ON A MIXTURE OF THE $D-$ AND D_1-OPTIMALITY CRITERION IN POLYNOMIAL REGRESSION

by

Holger Dette
Purdue University and University of Göttingen

Technical Report #91-23

Department of Statistics
Purdue University

May 1991

⋆

ON A MIXTURE OF THE $D-$ AND D_1-OPTIMALITY CRITERION IN POLYNOMIAL REGRESSION

by

Holger Dette
Purdue University and University of Göttingen

ABSTRACT

We consider a mixture of the D_1–optimality criterion (minimizing the variance of the estimate for the highest coefficient) and the D–optimality criterion (minimizing the volume of the ellipsoid of concentration for the unknown parameter vector) in the polynomial regression model of degree $n \in \mathbb{N}$. The mixture is defined as a weighted product of both optimality criteria and explicit solutions are given for the proposed criterion. The derived designs have excellent efficiencies compared to the $G-$, $D-$ and D_1–optimal design. This is illustrated in some examples for polynomial regression of lower degree. The optimal designs are calculated using the theory of canonical moments. Further applications are given in the field of model robust designs for polynomial regression models where only an upper bound for the degree of the polynomial regression model is known by the experimenter before the experiments are carried out.

Key Words and phrases: D–optimality, D_1–optimality, mixture of optimality criteria, model robust designs, ultraspherical polynomials, canonical moments.

AMS 1980 Subject Classifications: Primary 62K05; Secondary 62J05.

1. Introduction. Consider the polynomial regression model of degree $n \in \mathbb{N}$

$$g_n(x) = \alpha_0 + \alpha_1 x + \ldots + \alpha_n x^n, \qquad x \in [-1,1]. \tag{1.1}$$

For each $x \in [-1,1]$ a random variable $Y(x)$ with mean $g_n(x)$ and variance σ^2 can be observed. A design ξ is a probability measure on [-1,1] and the information matrix of ξ is defined by

$$M_n(\xi) = \int_{-1}^{1} (1, x, \ldots, x^n)^T (1, x, \ldots, x^n) d\xi(x). \tag{1.2}$$

If ξ is supported at k points $x_1, \ldots, x_k$ with masses $\frac{n_1}{N}, \ldots, \frac{n_k}{N}$ where $\sum_{j=1}^{k} n_j = N$, the experimenter takes n_k uncorrelated observations at each x_k. In this case the covariance matrix of the least squares estimator for the unknown parameter vector $(\alpha_0, \ldots, \alpha_n)^T$ is given by $(\sigma^2/N)M_n^{-1}(\xi)$.

An optimal design maximizes (or minimizes) an appropriate functional of the information matrix or its inverse. One of the more commonly used criteria for choosing a design ξ is the D–optimality criterion which maximizes the determinant of the information matrix $M_n(\xi)$. It was shown by Hoel (1958) that the D–optimal design puts equal mass at the zeros of the polynomial $(1-x^2)P_n'(x)$ where $P_n(x)$ denotes the n–th Legendre polynomial. A D–optimal design allows good estimates for the whole parameter vector $\alpha = (\alpha_0, \ldots, \alpha_n)^T$ because (under the additional assumption of normality) it minimizes the volume of the ellipsoid of concentration for the parameter vector $\alpha = (\alpha_0, \ldots, \alpha_n)^T$. By the celebrated equivalence theorem of Kiefer and Wolfowitz (1960) the D–optimal design is also G–optimal, i.e. it minimizes the maximum variance of the least squares estimator for $g_n(x)$ in the regression region $\mathcal{X} = [-1,1]$. However, a D–optimal design should not be used for the decision if the term x^n has to be included into the model or not. For this purpose the D_1–optimality criterion is appropriate which minimizes the variance of the estimator for the coefficient α_n and maximizes the power of the F–test for the hypotheses $\alpha_n = 0$. Kiefer and Wolfowitz (1959) showed that the D_1–optimal design (for polynomial regression of degree n) concentrates mass 1/(2n) on the boundary points $-1, 1$ and mass 1/n at the zeros of $T_n'(x)$ where $T_n(x)$ is the Chebyshev polynomial of the first kind. The disadvantage of the D_1–optimal design is that it does not allow to efficient estimates of the vector α and of the regression function $g_n(x)$ in the model of degree n where it is usually used when the experimenter rejects the hypotheses $\alpha_n = 0$. This fact is illustrated for the case $n = 3$ in Table 1.1 which shows the $D-$ and G–efficiencies

$$D(\xi) = \left(\frac{\det M_n(\xi)}{\sup_\eta \det M_n(\eta)} \right)^{1/n+1}, \quad G(\xi) = \frac{n+1}{\sup_{x\in[-1,1]} d(x,\xi)}$$

$(d_n(x,\xi) = (1,x,\ldots,x^n)M_n^{-1}(\xi)(1,x,\ldots,x^n)^T)$ and the D_1–efficiency

$$D_1(\xi) = \left(\frac{\det M_n(\xi)}{\det M_{n-1}(\xi)} \right) \Big/ \left(\sup_\eta \frac{\det M_n(\eta)}{\det M_{n-1}(\eta)} \right)$$

for the D–optimal and D_1–optimal design ξ_D and ξ_{D_1} in the cubic model $(n = 3)$

	$D(\xi)$	$G(\xi)$	$D_1(\xi)$	support	$\xi(\{\pm 1\})$	$\xi(\{\pm t\})$
ξ_D	1	1	0.8566	$-1, \pm 1/\sqrt{5}, 1$	1/4	1/4
ξ_{D_1}	0.9344	0.6667	1	$-1, \pm 1/2, 1$	1/6	1/3

Table 1.1: $D-$, $G-$ and D_1 efficiencies of the $D-$ and D_1–optimal designs ξ_D and ξ_{D_1} in the cubic model.

It can also be shown that the D_1–optimal design is not very efficient in the model of degree $n-1$ where it is usually used when the experimenter has decided (for example by a F–test) to omit the power x^n in the regression model (efficiency in the quadratic model 0.75).

In this paper we will determine designs which are very efficient with respect to all three optimality criteria. To this end we will introduce an optimality criterion which is a mixture of the $D-$ and D_1–optimality criterion. This criterion is defined in section 2 in a general context and an equivalence theorem is stated. In section 3 we go back to polynomial regression models and determine the support and the weights of the optimal design with respect to our criterion. For this purpose we use the theory of canonical moments which was introduced by Studden (1980, 1982a, 1982b) (see also Lau (1983), Lau and Studden (1985) and Lim and Studden (1988) for more details). In section 4 we give some examples for the polynomial regression models of lower degree and compare the efficiencies of the derived designs with the efficiencies of the $D-$ $(G-)$ and D_1–optimal designs. In section 5 we will apply the results of section 3 in the context of model robust design for polynomial regression which was considered by Läuter (1974, 1976), Cook and Nachtsheim (1982) and Dette (1990, 1991). Finally in the appendix a proof of the main theorem (stated in section 3) is given.

2. The mixture of the $D-$ and D_1–optimality criterion. Let $f_0(x), \dots, f_n(x)$ denote $n+1$ linear independent regression functions defined on a compact experimental domain $\mathcal{X}$ which contains at least $n+1$ points and define the model g by

$$g(x) = \alpha_0 f_0(x) + \ldots + \alpha_n f_n(x).$$

The information matrix for a design ξ on $\mathcal{X}$ is now given by

$$M_n(\xi) = \int_{\mathcal{X}} \tilde{f}_n(x) \tilde{f}_n^T(x) d\xi(x)$$

where $\tilde{f}_n(x) = (f_0(x), \dots, f_n(x))^T$. Let $c_n = (0, \dots, 0, 1)^T$, the variance of the least squares estimator for α_n is proportional to $c_n^T M_n^{-1}(\xi) c_n$ while the volume of the ellipsoid of concentration for $\alpha = (\alpha_0, \dots, \alpha_n)$ is proportional to $\det M_n^{-1}(\xi)$. We call a design ξ ψ_β–optimal if it maximizes the function

$$(2.1) \qquad \begin{cases} \psi_\beta(\xi) = (1-\beta) \log(c_n^T M_n^{-1}(\xi) c_n)^{-1} + \dfrac{\beta}{n+1} \log\det\, M_n(\xi) \\ \qquad\;\; = (1-\beta) \cdot \log \dfrac{\det\, M_n(\xi)}{\det\, M_{n-1}(\xi)} + \dfrac{\beta}{n+1} \log\det\, M_n(\xi) \end{cases}$$

where $\beta \in [0,1]$ reflects the desired weight of the experimenter for the D–optimality criterion. Note that a ψ_β–optimal design will always have a non singular information matrix and that ψ_β gives the $D-$ and D_1–optimality criterion for $\beta = 1$ and $\beta = 0$, respectively. The following theorem characterizes the ψ_β–optimal designs.

Theorem 2.1. Let $c_n = (0, \ldots, 0, 1)^T \in \mathbb{R}^n$, then the following three conditions are equivalent

(i) The design ξ_β is ψ_β–optimal

$$\text{(ii)} \qquad (1-\beta)\,\frac{(\tilde{f}_n^T(x)M_n^{-1}(\xi)c_n)^2}{c_n^T M_n^{-1}(\xi)c_n} + \frac{\beta}{n+1}\,\tilde{f}_n^T(x)M_n^{-1}(\xi)\tilde{f}_n(x) \;\leq\; 1 \quad \forall x \in \mathcal{X}$$

$$\text{(iii)} \qquad \frac{\beta}{n+1}\sum_{l=0}^{n-1}\frac{(\tilde{f}_l^T(x)M_l^{-1}(\xi)c_l)^2}{c_l^T M_l^{-1}(\xi)c_l} + \frac{n(1-\beta)+1}{n+1}\,\frac{(\tilde{f}_n^T(x)M_n^{-1}(\xi)c_n)^2}{c_n^T M_n^{-1}(\xi)c_n} \leq 1 \quad \forall x \in \mathcal{X}$$

Moreover, in (ii) and (iii) we have equality for the support points of the optimal design ξ_β and the expressions on the left of (ii) and (iii) are the same.

Proof: We consider the Frèchet derivative of ψ_β at M_1 in direction of M_2 (see Silvey (1980), p. 18) and obtain by straightforward calculations

$$F_{\psi_\beta}(M_1, M_2) = (1-\beta)\left\{\frac{c_n^T M_1^{-1} M_2 M_1^{-1} c_n}{c_n^T M_1^{-1} c_n} - 1\right\} + \frac{\beta}{n+1}\,\{\mathrm{tr}(M_2 M_1^{-1}) - (n+1)\}$$

(see also Silvey (1980), p. 21, p. 48). The assertion (ii) now follows by an application of Theorem 3.7 of Silvey (1980), p. 19, which gives a general equivalence for concave and differentiable optimality criteria. From

$$\log(\det\, M_n(\xi)) = \sum_{l=0}^{n}\log\frac{\det\, M_l(\xi)}{\det\, M_{l-1}(\xi)} = \sum_{e=0}^{n}\log(c_l^T M_l^{-1}(\xi)c_l)^{-1}$$

($\det\, M_{-1}(\xi) = 1$) we have for ψ_β

$$\psi_\beta(\xi) = \frac{\beta}{n+1}\sum_{l=0}^{n-1}\log(c_l^T M_l^{-1}(\xi)c_l)^{-1} + \frac{n(1-\beta)+1}{n+1}\log(c_n^T M_n^{-1}(\xi)c_n)^{-1}$$

and the condition (iii) is proved by a similar reasoning as given for (ii).

We should mention at this point, that Theorem 2.1 is also a consequence of a general equivalence theorem for mixtures of optimality criteria, which was considered by Gutmair (1991) and does not require the assumption of differentiability.

Theorem 2.1 allows to check if a given design is ψ_β–optimal and it is also the basis of any numerical procedure for the determination of ψ_β–optimal designs (see Fedorov (1972), Silvey (1980)). However, it is not a very useful tool for the determination of an analytical solution of the ψ_β–optimal design problem. In the case of polynomial regression there are more efficient methods for this purpose which are considered in the following section.

3. Polynomial regression and the theory of canonical moments. In this section we are going back to the polynomial regression model of section 1 and give a brief review of the theory of canonical moments which is used to determine the ψ_β–optimal designs. For a more detailed investigation we refer to the work of Skibinsky (1967, 1968, 1969, 1986), Studden (1980, 1982a, 1982b, 1989), Lau (1983, 1988), Lau and Studden (1985, 1988), Lim and Studden (1988) and Dette (1990, 1991). Let ξ denote a probability measure on [-1,1] with moments $c_i = \int_{-1}^{1} x^i d\xi(x)$. For a given set of moments $c_0, \ldots, c_{i-1}$ let c_i^+ denote the maximum of the i–th moment $\int_{-1}^{1} x^i d\mu(x)$ over the set of all probability measures μ having the given moments $c_0, \ldots, c_{i-1}$. Similarly let c_i^- denote the corresponding minimum value. The canonical moments are defined by

$$p_i = \frac{c_i - c_i^-}{c_i^+ - c_i^-} \quad , \quad i = 1, 2, \ldots$$

Note that $0 \leq p_i \leq 1$ and that the canonical moments are left undefined whenever $c_i^+ = c_i^-$. If i is the first index for which this equality holds, then $0 < p_k < 1$, $k = 1, 2, \ldots, i-2$, p_{i-1} must have the value 0 or 1 and ξ has a finite support (see Skibinsky (1986), section 1). As an example consider the Jacobi measure with density proportional to $(1-x)^\alpha (1+x)^\beta (\alpha > -1, \beta > -1)$. Skibinsky (1969) showed that the canonical moments of the Jacobi measure are given by

$$p_{2k} = \frac{k}{\alpha + \beta + 2k + 1} \qquad k \geq 1$$

$$p_{2k-1} = \frac{\beta + k}{\alpha + \beta + 2k} \qquad k \geq 1.$$

The uniform measure ($\alpha = \beta = 0$) has $p_{2k-1} = \frac{1}{2}$ and $p_{2k} = k/(2k+1)$ $(k \geq 1)$ and the arcsin–distribution ($\alpha = \beta = -\frac{1}{2}$) has $p_k = \frac{1}{2}$ for all k. The determinants of the information matrices (for the polynomial regression model) given in (1.2) can easily be expressed in terms of canonical moments (see Skibinsky (1968), Studden (1982b)).

Theorem 3.1. Let ξ denote a probability measure with canonical moments $p_1, p_2, \dots, q_j = 1 - p_j \ \ (j \geq 1), \zeta_0 = 1, \ \zeta_1 = p_1$ and $\zeta_j = q_{j-1} p_j \ \ (j \geq 2)$, then the information matrix of ξ (in the polynomial regression model of degree n) is given by

$$\det M_n(\xi) = 2^{n(n+1)} \prod_{i=1}^{n} (\zeta_{2i-1} \zeta_{2i})^{n+1-i}.$$

The maximization of the optimality criterion given in (2.1) can now be carried out by a maximization in terms of the canonical moments of the design ξ and we obtain.

Theorem 3.2. The ψ_β–optimal design ξ_β for the polynomial regression model of degree n is uniquely determined by its canonical moments which are given by

$$\text{(3.1)} \qquad \begin{cases} p_{2i-1} = \dfrac{1}{2} & i = 1, \dots, n \\ p_{2i} = \dfrac{n+1-i\beta}{2(n+1-i\beta) - \beta} & i = 1, \dots, n-1 \\ p_{2n} = 1 & \end{cases}$$

Proof: Maximizing the function $\psi_\beta(\xi)$ in terms of the canonical moments of ξ we obtain by straightforward algebra the moments of (3.1). The fact that ξ_β is the unique design with canonical moments (3.1) results from $p_{2n} = 1$ (see Skibinsky (1986)).

Note that the cases $\beta = 1$ and $\beta = 0$ give the canonical moments of the $D-$ and D_1-optimal design. If we let $\beta = \frac{n+1}{n+1+a}$ $(a \in [0, \infty])$ we obtain for the sequence $(p_1, \dots, p_{2n})$ of (3.1)

$$\text{(3.2)} \qquad \begin{cases} p_{2i-1} = \dfrac{1}{2} & i = 1, \dots, n \\ p_{2i} = \dfrac{n-i+a+1}{2(n-i+a)+1} & i = 1, \dots, n-1 \\ p_{2n} = 1 & \end{cases}$$

where $a = 0$ and $a = \infty$ correspond to the $D-$ and D_1–optimal design. In this notation we have a nice interpretation of the ψ_β–optimal sequences of canonical moments. The sequence of canonical moments of the D–optimal design $(p_1, \ldots, p_{2n})$ is given by (see Studden (1980))

$$\frac{1}{2}, \frac{n}{2n-1}, \frac{1}{2}, \frac{n-1}{2n-3}, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{4}{7}, \frac{1}{2}, \frac{3}{5}, \frac{1}{2}, \frac{2}{3}, \frac{1}{2}, 1$$

while the sequence of the D_1–optimal design is

$$\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{1}{2}, 1.$$

For $a = 1$ ($\beta = \frac{n+1}{n+2}$) we obtain for the canonical moments of ξ_β the "shifted" sequence

$$\frac{1}{2}, \frac{n+1}{2n+1}, \frac{1}{2}, \frac{n}{2n-1}, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{5}{9}, \frac{1}{2}, \frac{4}{7}, \frac{1}{2}, \frac{3}{5}, \frac{1}{2}, 1$$

while $a = 2$ (3.2) yields the sequence

$$\frac{1}{2}, \frac{n+2}{2n+3}, \frac{1}{2}, \frac{n+1}{2n+1}, \frac{1}{2}, \ldots, \frac{1}{2}, \frac{6}{11}, \frac{1}{2}, \frac{5}{9}, \frac{1}{2}, \frac{4}{7}, \frac{1}{2}, 1.$$

Thus by letting $a = 0, 1, 2, \ldots \to \infty$ we are shifting the sequence of canonical moments of the D–optimal design (to the right) into the sequence of the D_1–optimal design (note that if $a \to \infty$ all the canonical moments of (3.2) will tend to 1/2).

Theorem 3.2 gives, in a sense, a complete solution of the ψ_β–optimal design problem. The remainder of the problem is converting the optimum p_i to the support points and the weights of the corresponding design. There is a considerable amount of literature concerning the relationship between a sequence of canonical moments and the corresponding design ξ (see Studden (1982a,b) and Lau (1983)). In general the support and the weights of the design ξ have to be found numerically and analytic solutions are only possible for $n \leq 3$ or special sequences of canonical moments (see Dette (1990, 1991)). However, the sequences defined in (3.2) (or (3.1)) allow an explicit representation of the support points as the zeros of a weighted sum of orthogonal polynomials. In what follows let $C_n^{(\alpha)}(x)$ denote the n–th ultraspherical polynomial which is orthogonal with respect to the measure $(1 - x^2)^{\alpha - 1/2} dx$ (see Szegö (1959) or Abramowitz and Stegun (1964)), then we have the following theorem which proof is deferred to the appendix.

Theorem 3.3. Let $n \geq 2$ and $a+z > -1$, the design ξ corresponding to the sequence of canonical moments

$$(3.3) \qquad \begin{cases} p_{2i-1} = \dfrac{1}{2} \\ p_{2i} = \dfrac{n-i+a+z}{2(n-i+a)+z} & i = 1,\dots,n-1 \\ p_{2n} = 1 \end{cases}$$

is supported at the points $-1, 1$ and the $n-1$ zeros of the polynomial

$$(3.4) \quad F_{n-1}^{(a,z)}(x) = \sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^j \frac{\Gamma(a+j)}{\Gamma(a)\Gamma(j+1)} \; \frac{\Gamma(z+a+1+j)}{\Gamma(z+a+1)} \; \frac{\Gamma(n-j)}{\Gamma(n)} \; C_{n-2j-1}^{(\frac{z+2}{2}+a+j)}(x).$$

The masses at the support points $x_0, x_1, \dots, x_n$ are given by

$$(3.5) \qquad \xi(\{x_j\}) = \frac{x_j \cdot F_{n-1}^{(a+1,z-2)}(x_j) - \frac{a+1}{n-1} F_{n-2}^{(a+2,z-2)}(x_j)}{\frac{d}{dx}(x^2-1) F_{n-1}^{(a,z)}(x)|_{x=x_j}} \qquad j = 0, 1, \dots, n.$$

Note, that for $a = 0$, Theorem 3.3 gives the model robust designs considered in Dette (1990) while in the case $a = \infty$ we obtain the D_1–optimal design for all $z > -1$. By an application of Theorem 3.2 we immediately get the following result.

Corollary 3.4. Let $\beta \in [0,1]$ and $a = (n+1)\frac{1-\beta}{\beta}$. The ψ_β–optimal design (for the polynomial regression model of degree n is supported at the points $-1, 1$ and the zeros of the polynomial given in (3.4) for $z = 1$. The weights are given by (3.5) (for $z = 1$).

4. Examples (polynomial regression of degree 3 and 4).

a) We will consider the case of cubic regression given in the introduction. Let $a = 4\frac{1-\beta}{\beta}$, then we see from Corollary 3.4 that the support of the ψ_β–optimal design is given by -1 and 1 and the zeros of the polynomial

$$F_2^{(a,1)}(x) = C_2^{(a+3/2)}(x) - \tfrac{a(a+2)}{2} = (a+\tfrac{3}{2})(2a+5)x^2 - \tfrac{(a+3)(a+1)}{2}$$

where we have used the representation $C_2^{(\alpha)}(x) = \alpha[2(\alpha+1)x^2-1]$ for the ultraspherical polynomial of degree 2 (see Abramowitz and Stegun (1964), p. 794). By the same theorem we obtain for the weights at the support points

$$(4.1)\quad \begin{aligned} &x_1 = -1, \quad x_2 = -\sqrt{\frac{(a+1)(a+3)}{(2a+3)(2a+5)}}, \quad x_3 = \sqrt{\frac{(a+1)(a+3)}{(2a+3)(2a+5)}}, \quad x_4 = 1 \\ &\xi_\beta(\{x_j\}) = \frac{1}{2}\,\frac{(2a+3)(2a+5)x_j^2 - (3a^2+11a+9)}{2(2a+3)(2a+5)x_j^2 - (5a^2+20a+18)} \qquad j = 1,\ldots,4 \end{aligned}$$

which yields (note that $\beta = \frac{n+1}{n+1+a}$)

$$(4.2)\qquad \xi_\beta(\{\pm 1\}) = \frac{1}{6}\,\frac{a+3}{a+2}, \qquad \xi_\beta(\{x_{2/3}\}) = \frac{1}{6}\,\frac{2a+3}{a+2}.$$

The optimal designs for various values of a (β) are given in Table 4.1 which also contains the $D-$, $G-$ and D_1 efficiencies for the different designs in the cubic model. The $D-$ and D_1–efficiencies are calculated by an application of Theorem 3.1 and 3.2. For the G–efficiency we use

$$\sup_{x\in[-1,1]} d_2(x,\xi) = d_2(1,\xi) = 1 + \frac{1}{p_2} + \frac{q_2}{p_2p_4} + \frac{q_2q_4}{p_2p_4p_6}$$

which can be proved by a similar reasoning as given in Lim and Studden (1988, p. 1237) (see also Dette (1990), p. 1790).

a	β	$D(\xi)$	$G(\xi)$	$D_1(\xi)$	$x_{2/3}$	$\xi_\beta(x_{2/3})$	$\xi_\beta(\pm 1)$
0	1	1	1	0.8566	±0.4472	0.25	0.25
1	0.8	0.9908	0.8889	0.9074	±0.4781	0.2222	0.2778
2	0.6667	0.9809	0.8333	0.9674	±0.4880	0.2083	0.2917
3	0.5714	0.9735	0.8	0.9795	±0.4923	0.2	0.3
4	0.5	0.9681	0.7778	0.9858	±0.4947	0.1944	0.3056
6	0.4	0.9607	0.75	0.9921	±0.4971	0.1875	0.3125
10	0.2857	0.9527	0.7222	0.9965	±0.4986	0.1806	0.3194
∞	0	0.9344	0.6667	1	±0.5	0.1667	0.3333

Table 4.1: ψ_β–optimal designs and the $D-$, $G-$ and D_1- efficiencies for different values of $\beta = 4/(4+a)$

Note that Table 4.1 indicates the convergence of the ψ_β optimal design to the D_1–optimal design as β tends to zero ($a \to \infty$). In the range $\beta \in [0.5, 0.6667]$ ($a \in [2,4]$)

all designs ξ_β have good $D-$, $G-$ and D_1–efficiences which means that these designs are in some sense robust with respect to the ($D-$, $G-$ and D_1-) optimality criterion. Because of the simple form of its weights we would recommend the design ξ_β ($\beta = 0.5714$, $a = 3$) for practical applications in the cubic regression model.

b) As a further example we consider the case $n = 4$, $\beta = 1/2$ which corresponds to the case $a = 5$ in Corollary 3.4. The canonical moments of even order of the ψ_β–optimal design are given by $p_2 = 9/17$, $p_4 = 8/15$, $p_6 = 7/13$, $p_8 = 1$ and from the definition of the ultraspherical polynomials we obtain

$$C_3^{(13/2)}(x) = \frac{65}{2}x[17x^2 - 3], \qquad C_1^{(15/2)}(x) = 15x$$

and it follows by straightforward algebra

$$F^{(5,1)}(x) = \tfrac{5}{2}x[221x^2 - 109].$$

Thus the support of the optimal design is given by the points -1, $-\sqrt{\frac{109}{221}}$, 0, $\sqrt{\frac{109}{221}}$, 1. For the weights we apply (3.5) and obtain

$$\xi_\beta(\{x_j\}) = \frac{221\ x_j^4 - 213\ x_j^2 + \frac{128}{5}}{5(221\ x_j^4 - 198\ x_j^2 + \frac{109}{5})} \qquad (j = 0, 1, \ldots, 4)$$

which yields

$$\xi_\beta(\{\pm 1\}) = \frac{3}{20}, \qquad \xi_\beta\left(\pm\sqrt{\frac{109}{221}}\right) = \frac{3549}{15260} \approx 0.2326, \qquad \xi_\beta(\{0\}) = \frac{128}{545}.$$

The D–efficiency of this design is 0.9662, the G–efficiency is 0.75 while its D_1–efficiency is given by 0.9862. Note also that the D–optimal design has D_1 efficiency 0.8359 and that the D_1–optimal design has D–efficiency 0.9338 and G–efficiency 0.625.

5. Model robust designs. In this section we will consider some aspects of optimal design theory where the underlying model is not known by the experimenter before the experiments are carried out (see Atkinson (1972), Stigler (1971), Läuter (1974, 1976), Studden (1982a), Dette (1990, 1991) for more details). We will use the optimality criterion introduced by Läuter (1974) for a given class of models. To this end let $g_l(x)$ denote a polynomial regression model of degree l $(l = 1, \ldots, n)$ and collect all the polynomial regression models up to the degree n in the set

$$\mathcal{F}_n = \{g_1(x), \ldots, g_n(x)\}.$$

Let $\beta_l \geq 0$ denote a prior, which reflects the experimenters belief about the adequacy of the model g_l $\left(\sum_{l=1}^{n} \beta_l = 1\right)$, a design ξ is called optimal for the class of $\mathcal{F}_n$ with respect to the prior $\beta = (\beta_1, \ldots, \beta_n)$ if it maximizes the function

$$\Phi_\beta(\xi) = \sum_{l=1}^{n} \frac{\beta_l}{l+1} \log(\det M_l(\xi))$$

where $M_l(\xi)$ denotes the information matrix of the design ξ in the model g_l defined by (1.2). The canonical moments of the ξ_β–optimal design with respect to a given prior β were determined by Dette (1990) who also identified a one parametric class of priors yielding optimal designs of a very simple structure. We will consider here the following two parametric classes of priors $(a + z > -1)$

$$(5.1) \qquad \begin{cases} \beta_l = (l+1) \dfrac{z(z-1)}{\Sigma} \dfrac{\Gamma(n+a+z-l-1)}{\Gamma(n+a-l+1)} & l = 1, \ldots, n-2 \\[2ex] \beta_{n-1} = n \dfrac{\Gamma(z+a+1)}{\Gamma(a+2)} \dfrac{(z-a-1)}{\Sigma} & \\[2ex] \beta_n = (n+1) \dfrac{\Gamma(z+a+1)}{\Gamma(a+1)} \cdot \dfrac{1}{\Sigma} & \end{cases}$$

where Σ is a normalizing constant to ensure $\sum_{l=1}^{n} \beta_l = 1$. Note that this class contains the one parametric class investigated by Dette (1990) $(a = 0)$. For the case $a = 0,\ z = 2$ the prior defined in (5.1) puts weight proportional to $2{:}\,3{:} \ldots n{:}\,n+1$ on every model of $\mathcal{F}_n$. A very important case is obtained for the choice $z = 1,\ a \in [-1, 0]$, which reduces the prior of (5.1) to a prior which puts only non vanishing (positive) weight on the models of degree n and $n-1$, i.e.

$$(5.2) \qquad \begin{cases} \beta_1 = \ldots = \beta_{n-2} = 0 \\[1ex] \beta_{n-1} = \dfrac{-an}{a+n+1} \\[2ex] \beta_n = \dfrac{(a+1)(n+1)}{a+n+1}. \end{cases}$$

Note that when a varies between -1 and 0 β_{n-1} and $\beta_n = 1 - \beta_{n-1}$ vary in the same range. This prior is reasonable for an experimenter who wants to fit either a polynomial of degree n or of degree $n-1$. The optimal design for the class $\mathcal{F}_n$ with respect to a prior $\beta = (\beta_1, \ldots, \beta_n)$ can be easily expressed in terms of canonical moments (see Dette (1990), Theorem 3.1 and formula (4.1)) and by an application of Theorem 3.1 and Theorem 3.3 we obtain the following result.

Theorem 5.1. Let $a + z > -1$, then optimal design for the class $\mathcal{F}_n$ with respect to a prior defined by (5.1) has the canonical moments given in (3.3). The support points are given by -1,1 and the zeros of the polynomial $\mathcal{F}_{n-1}^{(a,z)}(x)$ defined by (3.4) while the weights can be calculated using the representation (3.5).

Note that in the case $z = 1$ the ψ_β–optimal design (Corollary 3.4) and the optimal design for the polynomial of degree n or $n-1$ (i.e. the optimal design for the class $\mathcal{F}_n$ with respect to the prior given in (5.2)) are completely different although they are given by the same sequence of canonical moments in (3.3) ($z = 1,\ a > -2$). This results from the fact that in the optimality criteria ψ_β and Φ_β reasonable priors should be positive which yields $a \in [-1, 0]$ for the Φ_β–criterion (prior in (5.2)) and $a > 0$ for the ψ_β–criterion. Therefore an ψ_β–optimal design for polynomial regression of degree n can not be used as an optimal design for the class $\mathcal{F}_n$ with respect to the prior in (5.2).

Example: (Cubic or quadratic regression). Assume that the experimenter wants to fit a quadratic or cubic model and that he has no prior knowledge about the adequacy of these models. In this case he would choose $\beta_1 = 0,\ \beta_2 = \beta_3 = \frac{1}{2}$ which corresponds to $a = -\frac{4}{7}$ in (5.2). This is the case $a = -\frac{4}{7},\ z = 1$ and $n = 3$ of Theorem 5.1 which is also treated in section 4. Thus we obtain from (4.1) and (4.2) that the optimal design for the quadratic and cubic polynomial (equally weighted) is supported at the points $-1 - \sqrt{\frac{17}{117}}, +\sqrt{\frac{17}{117}}, 1$ with masses $\frac{17}{60}, \frac{13}{60}, \frac{13}{60}, \frac{17}{60}$. The D–efficiency of this design in the cubic model is given by 0.9775 while its D–efficiency in the quadratic model is 0.9135. If we want to increase the efficiency of the optimal design in the quadratic polynomial we have to put more weight on the polynomial of degree 2. To give an example we choose $\beta_2 = 3/4$ and $\beta_3 = 1/4$ which corresponds to the case $a = -4/5$. The optimal design is supported at the points $-1,\ -\sqrt{\frac{11}{119}},\ \sqrt{\frac{11}{119}},\ 1$ and the weights are $\frac{11}{36},\ \frac{7}{36},\ \frac{7}{36},\ \frac{11}{36}$. This design has D–efficiencies 0.9120 and 0.9491 in the cubic and quadratic model, respectively.

6. Appendix

Proof of Theorem 3.3: a) The representation of the support: Let ξ denote the design corresponding to the sequence in (3.3), then the Stieltjes transform of ξ has the representation (see Perron (1954) or Dette (1989))

$$\begin{aligned}\Phi(x) &= \int_{-1}^{1} \frac{d\xi(t)}{x-t} \\ &= \frac{1}{1-x^2}\left[-x-1+2\gamma_1+\frac{4\gamma_1\gamma_2 \quad |}{|x+1-2(\gamma_2+\gamma_3)} - \ldots - \frac{4\gamma_{2n-3}\gamma_{2n-2} \quad |}{|x+1-2(\gamma_{2n-2}+\gamma_{2n-1})}\right]\end{aligned}$$

where $\gamma_1 = 1-p_1,\ \gamma_j = p_{j-1}(1-p_j)\ \ (j \geq 2)$. The support of ξ is now given by -1, 1 and the polynomial in the denominator of the above continued fraction. Defining

$$K\begin{pmatrix} & a_1 & & \ldots & a_n & \\ b_0 & & b_1 & \ldots & & b_n \end{pmatrix} = \det\begin{bmatrix} b_0 & -1 & & & \\ a_1 & b_1 & -1 & & \\ & & \ddots & & \\ & & & & -1 \\ & & & a_n & b_n \end{bmatrix}$$

(all other elements in the matrix are 0) we obtain for this polynomial (see Perron (1954), p. 4 or Wall (1948)) ($P_0(x) = 1,\ P_1(x) = x$)

$$P_{n-1}(x) = K\begin{pmatrix} & -\frac{(n-2+a)(n-1+a+z)}{(2n-4+2a+z)(2n-2+2a+z)} & \ldots & -\frac{(a+1)(a+2+z)}{(2a+2+z)(2a+4+z)} & \\ x & & x\ \ldots & & x \end{pmatrix}. \tag{6.1}$$

We will now show that this polynomial (of degree $n-1$) is proportional to the polynomial $F_{n-1}^{(a,z)}(x)$ defined in Theorem 3.3. More precisely we will show by induction that

$$\tilde{P}_{n-1}(x) = 2^{n-1}\frac{\Gamma(\frac{z}{2}+a+n)}{\Gamma(\frac{z+2}{2}+a)\Gamma(n)}P_{n-1}(x) = F_{n-1}^{(a,z)}(x) \tag{6.2}$$

where $F_{n-1}^{(a,z)}(x)$ is defined in (3.4). For $n=1$ the identity (6.2) is obvious while the case $n=2$ yields

$$\tilde{P}_1(x) = 2\left(\frac{z+2}{2}+a\right)x = C_1^{(\frac{z+2}{2}+a)}(x)$$

which is evident from the definition of the ultraspherical polynomials (see Abramowitz and Stegun (1964), p. 794). For the induction step from n to $n+1$ it is convenient to distinguish the cases of odd or even n and we will only consider the case of $n = 2m$ even (the case $n = 2m+1$ is treated in exactly the same way). From the induction hypotheses we thus obtain for $\tilde{P}_{2m}(x)$ (by an expansion of the determinant)

$$\tilde{P}_{2m}(x) = \frac{2^{2m}\Gamma(\frac{z+2}{2}+a+2m)}{\Gamma(\frac{z+2}{2}+a)\Gamma(2m+1)}\left[xP_{2m-1}(x) - \frac{(2m-1+a)(2m+a+z)}{(4m-2+2a+z)(4m+2a+z)}\, P_{2m-2}(x)\right]$$

$$= \sum_{j=0}^{m-1}(-1)^j \frac{\Gamma(a+j)}{\Gamma(a)\Gamma(j+1)}\ \frac{\Gamma(z+a+j+1)}{\Gamma(z+a+1)}\ \frac{\Gamma(2m-j-1)}{\Gamma(2m+1)}$$

$$\left\{(2m-j-1)(z+2a+4m)xC_{2m-1-2j}^{(\frac{z+2}{2}+a+j)}(x) - (2m-1+a)(2m+a+z)\ C_{2m-2-2j}^{(\frac{z+2}{2}+a+j)}(x)\right\}.$$

Now let $\tilde{P}_{2m}^{(f)}(x)$ denote the above sum where the summation is only performed over the indices $0,1,\dots,f$ ($f \in \{0,1,\dots,m-1\}$, i.e. $\tilde{P}_{2m}^{(m-1)}(x) = \tilde{P}_{2m}(x)$) and similarly define $F_{2m}^{(f)}(x)$ as the "truncated" sum of the polynomial $F_{2m}^{(z,a)}(x)$ defined in (3.4), i.e.

$$F_{2m}^{(f)}(x) = \sum_{j=0}^{f}(-1)^j \frac{\Gamma(a+j)}{\Gamma(a)\Gamma(j+1)}\ \frac{\Gamma(z+a+j+1)}{\Gamma(z+a+1)}\ \frac{\Gamma(2m+1-j)}{\Gamma(2m+1)}\ C_{2m-2j}^{(\frac{z+2}{2}+a+j)}(x)$$

($f \in \{0,\dots,m\}$) then we have the following Lemma which proof is given at the end of this appendix.

Lemma 6.1. The polynomials $F_{2m}^{(f)}(x)$ and $\tilde{P}_{2m}^{(f)}(x)$ satisfy the equation ($f = 0,\dots,m-1$)

$$\tilde{P}_{2m}^{(f)}(x) - \tilde{F}_{2m}^{(f)}(x) = (-1)^{f+1}\frac{\Gamma(a+f+1)}{\Gamma(a)\Gamma(f+1)}\ \frac{\Gamma(z+a+2+f)}{\Gamma(z+a+1)}\ \frac{\Gamma(2m-1-f)}{\Gamma(2m+1)}\cdot C_{2m-2f-2}^{(\frac{z+2}{2}+f+a)}(x).$$

Using Lemma 6.1 we obtain now for the differences of the polynomials $\tilde{P}_{2m}(x)$ and $F_{2m}^{(a,z)}(x)$

$$\begin{aligned}\tilde{P}_{2m}(x) - F_{2m}^{(a,z)}(x) &= \tilde{P}_{2m}^{(m-1)}(x) - F_{2m}^{(m-1)}(x)\\ &\quad - (-1)^m\ \frac{\Gamma(a+m)}{\Gamma(a)\Gamma(m+1)}\ \frac{\Gamma(z+a+m+1)}{\Gamma(z+a+1)}\ \frac{\Gamma(m+1)}{\Gamma(2m+1)} = 0\end{aligned}$$

which establishes (6.2) (in the case of even $n = 2m$). Because the odd case is shown in the same way we have proved the assertion of Theorem 3.3 concerning the support points of the ψ_β–optimal design.

b) The representation of the weights: The weights of $\xi_\beta(\{x_j\})$ at the support points $-1 = x_0 < x_1 < \ldots < x_n = 1$ can be calculated by a partial fraction expansion of another representation of the Stieltjes transform of ξ_β

$$\Phi(x) = \sum_{j=0}^{n} \frac{\xi_\beta(\{x_j\})}{x - x_j} = \int_{-1}^{1} \frac{d\xi_\beta(t)}{x-t}$$
$$= \frac{1 \quad |}{|x+1-2\zeta_1} - \frac{4\zeta_1\zeta_2 \quad |}{|x+1-2\zeta_2-2\zeta_3} - \ldots - \frac{4\zeta_{2n-1}\zeta_{2n} \quad |}{|x+1-2\zeta_{2n}}$$

(see Lau and Studden (1988)). This yields for $j = 0, \ldots, n$

$$\xi_\beta(\{x_j\}) = \Phi(x)(x-x_j)|_{x=x_j} = \frac{G_n(x_j)}{\frac{d}{dx} H_{n+1}(x)|_{x=x_j}}$$

where $H_{n+1}(x) = \prod_{i=0}^{n}(x - x_i)$ and the polynomial in the numerator is given by (see Perron (1954), p. 4)

$$G_n(x) = K \begin{pmatrix} 1 & & -4\zeta_1\zeta_2 & & \cdots & -4\zeta_{2n-1}\zeta_{2n} & \\ 0 & x+1-2\zeta_1 & & x+1-2\zeta_2-2\zeta_3 & \cdots & & x+1-2\zeta_{2n} \end{pmatrix}$$

$$= xK \begin{pmatrix} & -\frac{(n-2+a+z)(n-1+a)}{(2n-4+2a+z)(2n-2+2a+z)} & & \cdots & -\frac{(z+a+1)(a+2)}{(2a+2+z)(2a+4+z)} & \\ x & & x & \ldots x & & x \end{pmatrix}$$
$$- \frac{a+1}{2a+2+z} K \begin{pmatrix} & -\frac{(n-2+a+z)(n-1+a)}{(2n-4+2a+z)(2n-2+2a+z)} & & \cdots & -\frac{(z+a+2)(a+3)}{(2a+4+z)(2a+6+z)} & \\ x & & x & \ldots x & & x \end{pmatrix}$$

$$=: x\, \tilde{G}_{n-1}^{(a,z)}(x) - \frac{a+1}{2a+2+z}\, \tilde{G}_{n-2}^{(a+1,z)}(x)$$

where the last equation defines $\tilde{G}_{n-1}^{(a,z)}(x)$. By the reasoning given above we obtain from (6.1) and (6.2) the representation

$$F_{n-1}^{(a+1,z-2)}(x) = 2^{n-1} \frac{\Gamma(\frac{z}{2}+a+n)}{\Gamma(\frac{z+2}{2}+a)\Gamma(n)}\, \tilde{G}_{n-1}^{(a,z)}(x)$$
$$= \sum_{j=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^j \frac{\Gamma(a+1+j)}{\Gamma(j+1)\Gamma(a+1)} \frac{\Gamma(z+a+j)}{\Gamma(z+a)} \frac{\Gamma(n-j)}{\Gamma(n)}\, C_{n-1-2j}^{(\frac{z+2}{2}+a+j)}(x)$$

and for the polynomial $H_{n+1}(x)$ we have from (6.1)

$$H_{n+1}(x) = (x^2-1) \cdot P_{n-1}(x) = \left[2^{n-1} \frac{\Gamma(\frac{z}{2}+a+n)}{\Gamma(\frac{z+2}{2}+a)\Gamma(n)} \right]^{-1} (x^2-1) F_n^{(a,z)}(x).$$

This yields

$$\xi_\beta(\{x_j\}) = \frac{x_j F_{n-1}^{(a+1,z-2)}(x_j) - \frac{a+1}{n-1} F_{n-1}^{(a+2,z-2)}(x_j)}{\frac{d}{dx}(x^2-1)F_{n-1}^{(a,z)}(x)|_{x=x_j}} \qquad j = 0,\dots,n$$

and completes the proof of Theorem 3.3.

Proof of Lemma 6.1: The proof of Lemma 6.1 is performed by induction and straightforward. However, the calculations which have to be performed are extremely tedious. For the sake of simplicity we will only state the proof in the case $a = 1, z = 2$ which illustrates all essential arguments. The general case can be treated exactly in the same way as the proof given here for $a = 1, z = 2$. Let $f = 0$, then we have ($a = 1,\ z = 2$) using the recurrence formula for the ultraspherical polynomials (see Abramowitz and Stegun (1964), p. 782)

$$\begin{aligned}\tilde{P}_{2m}^{(0)}(x) - \tilde{F}_{2m}^{(0)}(x) &= \frac{1}{2m(2m-1)}\left[4(m+1)(2m-1)xC_{2m-1}^{(3)}(x) - 2m(2m+3)C_{2m-2}^{(3)}(x)\right] \\ &\quad - \frac{1}{2m}\left[4(m+1)xC_{2m-1}^{(3)}(x) - 2(m+2)C_{2m-2}^{(3)}(x)\right] \\ &= -\frac{4}{2m(2m-1)}C_{2m-2}^{(3)}(x)\end{aligned}$$

which shows the assertion for $f = 0$. To go the step from f to $f+1$ $(f \le m-2)$ define $D(f) = \tilde{P}_{2n}^{(f)}(x) - F_{2n}^{(f)}(x)$ and

$$\gamma = (-1)^{f+2}\frac{\Gamma(f+5)\Gamma(2m-f-2)}{\Gamma(4)\Gamma(2m+1)(2m-2f-2)}. \tag{6.3}$$

With this notation we obtain (using again the recurrence formula)

$$\begin{aligned}&\tilde{P}_{2m}^{(f+1)}(x) - F_{2m}^{(f+1)}(x) - D(f) = (-1)^{f+1}\frac{\Gamma(f+5)\Gamma(2m-f-2)}{\Gamma(4)\Gamma(2m+1)}\times \\ &\Big[4(m+1)(2m-f-2)xC_{2m-2f-3}^{(f+4)}(x) - 2m(2m+3)C_{2m-2f-4}^{(f+4)}(x) \\ &-\frac{(2m-f-2)(2m-f-1)}{2m-2f-2}\left\{(4m-2f+2)xC_{2m-2f-3}^{(f+4)}(x) - 2(m+2)C_{2m-2f-3}^{(f+4)}(x)\right\}\Big] \\ &= \gamma\left[(2m-2-f)(f+1)(2f+6)xC_{2m-2f-3}^{(f+4)}(x) - \lambda\, C_{2m-2f-3}^{(f+4)}(x)\right]\end{aligned}$$

where

$$(6.4) \qquad \lambda = (2m+3)[(f+1)(f+2) - 2m] + (2m-f-1)(2m-f-2).$$

Using the induction hypothesis and the recurrence formula of the ultraspherical polynomials (see Abramowitz and Stegun (1964), p. 782) it now follows

$$\tilde{P}_{2m}^{(f)}(x) - F_{2m}^{(f)}(x) = \gamma \Big[(2m-f-2)(f+1) \Big\{ x(2f+6)\, C_{2m-2f-3}^{(f+4)}(x)$$

$$- (2m-2f-2)\, C_{2m-2f-2}^{(f+3)}(x) \Big\} - \lambda\, C_{2m-2f-3}^{(f+4)}(x) \Big]$$

$$= \gamma \cdot \Big[(2m-f-2)(f+1) \Big\{ 2x \Big[(f+3)\, C_{2m-2f-3}^{(f+4)}(x) - (2m-f)\, C_{2m-2f-3}^{(f+3)}(x) \Big]$$

$$+ (2m+2)\, C_{2m-2f-4}^{(f+3)}(x) \Big\} - \lambda\, C_{2m-2f-4}^{(f+4)}(x) \Big]$$

$$= \gamma \cdot \Big[(2m-f-2)(f+1) \Big\{ 2(f+3) x C_{2m-2f-5}^{(f+4)}(x) + (2m+2)\, C_{2m-2f-4}^{(f+3)}(x) \Big\} - \lambda\, C_{2m-2f-4}^{(f+4)}(x) \Big]$$

where the last line results from the identity

$$(6.5) \qquad (n+\alpha)\, C_{n+1}^{(\alpha-1)}(x) = (\alpha - 1) \Big[C_{n+1}^{(\alpha)}(x) - C_{n-1}^{(\alpha)}(x) \Big]$$

for ultraspherical polynomials (Abramowitz and Stegun (1964), formula 22.7.23). By a further application of this formula, the recurrence formula for these polynomials and (6.5) we obtain (using the definitions of γ and λ given in (6.3) and (6.4))

$$\tilde{P}_{2m}^{(f)}(x) - F_{2m}^{(f)}(x) = \gamma \cdot \Big[(2m-f-2)(f+1) \Big\{ \frac{2(f+3)}{4m-2f-2} \Big[(2m-2f-4)\, C_{2m-2f-4}^{(f+4)}(x)$$

$$+ (2m+2)\, C_{2m-2f-6}^{(f+4)}(x) \Big] + 2(m+1)\, C_{2m-2f-4}^{(f+3)}(x) \Big\} - \lambda\, C_{2m-2f-4}^{(f+4)}(x) \Big]$$

$$= \gamma \cdot [2(2m-f-2)(f+1)(f+3) - \lambda]\, C_{2m-2f-4}^{(f+4)}(x)$$

$$= \gamma \cdot (f+5)(2m-2f-2)(f+2)\, C_{2m-2f-4}^{(f+4)}(x)$$

$$= (-1)^{f+2} \frac{\Gamma(f+6)\Gamma(2m-f-2)}{\Gamma(4)\Gamma(2m+1)}\, (f+2)\, C_{2m-2f-4}^{(f+4)}(x)$$

which shows the assertion of Lemma 6.1 ($a = 1,\ z = 2$) for $f + 1$ and completes its proof.

References

M. Abramowitz, I.A. Stegun (1964), Handbook of mathematical functions, Dover Publications, New York.

A.C. Atkinson (1972), Planning experiments to detect inadequate regression models, *Biometrika* **59**, 275–293.

R.D. Cook and C.J. Nachtsheim (1982), Model robust, linear optimal designs, *Technometrics* **24**, 49–54.

H. Dette (1989), Optimale Versuchspläne für mehrere konkurrierende Polynommodelle bei einer gegebenen a–priori Gewichtung. Dissertation Universität Hannover (in German).

H. Dette (1990), A generalization of $D-$ and D_1–optimal design in polynomial regression, *Ann. Statist.* **18**, 1784–1804.

H. Dette (1991), A note on model robust design in polynomial regression, *Journal of Statistical Planning and Inference* **20**, 223–232.

V.V. Fedorov (1971), Theory of optimal experiments, Academic Press, New York.

S. Gutmair (1991), Polar and subgradients of mixtures of information functions. Technical Report No. 286, Schwerpunktprogramm der DFG "Anwendungsbezogene Optimierung und Steuerung." Universität Augsburg.

P.G. Hoel (1958), Efficiency problems in polynomial estimation, *Ann. Math. Statist.* **29**, 1134–1145.

J. Kiefer and J. Wolfowitz (1959), Optimum designs in regression problems, *Ann. Math. Statist.* **30**, 271–294.

J. Kiefer and J. Wolfowitz (1960), The equivalence of two extremum problems, *Canad. J. Math.* **12**, 363–366.

E. Läuter (1974), Experimental design in a class of models, *Math. Operat. Statist.* **5**, 379–398.

E. Läuter (1976), Optimal multipurpose designs for regression models, *Math. Operat. Statist.* **7**, 51–68.

T.S. Lau (1983), Theory of canonical moments and its application in polynomial regression, I and II, Technical Reports 83–23 and 83–24, Dept. of Statistics, Purdue University.

T.S. Lau (1988), D–optimal designs on the unit q–ball, *J. Statist. Plan. Inf.* **19**, 299–315.

T.S. Lau and W.J. Studden (1985), Optimal designs for trigonometric and polynomial regression using canonical moments, *Ann. Statist.* **13**, 383–394.

T.S. Lau and W.J. Studden (1988), On an extremal problem of Fejèr, *Journal of Approximation Theory* **53**, 184–194.

Y.B. Lim and W.J. Studden (1988), Efficient D_s–optimal design for multivariate polynomial regression on the q–cube, *Ann. Statist.* **16**, 1225–1240.

O. Perron (1954), Die Lehre von den Kettenbrüchen (Band I & II), B.G. Teubner, Stuttgart.

S.D. Silvey (1980), Optimal design, Chapman and Hall, London.

M. Skibinsky (1967), The range of the $(n+1)$–th moment for distributions on [0,1], *J. Appl. Probability* **4**, 543–552.

M. Skibinsky (1968), Extreme n–th moments for distributions on [0,1] and the inverse of a moment space map, *J. Appl. Probability* **5**, 693–701.

M. Skibinsky (1969), Some striking properties of binomial and beta moments, *Ann. Math. Statist.* **40**, 1753–1764.

M. Skibinsky (1986), Principal representations and canonical moment sequences for distribution on an interval, *J. Math. Anal. and Applications* **120**, 95–120.

S.M. Stigler (1971), Optimal experimental design for polynomial regression, *J. Amer. Statist. Assoc.* **66**, 311–318.

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

W.J. Studden (1982a), Some robust–type D–optimal designs in polynomial regressions, *J. American Statist. Assoc.* **77**, 916–921.

W.J. Studden (1982b), Optimal designs for weighted polynomial regression using canonical moments, in *Statistical Decision Theory and Related Topics III*, Academic Press, New York, 335–350.

W.J. Studden (1989), Note on some Φ_p optimal design for polynomial regression, *Ann. Statist.* **17**, 618–623.

G. Szegö (1959), Orthogonal polynomials, *American Mathematical Society, Colloquium Publications* **23**.

H.S. Wall (1948), Analytic theory of continued fractions, Van Nostrand, New York.