Herren,

Sie kennen uns als kompetenten und zuverlässigen Dokumentlieferanten mit überdurchschnittlichem Service und fairem Preis/Leistungsverhältnis. Unser Dienstleistungsangebot orientiert sich stets an den Wünschen und Bedürfnissen u Kunden unter dem Einsatz neuester Techniken.

Eine solche Flexibilität ist uns bei der Preisgestaltung aufgrund gesetzlicher Vorga jedoch nicht möglich. Nach neuesten urheberrechtlichen Bestimmungen muss für Dokumentlieferung (Aufsatzkopie) an Sie ab dem 1. September 2000 eine Vergütu die *Verwertungsgesellschaft Wort (*s.a.: http://www.bdbibl.de/dbv/urheberrecht.htm http://www.vgwort.de) abgeführt werden (sog. „Verlegertantieme"). Das bedeutet, c unsere Preise, die wir über Jahre für Sie konstant halten konnten, neu kalkulieren anpassen mussten.

Unserer neuen Preistabelle auf der Rückseite dieses Schreibens können Sie entr wie sich die Preiserhöhungen auswirken: am deutlich günstigsten bestellen Sie zu über unseren elektronischen Dokumentlieferdienst **TIBORDER**-Online (Internet-Adresse:http://tiborder.tib.uni-hannover.de), über den Sie Ihre benötigten Veröffentlichungen auf dem schnellsten Wege direkt erhalten.

Selbstverständlich stehen Ihnen auch weiterhin die konventionellen Bestellwege p Fax, vorab gekaufter Bestellscheine oder formloser e-mail zur Verfügung.
Wir möchten Sie auch weiterhin als zufriedene Kunden bedienen und bitten um Ih Verständnis.

Mit freundlichen Grüßen

Ihre UB/TIB Hannover

Forschungseinrichtungen			Untern./Selbst., Privatpersonen								
Normal	Eilig	Sehr eilig	Normal	Eilig	Sehr eilig	Normal	Eilig	Sehr eilig	Normal	Eilig	Sehr eilig
7	19	31	25	37	49	25	37	49	25	37	49
10	22	34	28	40	62	28	40	52	28	40	52
12	24	36	33	45	57	40	52	64	40	52	64
13	25	37	23	35	47	23	35	47	43	55	67

e konventionell aufgeben, also per Post, Fax, formloser e-mail, vorab gekaufter Bestellscheine

Deutschland: Hochschulen u. Forschungseinrichtungen			Deutschland: Gewerbliche Untern./Selbst., Privatpersonen			Europa: alle Kunden			Übersee: alle Kunden		
Normal	Eilig	Sehr eilig	Normal	Eilig	Sehr eilig	Normal	Eilig	Sehr eilig	Normal	Eilig	Sehr eilig
12	24	36	30	42	54	30	42	54	30	42	54
15	27	39	33	45	57	33	45	57	33	45	57
17	29	41	38	50	62	45	57	69	45	57	69
18	30	42	28	40	52	28	40	52	48	60	72