The Annals of Statistics
1997, Vol. 25, No. 3, 1161–1175

OPTIMAL DISCRIMINATION DESIGNS FOR MULTIFACTOR EXPERIMENTS

By Holger Dette[1] and Ingo Röder

Ruhr-Universität Bochum and Universität Leipzig

In this paper efficient designs are determined when Anderson's procedure is applied in order to identify the degree of a multivariate polynomial regression model. It is shown that the optimal designs are very closely related to model robust designs which maximize a weighted p-mean of D-efficiencies. As a consequence we obtain designs with high efficiency for model discrimination and for the statistical analysis in the identified model.

1. Introduction. The construction of optimal designs for multifactor experiments has received considerable attention in the recent literature [Farell, Kiefer and Walbran (1967), Lim and Studden (1988), Rafajlowicz and Myszka (1988, 1992) and Wong (1994)]. While most of this work considers the problem of designing experiments for a given model, the optimal design problem for the identification of the relevant parameters in a multifactor model is more complicated and has not been discussed so far. In this paper efficient designs are constructed for discrimination between several multifactor models. More precisely, we concentrate on a multivariate regression of degree m in q variables,

$$(1.1)\qquad \begin{aligned} h_m(x) = a_0 + \sum_{i=1}^{q} a_i x_i + \sum_{1\le i_1\le i_2\le q} a_{i_1,i_2} x_{i_1} x_{i_2} \\ + \cdots + \sum_{1\le i_1\le \cdots \le i_m\le q} a_{i_1,\dots,i_m} x_{i_1}\cdots x_{i_m}, \end{aligned}$$

where $x = (x_1,\dots,x_q)^T$ denotes the independent variable which varies in the design space $\mathscr{X} = [-1,1]^q$. The regression functions in the model (1.1) are the $N_{q,m} := (q+1)\cdots(q+m)/m!$ multiple monomials of the form $\prod_{i=1}^q x_i^{m_i}$, where $\sum_{i=1}^q m_i \le m$. A first problem in regression analysis is to determine the degree of the regression which adequately fits the data.

Assume that the experimenter observes independent, normally distributed responses $Y_1,\dots,Y_n$ with common variance $\sigma^2 \ge 0$ and mean $\mathbb{E}[Y_j] = g(x_j)$, $j = 1,\dots,n$, where $g(x)$ is a regression function which belongs to the class of

---

Received September 1995; revised July 1996.

[1] Research partially supported by the Deutsche Forschungsgemeinschaft.

AMS 1991 *subject classifications*. Primary 62K05; secondary 62G10.

Key words and phrases. Multifactor experiments, model discrimination, optimal designs, invariance.

multivariate polynomials up to degree m, that is,

$$g \in \mathscr{F}_{m,q} = \{h_1, \dots, h_m\}. \tag{1.2}$$

Anderson (1962) studied the following decision rule for the determination of the degree of the regression. For a given set of levels $(\alpha_1, \dots, \alpha_m)$ choose the largest integer l in $\{1, \dots, m\}$, for which the F-test in the model $h_l(x)$ rejects the hypothesis

$$H_0^{(l)}\colon a_{i_1,\dots,i_l} = 0 \quad \text{for all } 1 \le i_1 \le \cdots \le i_l \le q \tag{1.3}$$

at the level α_l. Note that the coefficients specified in (1.3) correspond to the multiple monomials $\prod_{i=1}^q x^{m_i}$ with exact degree $l = \sum_{i=1}^q m_i$. Anderson's procedure satisfies several optimality properties [see Anderson (1962) or Spruill (1990)]. Roughly speaking, it minimizes the probability of the error of choosing a too high degree polynomial.

The F-test for the hypothesis $H_0^{(l)}$ can be obtained from the confidence ellipsoid $\mathscr{E}_l$ for $\mathbf{K}_l^T b_l$, where $\mathbf{K}_l$ is the parameter matrix corresponding to the hypothesis $H_0^{(l)}$ and b_l is the vector of parameters in the model of degree l. The hypothesis $H_0^{(l)}$ is rejected if and only if $0 \notin \mathscr{E}_l$. The volume of the ellipsoid is proportional to $\delta_l^{-1/2} = |\mathbf{K}_l^T(\mathbf{X}_l^T\mathbf{X}_l)^{-1}\mathbf{K}_l|^{1/2}$, where $\mathbf{X}_l^T\mathbf{X}_l$ denotes the design matrix. The quality of the test is improved if the volume of the ellipsoid is reduced and consequently a "good" discrimination design will make $\delta_1, \dots, \delta_m$ as large as possible. In the cases of practical interest a simultaneous maximization of these quantities is not possible and it is common practice to maximize real-valued functions of $\delta_1, \dots, \delta_m$ with respect to the choice of the design which are called *optimality criteria*.

This paper deals with the approximate theory of optimal design [see Kiefer (1974)], which means that a design is treated as a probability measure on the cube $[-1,1]^d$ with finite support. Thus an approximate design η requires the observations to be taken at the support points of the probability measure and in proportion to the masses at the corresponding support points. If η denotes a design on the cube $[-1,1]^q$ we define

$$\mathbf{M}_l(\eta) = \int_{[-1,1]^q} f_l(x) f_l(x)^T \, d\eta(x) \in \mathbb{R}^{N_{q,l} \times N_{q,l}}, \qquad l = 1, \dots, m,$$

as the information matrix of η, where $f_l(x)$ denotes the vector of the $N_{q,l}$ multiple monomials $\prod_{i=1}^q x_i^{m_i}$ of degree less than or equal to l; $\mathbf{M}_l(\eta)$ is the continuous analogue of the design matrix $\mathbf{X}_l^T\mathbf{X}_l$, and the quantities corresponding to the determinants δ_l are given by

$$\delta_l(\eta) = \left|\mathbf{K}_l^T \mathbf{M}_l(\eta)^{-1} \mathbf{K}_l\right|^{-1} = \frac{\det \mathbf{M}_l(\eta)}{\det \mathbf{M}_{l-1}(\eta)}, \qquad l = 1, \dots, m. \tag{1.4}$$

Following the discussion of the previous paragraph a "good" approximate discrimination design η maximizes a real-valued function of $\delta_1(\eta), \dots, \delta_m(\eta)$.

In this paper we determine efficient designs with respect to various optimality criteria based on p-means of $\delta_1(\eta), \dots, \delta_m(\eta)$. We restrict our investigations to the class of product designs [see Lim and Studden (1988)] which is introduced in Section 2. This set is not convex, thus standard invariance arguments for multifactor experiments [see, e.g., Kiefer (1974) or Lim and Studden (1988)] cannot be applied for the reduction of the dimensionality of the optimization problem. Nevertheless, we will demonstrate that for many optimality criteria a permutation invariance property can still be established. Our argument is based on the theory of canonical moments which was introduced by Studden (1980) and allows us to consider the optimization problem in the (nonconvex) class of product measures as a maximization of a concave function over a convex space. This enables us to reduce the optimization to a composite design problem for a univariate polynomial regression which is then solved in Section 3.

Finally, it is shown that for $q \geq 2$ the factors of the optimal discrimination design are precisely the factors of a model robust design for multivariate regression models up to degree m in $q-1$ variables. This result generalizes recent findings of Dette and Studden (1995) to the present multivariate case and also explains the relation between the D- and D_1-optimal product designs found by Lim and Studden (1988). Some examples for discrimination designs between linear, quadratic and cubic multivariate regression are presented in Section 4. The determined designs are highly efficient for model discrimination and also have excellent efficiencies for the statistical analysis in the identified model.

2. Product designs, the optimality criterion and permutation invariance. For moderate degree the situation is complicated and numerical methods are applied [see Farell, Kiefer, Walbran (1967)] for determining D-optimal designs. However, as pointed out by Lim and Studden (1988), these algorithms fail to converge if the number of parameters in the model is too large. For this reason these authors restricted the optimization to the class of all product designs, say Ξ, on $[-1, 1]^q$. Recalling the definition of $\delta_l(\eta)$, in (1.4) we call a (product) design $\eta \in \Xi$ a Φ_p-*optimal discrimination design with respect to the prior* $\beta = (\beta_1, \dots, \beta_m)$ if η maximizes the function

$$\Phi_p(\eta) = \left[\sum_{l=1}^{m} \beta_l \left\{\mathrm{eff}_l^{D_1}(\eta)\right\}^p\right]^{1/p}. \tag{2.1}$$

Here $-\infty \leq p \leq 0$ and the prior $\beta = (\beta_1, \dots, \beta_m)$ reflects the experimenter's belief about the adequacy of the different models $h_1(x), \dots, h_m(x)$. The quantity

$$\mathrm{eff}_l^{D_1}(\eta) = \left(\frac{\delta_l(\eta)}{\max_{\mu \in \Xi} \delta_l(\mu)}\right)^{1/N_{q-1,l}} \tag{2.2}$$

is called D_1-*efficiency* and measures the loss of efficiency when the product design η is used instead of the optimal product design maximizing $\delta_l(\mu)$. In order to simplify the notation we define

$$s_l = \frac{p}{N_{q-1,l}}, \qquad l = 1, \dots, m, \tag{2.3}$$

$$\gamma_l = \beta_l \Big(\max_{\mu \in \Xi} \delta_l(\mu) \Big)^{-s_l}, \qquad l = 1, \dots, m, \tag{2.4}$$

and obtain

$$\Phi_p(\eta) = \left(\sum_{l=1}^{m} \gamma_l [\delta_l(\eta)]^{s_l} \right)^{1/p}. \tag{2.5}$$

The cases $p = 0$ and $p = -\infty$ are understood as the corresponding limits [see Dette and Studden (1995)]. We further remark that Φ_p is a concave function on the set of all probability measures on the q-dimensional cube $[-1, 1]^q$ and that the set Ξ of product designs is *not* convex.

A standard tool for the reduction of the dimensionality of the design problem in a multivariate regression are invariance arguments [see Kiefer (1974)]. More precisely, let $\pi = (\pi_1, \dots, \pi_q)$ denote a permutation of $\{1, \dots, q\}$ and define $\pi x = (x_{\pi_1}, \dots, x_{\pi_q})$. Then π induces an obvious permutation of the factors of the product design $\eta = \xi_1 \times \cdots \times \xi_q$, that is, $\eta_\pi = \xi_{\pi_1} \times \cdots \times \xi_{\pi_q}$. Because $\delta_l(\eta)$ is a ratio of determinants [see (1.4)] it follows that $\delta_l(\eta_\pi) = \delta_l(\eta)$, $l = 1, \dots, m$, and consequently we obtain

$$\Phi_p(\eta_\pi) = \Phi_p(\eta). \tag{2.6}$$

Unfortunately, the standard argument for proving the existence of a permutation-invariant optimal product design is not available, because the set Ξ of product measures is not convex. Nevertheless, the existence of a permutation-invariant product design can still be established by mapping the nonconvex set of product measures Ξ in a one-to-one manner onto a convex set.

To this end we need some basic facts about the theory of canonical moments [see Studden (1980)]. For a probability measure on the interval $[-1, 1]$ with moments $c_j = \int_{-1}^{1} x^j \, d\xi(x)$, $j = 1, 2, \dots$, define c_j^+ as the maximum of the jth moment $\int_{-1}^{1} x^j \, d\mu(x)$ over the set of all probability measures μ having the given moments $c_1, \dots, c_{j-1}$. Similarly, let c_j^- denote the corresponding minimum. If $c_j^- < c_j^+$, then the jth canonical moment is defined by

$$p_j = \frac{c_j - c_j^-}{c_j^+ - c_j^-}, \qquad j = 1, 2, \dots,$$

otherwise $(c_j^- = c_j^+)$ it is left undefined. Note that $0 \le p_j \le 1$ whenever the canonical moments are defined. If i is the first index for which $c_j^- = c_j^+$, then $0 < p_j < 1$ for $j = 1, \dots, i-2$, p_{i-1} must have the value 0 or 1 and the design ξ is supported at a finite number of points [see Skibinsky (1986)]. It is also shown in Skibinsky (1986) that there is a one-to-one correspondence between the set of probability measures on the interval $[-1, 1]$ and the set of canonical moment sequences. For more details about the theory of canonical moments the reader is referred to the recent monograph of Dette and Studden (1997).

If $\eta = \xi_1 \times \cdots \times \xi_q$ is a product design on the cube $[-1,1]^q$, let $p_1^{(j)}, p_2^{(j)}, \dots$ denote the canonical moments of the jth factor ξ_j, $j = 1, \dots, q$, and let $W_{i(j)}(t)$ denote the ith monic orthogonal polynomial with respect to the measure $d\xi_j(t)$. Further, for $j = 1, \dots, q$, let $\kappa_j = (p_1^{(j)}, \dots, p_{2m}^{(j)})$ denote the vector of canonical moments of ξ_j up to the order $2m$, $j = 1, \dots, q$, and define $\kappa = (\kappa_1, \dots, \kappa_q)$. Applying (1.4) and Lemmas 4.3 and 5.1 of Lim and Studden (1988) we obtain, for $l = 1, \dots, m$,

$$\begin{aligned} \delta_l(\eta) &= \prod_{j=1}^{q} \prod_{i=1}^{l} \left[\int_{-1}^{1} W_{i(j)}^2(t)\, d\xi_j(t) \right]^{N_{q-2,\, l-i}} \\ &= C_{l,q} \prod_{j=1}^{q} \prod_{i=1}^{l} \left[\prod_{k=1}^{i} q_{2k-2}^{(j)} p_{2k-1}^{(j)} q_{2k-1}^{(j)} p_{2k}^{(j)} \right]^{N_{q-2,\, l-i}} =: \delta_l(\kappa), \end{aligned} \tag{2.7}$$

where $q_0^{(j)} = 1$, $j = 1, \dots, q$, and $q_i^{(j)} = 1 - p_i^{(j)}$, $j = 1, \dots, q$, $i \geq 1$. Note that $N_{-1,j} = 0$, $1 \leq j \leq m$, $N_{-1,0} = 1$ and that the constant in (2.7) is given by

$$C_{l,q} = 2^{q \sum_{i=1}^{l} 2i N_{q-2,\, l-i}}, \qquad l = 1, \dots, m. \tag{2.8}$$

Observing (2.7) it follows that the function $\Phi_p(\eta)$ depends on $\eta = \xi_1 \times \cdots \times \xi_q$ only through the canonical moments of $\xi_1, \dots, \xi_q$ up to the order $2m$. Moreover, an optimal product design must satisfy $p_i^{(j)} \in (0,1)$, $j = 1, \dots, q$, $i = 1, \dots, 2m-1$. Because of the one-to-one correspondence between probability measures on the interval $[-1,1]$ and canonical moment sequences we can now consider the optimality criterion Φ_p as a concave function on the convex set $\mathscr{B} = \{(0,1)^{2m-1} \times [0,1]\}^q$. If $\kappa = (\kappa_1, \dots, \kappa_q) \in \mathscr{B}$ is a vector of canonical moment sequences (each component of length $2m$) corresponding to a product design $\eta = \xi_1 \times \cdots \times \xi_q$ [i.e., $\kappa_j = (p_1^{(j)}, \dots, p_{2m}^{(j)})$ are the first $2m$ canonical moments of the factor ξ_j, $j = 1, \dots, q$], then a permutation π on Ξ induces an obvious permutation on $\mathscr{B}$. More precisely, if $\eta_\pi = \xi_{\pi_1} \times \cdots \times \xi_{\pi_q}$, then we define $\kappa_\pi = (\kappa_{\pi_1}, \dots, \kappa_{\pi_q})$, which means that we are permuting the canonical moment sequences κ_j with their corresponding factors ξ_j, $j = 1, \dots, q$.

THEOREM 2.1. *The Φ_p-optimal product design η_p is invariant with respect to permutations of its factors, that is, $\eta_p = \xi_p \times \cdots \times \xi_p$, where ξ_p is a design on the interval $[-1,1]$.*

PROOF. Let $-\infty < p < 0$. By the preceding discussion the maximization of Φ_p in Ξ is equivalent to the minimization of the function

$$H(\kappa) = \sum_{l=1}^{m} \gamma_l [\delta_l(\kappa)]^{s_l} = [\Phi_p(\eta)]^p \tag{2.9}$$

on the cube $\mathscr{B}$. Observing (2.7) it is easy to see that the function $\delta_l(\kappa)^{-s_l}$ is strictly logarithmic concave. By taking second derivatives it therefore follows that $\delta_l(\kappa)^{s_l}$ is strictly convex on $\mathscr{B}$. Thus the function $H(\kappa)$ is strictly convex on $\mathscr{B}$ and satisfies [by (2.6)] $H(\kappa_\pi) = H(\kappa)$ for any permutation π of $(1, \dots, q)$. A standard argument in decision theory now shows that there exists a unique minimum of H at some point $(\kappa_*, \dots, \kappa_*) \in \mathscr{B}$. The function $H(\kappa)$ is decreasing in δ_l, $l = 1, \dots, m$. Therefore we obtain from (2.7) that the $2m$th entry of $\kappa_* = (p_1, \dots, p_{2m})$ has to be $p_{2m} = 1$ [in order to maximize $\delta_m(\eta)$], which means that the design ξ_p corresponding to κ_* is unique and supported at $m+1$ points including -1 and 1 [see Skibinsky (1986)]. Consequently the optimal product design is unique and of the form $\xi_p \times \cdots \times \xi_p$, which proves the assertion for $-\infty < p < 0$. The cases $p = 0$ follows directly from (2.7) while the case $p = -\infty$ is obtained by the same reasoning, observing that the function

$$H_{-\infty}(\kappa) = \Phi_{-\infty}(\eta) = \min\left\{\mathrm{eff}_l^{D_1}(\eta) \mid \beta_l > 0\right\}$$

is logarithmic concave on the set $\mathscr{B}$. □

REMARK 2.2. Note that p-means are usually defined for $-\infty \le p \le 1$ [see Pukelsheim (1993)] and it is natural to ask if there is a similar invariance property in the case $0 < p \le 1$. Unfortunately this question cannot be answered in general. The reason is that the functions $[H(\kappa)]^{1/p}$ or $H(\kappa)$ (which has to be maximized for $0 < p \le 1$) is not necessarily concave on the $2mq$-dimensional cube $\mathscr{B}$. Therefore the argument used in the proof of Theorem 2.1 does not work any longer. As an example consider the case $m = 2$, where $s_1 = p/q$, $s_2 = 2p/(q(q+1))$. Define $\kappa^x = (\kappa_1^x, \dots, \kappa_q^x) \in [0,1]^{4q}$, where $\kappa_j^x = (1/2, x, 1/2, 1)$, $j = 1, \dots, q$, and $x \in (0,1)$. If $[H(\kappa)]^{1/p}$ were concave on $[0,1]^{4q}$, then it follows that the function

$$h(x) = \left[H(\kappa^x)\right]^{1/p} = \left[\gamma_1 x^p + \gamma_2 x^{2pq/(q+1)}(1-x)^{2p/(q+1)}\right]^{1/p}$$

would also be concave on $(0,1)$. But this is not true in general (e.g., $\gamma_1 = 1 - \gamma_2 = 0.1$, $q = 5$, $p = 3/4$).

The discussion of the previous paragraph reduced the discrimination design problem in the class of all product designs on the cube $[-1,1]^q$ to the problem of maximizing a function in the set $\mathscr{P}$ of all designs on the interval $[-1,1]$. Consequently, we can now restrict ourselves to product designs of the form $\eta_\xi = \xi \times \cdots \times \xi$, where $\xi \in \mathscr{P}$. Let $\Delta_l(\xi)$ denote the information matrix of a design $\xi \in \mathscr{P}$ in a univariate polynomial regression of degree l, that is,

$$\Delta_l(\xi) = \int_{-1}^{1} (1, t, \dots, t^l)^T (1, t, \dots, t^l)\, d\xi(t), \qquad l = 1, \dots, m.$$

It follows from Lemma 4.3 and formula (4.5) in Lim and Studden (1988) that

$$(2.10)\quad \int_{-1}^{1} W_{i(j)}^2(t)\, d\xi_j(t) = \frac{\det \Delta_i(\xi_j)}{\det \Delta_{i-1}(\xi_j)} = 2^{2i} \prod_{k=1}^{i} \left(q_{2k-2}^{(j)} p_{2k-1}^{(j)} q_{2k-1}^{(j)} p_{2k}^{(j)} \right)$$

and for designs $\eta_\xi = \xi \times \cdots \times \xi$ the optimality criterion $\Phi_p(\eta)$ can be rewritten as [using the representation (2.7)]

$$(2.11)\quad \begin{aligned} \Phi_p(\eta_\xi) &= \left(\sum_{l=1}^{m} \gamma_l [\delta_l(\eta_\xi)]^{s_l} \right)^{1/p} \\ &= \left[\sum_{l=1}^{m} \gamma_l \prod_{i=1}^{l} \left(\frac{\det \Delta_i(\xi)}{\det \Delta_{i-1}(\xi)} \right)^{q s_l N_{q-2,\, l-i}} \right]^{1/p} =: F_p^\gamma(\xi). \end{aligned}$$

In the case $p = 0$, (2.11) gives

$$(2.12)\quad F_0^\gamma(\xi) = \prod_{j=1}^{m} \left(\frac{\det \Delta_j(\xi)}{\det \Delta_{j-1}(\xi)} \right)^{\bar\gamma_j} = \prod_{j=1}^{m} \left(\prod_{i=1}^{j} q_{2i-2} p_{2i-1} q_{2i-1} p_{2i} \right)^{\bar\gamma_j},$$

where

$$(2.13)\quad \bar\gamma_j = q \sum_{l=j}^{m} \frac{N_{q-2,\, l-j}}{N_{q-1,\, l}} \gamma_l, \qquad j = 1, \dots, m,$$

and $p_1, p_2, \dots$ are the canonical moments of ξ. This allows an elementary maximization in terms of canonical moments. For example, the design $\eta = \xi \times \cdots \times \xi$ maximizing $\delta_m(\eta)$ is obtained by setting $\gamma_1 = \cdots = \gamma_{m-1} = 0$, $\gamma_m > 0$ and satisfies

$$p_{2i-1} = \tfrac{1}{2}, \qquad p_{2i} = \frac{q-1+m-i}{q-1+2(m-i)}, \qquad i = 1, \dots, m$$

[see Lim and Studden (1988)]. This gives [observing (2.7) and (2.8)]

$$(2.14)\quad \begin{aligned} \delta_{m,q}^* &= \max_{\mu \in \Xi} \delta_m(\mu) = \max_{\xi \in \mathscr{P}} \delta_m(\eta_\xi) \\ &= \left(\frac{q-2+m}{q-3+2m} \right)^{q N_{q-1,\, m-1}} \\ &\quad \times \prod_{k=2}^{m} \left(\frac{(m-k+1)(q-1+m-k)}{(q+1+2(m-k))(q-1+2(m-k))} \right)^{q N_{q-1,\, m-k}}. \end{aligned}$$

The maximization of Φ_p in the general case is complicated, and more sophisticated tools are required for that purpose. The following result is the main tool for deriving optimal product designs for the criterion (2.6). The proof is obtained by an application of results in Pukelsheim (1993) and can be found in a technical report by Dette and Röder (1995).

THEOREM 2.3. *For* $-\infty < p \le 1$ *a design* ξ *on the interval* $[-1,1]$ *maximizes* F_p^{γ} *if and only if it maximizes* $F_0^{\gamma^*}$ *for the weights* $\gamma^* = (\gamma_1^*, \dots, \gamma_m^*)$, *where* γ_l^* *is proportional to*

$$(2.15)\qquad \gamma_l \prod_{j=1}^{l} \left(\frac{\det \Delta_j(\xi)}{\det \Delta_{j-1}(\xi)} \right)^{q s_l N_{q-2,l-j}}, \qquad l = 1, \dots, m.$$

Let

$$\mathscr{N}(\xi) = \Bigg\{ l \,|\, \beta_l > 0, \left[\frac{1}{\delta_{l,q}^*} \prod_{j=1}^{l} \left(\frac{\det \Delta_j(\xi)}{\det \Delta_{j-1}(\xi)} \right)^{q N_{q-2,l-j}} \right]^{1/N_{q-1,l}} = \min_{i=1}^{l} \left[\frac{1}{\delta_{i,q}^*} \prod_{j=1}^{i} \left(\frac{\det \Delta_j(\xi)}{\det \Delta_{j-1}(\xi)} \right)^{q N_{q-2,i-j}} \right]^{1/N_{q-1,i}} \Bigg\},$$

where $\delta_{i,q}^*$ *is defined by* (2.14). *A design* ξ *maximizes* $F_{-\infty}$ *if and only if there exist weights* $\gamma^* = (\gamma_1^*, \dots, \gamma_m^*)$ *with* $\gamma_j^* = 0$ *if* $j \notin \mathscr{N}(\xi)$ *such that* ξ *maximizes* $F_0^{\gamma^*}$.

3. Φ_p-optimal discrimination and model robust designs. In this section we determine the optimal product designs with respect to the criterion defined in (2.1). Note that the maximization of (2.1) is equivalent to the maximization of (2.5). This optimization problem was reduced by an invariance argument to a maximization problem in the class of all probability measures on the compact interval $[-1,1]$.

THEOREM 3.1. *Let* s_l *and* $\delta_{l,q}^*$ *be given by* (2.3) *and* (2.14), *respectively*, $l = 1, \dots, m$. *If* $-\infty < p \le 0$, *then the* Φ_p-*optimal discrimination design is given by* $\eta_p = \xi_p \times \cdots \times \xi_p$; ξ_p *is a symmetric design on the interval* $[-1,1]$ *uniquely determined by its canonical moments* $p_1, \dots, p_{2m-1}, p_{2m} = 1$, *where* $p_{2j-1} = 1/2$, $j = 1, \dots, m$, *and* $(p_2, \dots, p_{2m-2}) \in (0,1)^{m-1}$ *is the unique solution of the system of* $m - 1$ *equations*

$$(3.1)\qquad \frac{(l+1)\sum_{j=0}^{m-l}(-1)^j \binom{q-1}{j} \prod_{i=1}^{j+l-1} (q_{2i}/p_{2i}) \big(1 - q_{2(j+l)}/p_{2(j+l)}\big)}{(q+l)\sum_{j=0}^{m-l-1}(-1)^j \binom{q-1}{j} \prod_{i=1}^{j+l} (q_{2i}/p_{2i}) \big(1 - q_{2(j+l+1)}/p_{2(j+l+1)}\big)} = \frac{\beta_l}{\beta_{l+1}} \frac{\left(\delta_{l+1,q}^*\right)^{s_{l+1}}}{\left(\delta_{l,q}^*\right)^{q_l}} \prod_{j=1}^{l+1} \left(q_{2j-2}\, p_{2j}\right)^{-jpN_{q-2,l-j+1}/N_{q,l}},$$

$l = 1, \dots, m-1$, *such that*

$$(3.2) \qquad \sum_{j=0}^{m-l} (-1)^j \binom{q-1}{j} \prod_{i=1}^{l+j-1} \frac{q_{2i}}{p_{2i}} \left(1 - \frac{q_{2(l+j)}}{p_{2(l+j)}}\right) \ge 0, \qquad l = 1, \dots, m-1.$$

If $p = -\infty$, then the $\Phi_{-\infty}$-optimal discrimination design is given by $\eta_{-\infty} = \xi_{-\infty} \times \cdots \times \xi_{-\infty}$; $\xi_{-\infty}$ is a symmetric design on the interval $[-1, 1]$ uniquely determined by its canonical moments $p_1, \dots, p_{2m-1}, p_{2m} = 1$, where $p_{2j-1} = 1/2$, $j = 1, \dots, m$, and $(p_2, \dots, p_{2m-2}) \in (0,1)^{m-1}$ is the unique solution of the system of $m-1$ equations

$$(3.3) \qquad \prod_{j=1}^{l+1} (q_{2j-2}\, p_{2j})^{jN_{q-2,l-j+1}/N_{q,l}} = \frac{\left(\delta^*_{l+1,q}\right)^{1/N_{q-1,l+1}}}{\left(\delta^*_{l,q}\right)^{1/N_{q-1,l}}},$$

$l = 1, \dots, m-1$, *such that* (3.2) *is satisfied.*

Whenever $\beta_{l+1} = 0$, the corresponding equation in (3.1) *and* (3.3) *has to be replaced by the equation*

$$(3.4) \qquad \sum_{j=0}^{m-l-1} (-1)^j \binom{q-1}{j} \prod_{i=1}^{j+l} \frac{q_{2i}}{p_{2i}} \left(1 - \frac{q_{2(l+j+1)}}{p_{2(l+j+1)}}\right) = 0.$$

PROOF. By Theorem 2.1 the Φ_p-optimal product design is of the form $\eta_p = \xi_p \times \cdots \times \xi_p$ where ξ_p is the design on the interval $[-1, 1]$ which maximizes the function F_p^γ in (2.11) with weights $(\gamma_1, \dots, \gamma_m)$ defined by (2.4). Let $p_1, p_2, \dots$ denote the canonical moments of ξ_p. It follows from (2.11) and (2.10) that the odd canonical moments appear only in terms of the form $x(1-x)$ in F_p^γ, which is an increasing function of $\det \Delta_i(\xi)/\det \Delta_{i-1}(\xi)$. Consequently the canonical moments of ξ_p satisfy $p_{2j-1} = 1/2$, $j = 1, \dots, m$; $p_{2m} = 1$. By a result of Lau and Studden (1985) this means that ξ_p is symmetric and supported at $m+1$ points including the boundary points -1 and 1.

Consider at first the case $-\infty < p \le 0$. Dette (1994) proved that the symmetric design ξ_p maximizes the weighted geometric mean (2.12) over the set $\mathscr{P}$ of all probability measures on the interval $[-1, 1]$ if and only if

$$\bar{\gamma}_l = \prod_{j=1}^{l-1} \frac{q_{2j}}{p_{2j}} \left(1 - \frac{q_{2l}}{p_{2l}}\right), \qquad l = 1, \dots, m,$$

where $p_2, \dots, p_{2m-2}, p_{2m} = 1$ denote the canonical moments of even order of ξ_p. An inversion of (2.13) now yields that ξ_p maximizes $F_0^{\tilde{\gamma}}$, where the weights $\tilde{\gamma} = (\tilde{\gamma}_1, \dots, \tilde{\gamma}_m)$ are given by

$$(3.5) \qquad \tilde{\gamma}_l = \frac{N_{q-1,l}}{q} \sum_{j=0}^{m-l} (-1)^j \binom{q-1}{j} \prod_{i=1}^{l+j-1} \frac{q_{2i}}{p_{2i}} \left(1 - \frac{q_{2(l+j)}}{p_{2(l+j)}}\right), \qquad l = 1, \dots, m.$$

On the other hand, Theorem 2.3 and (2.10) prove that the $\tilde{\gamma}_l$ must be proportional to

$$\text{(3.6)} \quad \alpha_l = \gamma_l \prod_{j=1}^{l} \left(\frac{\det \Delta_j(\xi_p)}{\det \Delta_{j-1}(\xi_p)} \right)^{s_l q N_{q-2,l-j}} = \gamma_l \prod_{j=1}^{l} (q_{2j-2}\, p_{2j})^{s_l q N_{q-1,l-j}}.$$

Note that this proportionality shows that the $\tilde{\gamma}_l$ in (3.5) are nonnegative and consequently the canonical moments of the optimal design have to satisfy (3.2). Moreover, because $\beta_{l+1} = 0$ implies $\gamma_{l+1} = 0$ [by definition (2.4)], we obtain (3.4) from $\tilde{\gamma}_{l+1} = \alpha_{l+1} = 0$. If $\beta_{l+1} > 0$, it follows from the above discussion that $\alpha_l/\alpha_{l+1} = \tilde{\gamma}_l/\tilde{\gamma}_{l+1}$, which gives (observing $s_l = p/N_{q-1,l}$)

$$\begin{aligned} \frac{\tilde{\gamma}_l}{\tilde{\gamma}_{l+1}} &= \frac{\gamma_l}{\gamma_{l+1}} \frac{\prod_{j=1}^{l}(q_{2j-2}\, p_{2j})^{s_l q N_{q-1,l-j}}}{\prod_{j=1}^{l+1}(q_{2j-2}\, p_{2j})^{s_{l+1} q N_{q-1,l-j+1}}} \\ &= \frac{\gamma_l}{\gamma_{l+1}} \prod_{j=1}^{l+1} (q_{2j-2}\, p_{2j})^{-jpN_{q-2,l-j+1}/N_{q,l}}. \end{aligned}$$

From (2.4), (2.14) and (3.5) it is now easy to see that this system of equations is precisely (3.1). This means that the canonical moments of the Φ_p-optimal discrimination design ξ_p have to satisfy (3.1) and (3.2). Reversing these arguments shows that any solution of (3.1) subject to (3.2) yields a Φ_p-optimal discrimination design for the class $\mathscr{F}_{m,q}$ with respect to the prior β. Moreover, there exists a unique solution of the maximization problem (2.11) because of the strict concavity of $(F_p^{\gamma})^{1/q}$, and for this reason there exists a unique solution $(p_2, \dots, p_{2m-2}) \in (0,1)^{m-1}$ of the $(m-1)$ equations in (3.1) satisfying (3.2). This proves the case $p > -\infty$. The remaining case $p = -\infty$ is obtained from (3.1) considering the limit $p \to -\infty$. $\square$

Note that Theorem 3.1 extends recent findings for the univariate case $(q = 1)$ of Dette (1994, 1995) to the present multivariate case. The optimal discrimination design is completely determined by the solution of the system of equations in (3.1), which usually has to be solved numerically. The factor ξ_p of the optimal design $\eta_p = \xi_p \times \cdots \times \xi_p$ can be found by standard methods [see, e.g., Lim and Studden (1988), Lemma 4.4, or Dette (1994), Lemma 3.1] and some examples are presented in Section 4. However, if a design of this type is used for model discrimination and an appropriate model has been identified using Anderson's procedure, then the next step in the regression analysis is a statistical inference in the identified model. For this reason it is natural to investigate how our discrimination designs behave for estimating the parameters in the multivariate regression models $h_1(x), \dots, h_m(x)$. Some numerical results in this direction are presented in Section 4, a theoretical result will be given in the remaining part of this section.

Following Dette and Studden (1995) a model robust design for the class $\mathscr{F}_{m,q}$ should yield reasonable D-efficiencies in all multivariate polynomials up to degree m. Thus we call a product design $\eta = \xi_1 \times \cdots \times \xi_q$ Ψ_p-optimal for the class $\mathscr{F}_{m,q}$ with respect to the prior $\beta = (\beta_1, \dots, \beta_m)$ if η maximizes the function

$$\Psi_p(\eta) = \left[\sum_{l=1}^{m} \beta_l \{\mathrm{eff}_l^D(\eta)\}^p \right]^{1/p}, \tag{3.7}$$

which is a weighted p-mean of the D-efficiencies

$$\mathrm{eff}_l^D(\eta) = \left\{ \frac{\det \mathbf{M}_l(\eta)}{\max_{\mu \in \Xi} \det \mathbf{M}_l(\mu)} \right\}^{1/N_{q,l}}, \qquad l = 1, \dots, m. \tag{3.8}$$

The following result shows that Ψ_p-optimal designs can be obtained from the optimal discrimination designs determined in Theorem 3.1. More precisely, the factors of a Φ_p-optimal discrimination design for a regression in $q \geq 2$ variables are the factors of a Ψ_p-optimal design for the class $\mathscr{F}_{m,q-1}$ of models with $(q-1)$ variables with respect to the same prior but for a different value of p.

THEOREM 3.2. *Let* $-\infty \leq p \leq 0$, $q \geq 1$. *Then the factor* ξ_Ψ *of the* Ψ_p*-optimal product design* $\eta_\Psi = \xi_\Psi \times \cdots \times \xi_\Psi$ *for the class* $\mathscr{F}_{m,q}$ *with respect to the prior* $\beta = (\beta_1, \dots, \beta_m)$ *coincides with the factor* ξ_Φ *of the* $\Phi_{p'}$*-optimal discrimination design* $\eta_\Phi = \xi_\Phi \times \cdots \times \xi_\Phi$ *for the class* $\mathscr{F}_{m,q+1}$ (*of models up to degree* m *with* $q+1$ *independent variables*) *with respect to the prior* β, *where* $p' = qp/(q+1)$.

PROOF. By reasoning similar to that in Section 2, it follows that the Ψ_p-optimal product design for the class $\mathscr{F}_{m,q}$ is invariant with respect to permutations of the factors, which means that we can restrict the optimization to permutation-invariant product designs of the form $\eta_\xi = \xi \times \cdots \times \xi$. By Lemma 5.1 in Lim and Studden (1988) and (2.10) we thus obtain, for $l = 1, \dots, m$,

$$\begin{aligned} \det \mathbf{M}_l(\eta_\xi) &= \prod_{i=1}^{l} \left(\int_{-1}^{1} W_i^2(t)\, d\xi(t) \right)^{qN_{q-1,l-i}} \\ &= \prod_{i=1}^{l} \left(\frac{\det \Delta_i(\xi)}{\det \Delta_{i-1}(\xi)} \right)^{qN_{q-1,l-i}} \end{aligned} \tag{3.9}$$

[note that there is a mistake in the formulation of Lim and Studden (1988)]. Lemma 4.3 and Theorem 5.1 in the same reference and (2.14) yield the following for the denominator in (3.8):

$$\begin{aligned}\nu_{l,q} &= \max_{\mu \in \Xi} \det \mathbf{M}_l(\mu) \\ &= \left(\frac{q+l-1}{q+2l-2}\right)^{qN_{q,l-1}} \prod_{k=2}^{l} \left(\frac{(l-k+1)(q+l-k)}{(q+2(l-k+1))(q+2(l-k))}\right)^{qN_{q,l-k}} \\ &= (\delta^*_{l,q+1})^{q/(q+1)}.\end{aligned}$$

Now, observing (3.9), the maximization of the function Ψ_p in (3.7) can be performed by a maximization of the function

$$\begin{aligned}&\left[\sum_{l=1}^{m} \beta_l \left\{ (\delta^*_{l,q+1})^{-q/(q+1)} \prod_{i=1}^{l} \left(\frac{\det \Delta_i(\xi)}{\det \Delta_{i-1}(\xi)}\right)^{qN_{q-1,l-i}} \right\}^{p/N_{q,l}}\right]^{1/p} \\ &\qquad = \left[\sum_{l=1}^{m} \beta_l (\delta^*_{l,q+1})^{-s'_l} \prod_{i=1}^{l} \left(\frac{\det \Delta_i(\xi)}{\det \Delta_{i-1}(\xi)}\right)^{(q+1)s'_l N_{q-1,l-i}}\right]^{1/p},\end{aligned}$$

where

$$s'_l = \frac{pq}{(q+1)N_{q,l}}.$$

By (2.4) this is equivalent to the maximization of the function F_p^{γ} in (2.11), where q and p have to be replaced by $q+1$ and $pq/(q+1)$, respectively. The discussion in Section 2 shows that maximizing $F_{p'}^{\gamma}$ with $p' = qp/(q+1)$ and $\gamma = (\gamma_1, \dots, \gamma_m)$, $\gamma_l = \beta_l (\delta^*_{l,q+1})^{-s'_l}$ in the class of all probability measures on the interval $[-1, 1]$ yields the $\Phi_{p'}$-optimal discrimination design for the class $\mathscr{F}_{m,q+1}$, and the assertion follows. $\square$

Theorem 3.2 indicates that the optimal discrimination designs should also have good efficiencies for estimating the parameters in the identified model. Two cases of this result are worth mentioning here. First, consider the prior $\beta = (0, \dots, 0, 1)$. Then (2.1) and (3.7) reduce to the D_1- and D-optimality criteria and Theorem 3.2 shows that the D-optimal product design in a multivariate polynomial regression with $q \geq 1$ independent variables can be obtained from the D_1-optimal product design for a regression in $q+1$ variables. Second, the same relationship holds for the Φ_p- and Ψ_p-optimality criterion in the case $p = 0$ and $p = -\infty$, independently of the prior β.

4. Numerical results. Tables 1 and 2 show the Φ_p-optimal discrimination designs for the class of polynomial models up to degree $m = 2, 3$ with respect to a uniform prior for various values of the dimension q and the

TABLE 1
Support points and weights of the factor ξ_p of the Φ_p-optimal discrimination design $\eta_p = \xi_p \times \cdots \times \xi_p$ for the class $\mathscr{F}_{2,q}$ with respect to a uniform prior

	$m = 2$								
$p \backslash q$	2			3			4		
0	−1	0	1	−1	0	1	−1	0	1
	0.3889	0.222	0.3889	0.4167	0.1666	0.4167	0.4334	0.1332	0.4334
−1	−1	0	1	−1	0	1	−1	0	1
	0.3948	0.2104	0.3948	0.4201	0.1598	0.4201	0.4356	0.1288	0.4356
−2	−1	0	1	−1	0	1	−1	0	1
	0.3990	0.2020	0.3990	0.4228	0.1544	0.4228	0.4374	0.1252	0.4374
−10	−1	0	1	−1	0	1	−1	0	1
	0.4111	0.1778	0.4111	0.4322	0.1356	0.4322	0.4449	0.1102	0.4449
−∞	−1	0	1	−1	0	1	−1	0	1
	0.4191	0.1618	0.4191	0.4401	0.1198	0.4401	0.4524	0.0952	0.4524

parameter p. The canonical moments of the optimal discrimination designs were obtained by solving the system of equations in Theorem 3.1 numerically. The corresponding designs were calculated by Lemma 4.4 in Lim and Studden (1988). The first line in Tables 1 and 2 contains the support points, the second line the weights of the factor ξ_p of the Φ_p-optimal discrimination design $\eta_p = \xi_p \times \cdots \times \xi_p$.

Theorem 3.2 shows that the Φ_p-optimal discrimination designs give also Ψ_{p^*}-optimal designs for the class $\mathscr{F}_{m,q-1}$ with respect to the same prior, where $p^* = qp/(q-1)$. For example, the Ψ_{-3}-optimal design for the class $\mathscr{F}_{2,2}$ is given by $\xi^* \times \xi^*$, where the design ξ^* has masses 0.4228, 0.1544 and

TABLE 2
Support points and weights of the factor ξ_p of the Φ_p-optimal discrimination design $\eta_p = \xi_p \times \cdots \times \xi_p$ for the class $\mathscr{F}_{3,q}$ with respect to a uniform prior

	$m = 3$					
$p \backslash q$	2		3		4	
0	±1	±0.4010	±1	±0.3648	±1	±0.3366
	0.3195	0.1805	0.3583	0.1417	0.3833	0.1167
−1	±1	±0.4005	±1	±0.3662	±1	±0.3382
	0.3271	0.1729	0.3631	0.1369	0.3865	0.1135
−2	±1	±0.4009	±1	±0.3682	±1	±0.3403
	0.3330	0.1670	0.3671	0.1329	0.3893	0.1107
−10	±1	±0.4092	±1	±0.3888	±1	±0.3622
	0.3528	0.1472	0.3822	0.1178	0.4010	0.0990
−∞	±1	±0.4270	±1	±0.5367	±1	±0.5404
	0.3663	0.1337	0.3761	0.1239	0.4017	0.0975

TABLE 3
D-efficiencies of the Φ_p-optimal discrimination design for the class $\mathscr{F}_{2,q}$ with respect to the uniform prior in the linear and quadratic model

	$m = 2$					
	2		**3**		**4**	
$p \backslash q$	$\mathbf{eff}_1^D$	$\mathbf{eff}_2^D$	$\mathbf{eff}_1^D$	$\mathbf{eff}_2^D$	$\mathbf{eff}_1^D$	$\mathbf{eff}_2^D$
0	0.8457	0.9971	0.8722	0.9943	0.8918	0.9928
-1	0.8542	0.9940	0.8756	0.9916	0.8955	0.9906
-2	0.8603	0.9910	0.8818	0.9890	0.8985	0.9885
-10	0.8777	0.9785	0.8965	0.9765	0.9108	0.9774
$-\infty$	0.8890	0.9667	0.9087	0.9616	0.9231	0.9611

0.4228 at the points -1, 0 and 1, respectively. The D-efficiencies of the optimal discrimination designs in the different models up to degree m are given in Tables 3 and 4. The results show that the discrimination designs have reasonable D-efficiencies for estimating the parameters in all models of the class $\mathscr{F}_{m,q}$. For example, assume that $m = 3$, $q = 4$ and that the experimenter uses the Φ_{-2}-optimal discrimination design η_{-2} (with respect to the uniform prior) in order to identify the appropriate model in the class $\mathscr{F}_{3,4}$. In this case η_{-2} allows the estimation of the parameters in the identified model with at least 84% D-efficiency (which corresponds to the linear model). If Anderson's procedure decides for the quadratic or cubic regression, the efficiencies are 95.5% and 97.9%, respectively. These results indicate that the proposed discrimination designs are very efficient for two important purposes of statistical inference in a multivariate regression: the identification of the model and the estimation of the parameters in the identified model.

TABLE 4
D-efficiencies of the Φ_p-optimal discrimination design for the class $\mathscr{F}_{3,q}$ with respect to the uniform prior in the linear, quadratic and cubic models

	$m = 3$								
	2			**3**			**4**		
$p \backslash q$	$\mathbf{eff}_1^D$	$\mathbf{eff}_2^D$	$\mathbf{eff}_3^D$	$\mathbf{eff}_1^D$	$\mathbf{eff}_2^D$	$\mathbf{eff}_3^D$	$\mathbf{eff}_1^D$	$\mathbf{eff}_2^D$	$\mathbf{eff}_3^D$
0	0.7861	0.9079	0.9953	0.8094	0.9351	0.9899	0.8307	0.9518	0.9862
-1	0.7955	0.9135	0.9909	0.8162	0.9380	0.9855	0.8356	0.9537	0.9826
-2	0.8031	0.9164	0.9864	0.8221	0.9399	0.9811	0.8400	0.9551	0.9789
-10	0.8290	0.9197	0.9636	0.8458	0.9391	0.9567	0.8598	0.9549	0.9568
$-\infty$	0.8484	0.9185	0.9388	0.8645	0.9059	0.9169	0.8656	0.8915	0.8955

REFERENCES

ANDERSON, T. W. (1962). The choice of the degree of a polynomial regression as a multiple decision problem. *Ann. Math. Statist.* **33** 255–265.

DETTE, H. (1994). Discrimination designs for polynomials regression on compact intervals. *Ann. Statist.* **22** 890–903.

DETTE, H. (1995). Optimal designs for identifying the degree of a polynomial regression. *Ann. Statist.* **23** 1248–1267.

DETTE, H. and RÖDER, I. (1995). *Optimal Discrimination Designs for Multifactor Experiments. Dresdner Schriften zur Stochastik* **15**.

DETTE, H. and STUDDEN, W. J. (1995). Optimal designs for polynomial regression when the degree is not known. *Statist. Sinica* **5** 459–473.

DETTE, H. and STUDDEN, W. J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability and Analysis*. Wiley, New York. To appear.

FARREL, H. R., KIEFER, J. and WALBRAN, J. (1967). Optimum multivariate designs. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 113–138. Univ. California Press, Berkeley.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879.

LAU, T. S. and STUDDEN, W. J. (1985). Optimal designs for trigonometric and polynomial regression using canonical moments. *Ann. Statist.* **13** 383–393.

LIM, Y. B. and STUDDEN, J. (1988). Efficient D_s-optimal designs for multivariate polynomial regression on the q-cube. *Ann. Statist.* **16** 1225–1240.

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

RAFAJLOWICZ, E. and MYSZKA, W. (1988). Optimum experimental design for a regression on a hypercube—a generalization of Hoel's result. *Ann. Inst. Statist. Math.* **40** 821–828.

RAFAJLOWICZ, E. and MYSZKA, W. (1992). When product type optimal design is optimal? Brief survey and new results. *Metrika* **39** 321–333.

SILVEY, D. S. (1980). *Optimal Designs*. Chapman and Hall, London.

SKIBINSKY, M. (1986). Principal representations and canonical moment sequences for distributions on an interval. *J. Math. Anal. Appl.* **5** 693–701.

SPRUILL, M. C. (1990). Good designs for testing the degree of a polynomial mean. *Sankhyā Ser. B* **52** 67–74.

STUDDEN, W. J. (1980). D_s-optimal designs for polynomial regression using continued fractions. *Ann. Statist.* **8** 1132–1141.

WONG, W. K. (1994). G-optimal designs for multi-factor experiments. *J. Statist. Plann. Inference* **40** 127–133.

FAKULTÄT UND INSTITUT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
UNIVERSITÄTSSTRASSE 150, GEBÄUDE NA
44780 BOCHUM
GERMANY
E-MAIL: holger.dette@rz.ruhr-uni-bochum.de

INSTITUT FÜR MEDIZINISCHE INFORMATIK
STATISTIK UND EPIDEMIOLOGIE
UNIVERSITÄT LEIPZIG
LIEBIGSTRASSE 27
04103 LEIPZIG
GERMANY
E-MAIL: ingo@imise.uni-leipzig.de