OPTIMAL DESIGNS FOR POLYNOMIAL REGRESSION WHEN THE DEGREE IS NOT KNOWN

Holger Dette and William J. Studden

Technische Universität Dresden and Purdue University

Abstract. This paper considers the problem of determining efficient designs for polynomial regression models when only an upper bound for the degree of the polynomial is known by the experimenter before the experiments are carried out. The optimality criterion maximizes a weighted p-mean of the relative D-efficiencies in the different models. The optimal (model robust) design is completely determined in terms of its canonical moments which form the unique solution of a system of nonlinear equations. The efficiency of the optimal designs with respect to different criteria is investigated by several examples.

Key words and phrases: Canonical moments, D-efficiency, equivalence theorem, mixture of optimality criteria, polynomial regression.

1. Introduction

Consider the polynomial regression model

$$g_\ell(x) = \sum_{j=0}^{\ell} a_j x^j,$$

where $x \in [-1, 1]$ and $1 \leq \ell \leq n$. The experimenter chooses experimental conditions $x \in [-1, 1]$ and then observes a real valued response with expectation $g_\ell(x)$ and variance σ^2, where different observations are assumed to be uncorrelated. An experimental design is a probability measure on $[-1, 1]$ and the performance of a design ξ in the model g_ℓ, is evaluated through its information matrix

$$M_\ell(\xi) = \int_{-1}^{1} f_\ell(x) f_\ell^T(x) d\xi(x),$$

where $f_\ell(x) = (1, x, \ldots, x^\ell)^T$, $\ell = 1, \ldots, n$. It is well known (Hoel (1958)) that the D-optimal design ξ_ℓ^D in the model $g_\ell(x)$ puts equal masses at the zeros of the polynomial $(x^2 - 1)P_\ell'(x)$ where $P_\ell(x)$ is the ℓth Legendre polynomial. In

order to examine how a given design ξ behaves in the model g_ℓ with respect to the D-optimality criterion one uses the D-efficiency

$$\text{eff}_\ell(\xi) = \left(\frac{|M_\ell(\xi)|}{|M_\ell(\xi_\ell^D)|}\right)^{\frac{1}{\ell+1}} . \tag{1.1}$$

An obvious drawback of the D-optimal design ξ_ℓ^D is that it is not necessarily very efficient in polynomial regression models with degree different from ℓ. As an example consider the D-optimal design ξ_1^D for linear regression which puts equal masses at the points -1 and 1 and has efficiency 0 in the quadratic model. Conversely, the D-optimal design ξ_2^D for the quadratic model has only 82% efficiency in the linear model. Because in many applications of polynomial regression models the degree of the polynomial is not known before the experiments are carried out, the D-optimal design ξ_ℓ^D is not used very often in practice.

In this paper we consider the somewhat more realistic situation that the experimenter knows an upper bound for the polynomial regression, say $n \in I\!N$. In order to find a design which has good efficiencies in all polynomials up to degree n we maximize a concave function of the efficiencies in (1.1). More precisely, we define

$$\Phi_{p,\beta}(\xi) = \left[\sum_{\ell=1}^{n} \beta_\ell (\text{eff}_\ell(\xi))^p\right]^{\frac{1}{p}} , \tag{1.2}$$

where $p \in [-\infty, 1]$ and $\beta = (\beta_1, \ldots, \beta_n)$ is a prior distribution on the set $\{1, \ldots, n\}$ with $\beta_n > 0$ which reflects the experimenters belief about the adequacy of the different models. Here the cases $p = -\infty$ and $p = 0$ have to be understood as the corresponding limits, that is

$$\Phi_{-\infty}(\xi) = \min_{\ell=1}^{n}\{\text{eff}_\ell(\xi) \mid \beta_\ell > 0\}, \qquad \Phi_{0,\beta}(\xi) = \prod_{\ell=1}^{n} (\text{eff}_\ell(\xi))^{\beta_\ell} . \tag{1.3}$$

A design $\xi_{p,\beta}$ is called $\Phi_{p,\beta}$-optimal (with respect to the prior β) if it maximizes the function in (1.2) or (1.3). The case of the geometric mean $p = 0$ was introduced by Läuter (1974) and a solution of this problem in the case of polynomial regression models can be found in Dette (1990).

In this paper we present a complete solution of the $\Phi_{p,\beta}$-optimal design problem for all $p \in [-\infty, 1]$. The $\Phi_{p,\beta}$-optimal design with respect to the prior β is determined as the design whose canonical moments form the unique solution of a system of $n-1$ nonlinear equations. These equations can be solved very easily by standard numerical methods as the Newton Raphson algorithm. The proofs are based on a combination of equivalence theorems for mixtures of information functions (see e.g. Pukelsheim (1993, p. 283-293)), the theory of canonical moments

(see e.g. Studden (1980, 1982) or Lau (1983)) and a one to one correspondence between the set of (symmetric) probability measures on $[-1,1]$ and the set of optimality criteria in (1.2) (see e.g. Dette (1991)). In Section 2 some preliminary results are given which will be needed throughout the paper. Section 3 deals with the case $p > -\infty$. The case $p = -\infty$ (for which the solution of the optimal design problem is more transparent) is treated in Section 4 and some examples are given in Section 5.

2. Preliminaries

The general equivalence theory for mixtures of optimality criteria is described in Pukelsheim (1993, p. 283-293). For the $\Phi_{p,\beta}$-optimality criterion we obtain from these results the following Lemma.

Lemma 2.1. *A design ξ^* is $\Phi_{p,\beta}$-optimal (for some given $p > -\infty$) if and only if it is $\Phi_{0,\beta'}$-optimal with respect to the prior $\beta' = (\beta_1', \ldots, \beta_n')$ where $\beta_l' = \beta_l(\mathrm{eff}_l(\xi^*))^p / \sum_{j=1}^n \beta_j(\mathrm{eff}_j(\xi^*))^p$.*

Let $\mathcal{N}(\xi_{-\infty}) = \{1 \le j \le n \mid \beta_j > 0, \ \Phi_{-\infty}(\xi_{-\infty}) = \mathrm{eff}_j(\xi_{-\infty})\}$, then a design $\xi_{-\infty}$ is $\Phi_{-\infty}$-optimal if and only if there exists a prior $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_l = 0$ for all $l \notin \mathcal{N}(\xi_{-\infty})$ such that $\xi_{-\infty}$ is $\Phi_{0,\alpha}$-optimal.

Equivalence theorems provide a general method for investigating if a given design is optimal and are the basis for many numerical algorithms (see e.g. Wynn (1972), Läuter (1974)). For the special case of polynomial regression the theory of canonical moments provides a very useful tool for the determination of optimal designs (see e.g. Studden (1980, 1982) or Lau (1983)). For a given probability measure ξ on the interval $[-1,1]$ let $c_j = \int_{-1}^1 x^j d\xi(x)$, $j = 0, 1, 2, \ldots$, denote the ordinary moments. If $c_0, \ldots, c_{i-1}$ is a given set of moments (of ξ) define c_i^+ as the maximum of the ith moment over set of all probability measures η with given moments $c_0, \ldots, c_{i-1}$. Similarly let c_i^- denote the corresponding minimum value. The canonical moments are defined by

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

if $c_i^+ > c_i^-$ and are undefined whenever $c_i^+ = c_i^-$. A design ξ on $[-1,1]$ is symmetric if and only if $p_{2i-1} = \frac{1}{2}$ for all $i \in \mathbb{N}$ for which p_{2i-1} is defined (see e.g. Lau (1983)). The determinants of the information matrices $M_\ell(\xi)$ can easily be expressed in terms of the canonical moments of ξ (see Studden (1980) or Lau (1983)) and for a symmetric design on the interval $[-1,1]$ we have as a special case

$$|M_\ell(\xi)| = \prod_{j=1}^{\ell} (q_{2j-2}p_{2j})^{\ell+1-j}, \qquad \text{if } \xi \text{ is symmetric,} \qquad (2.1)$$

where $q_{2j} = 1 - p_{2j}$, $j \geq 1$, and $q_0 = 1$. The canonical moments of the D-optimal design ξ_ℓ^D for the model g_ℓ are given by

$$p_{2j} = \frac{\ell - j + 1}{2(\ell - j) + 1}, \quad p_{2j-1} = \frac{1}{2}, \quad j = 1, \ldots, \ell,$$

and we obtain from (2.1)

$$|M_\ell(\xi_\ell^D)| = \left(\frac{\ell}{2\ell - 1}\right)^\ell \prod_{j=2}^\ell \left(\frac{(\ell - j + 1)^2}{(2(\ell - j) + 1)(2(\ell - j) + 3)}\right)^{\ell + 1 - j}. \tag{2.2}$$

It is well known (see e.g. Lau (1983)) that ξ has canonical moments $0 < p_j < 1$, $j = 1, \ldots, 2n - 1$, $p_{2n} = 1$ if and only if ξ is supported at $n + 1$ points including -1 and 1 (which means that ξ_ℓ^D has $\ell - 1$ support points in the open interval $(-1, 1)$). The following result shows that there exists an intimate relation between these probability measures and the solutions of the $\Phi_{0,\beta}$-optimal design problem and this is an immediate consequence of Theorem 2.3 in Dette (1991).

Theorem 2.2. *Let $\Xi^{(n)}$ denote the class of all symmetric probability measures ξ on $[-1, 1]$ with $n + 1$ support points including -1 and 1 such that*

$$1 - 2\frac{q_{2\ell}}{p_{2\ell}} + \frac{q_{2\ell} q_{2\ell+2}}{p_{2\ell} p_{2\ell+2}} \geq 0, \qquad \ell = 1, \ldots, n, \tag{2.3}$$

(here p_j denote the canonical moments of ξ and $q_{2n+2} = 0$). The mapping

$$\psi: \quad (\beta_1, \ldots, \beta_n) \longrightarrow \xi_{0,\beta} = \arg \max_\xi \Phi_{0,\beta}(\xi)$$

is one to one from the set of all prior distributions $(\beta_1, \ldots, \beta_n)$ on $\{1, \ldots, n\}$ with $\beta_n > 0$ onto the set $\Xi^{(n)}$. Moreover, if $\xi \in \Xi^{(n)}$ has canonical moments (of even order) $p_2, \ldots, p_{2n-2}, p_{2n} = 1$, then the inverse of ψ is given by $\psi^{-1}(\xi) = (\beta_1^*, \ldots, \beta_n^*)$ where

$$\beta_\ell^* = \frac{\ell + 1}{2 - (q_2/p_2)} \prod_{j=1}^{\ell-1} \frac{q_{2j}}{p_{2j}} \left(1 - 2\frac{q_{2\ell}}{p_{2\ell}} + \frac{q_{2\ell} q_{2\ell+2}}{p_{2\ell} p_{2\ell+2}}\right), \qquad \ell = 1, \ldots, n. \tag{2.4}$$

3. $\Phi_{p,\beta}$-Optimal Designs

In this section we consider the criterion (1.2) for all $p \in (-\infty, 1]$. The case $p = 0$ was already solved by Dette (1990), the general case $(p \neq 0)$ is more complicated and stated in the following theorem.

Theorem 3.1. *Let $p \in (-\infty, 1]$, then the $\Phi_{p,\beta}$-optimal design is uniquely determined by its canonical moments $(\frac{1}{2}, p_2, \frac{1}{2}, \ldots, p_{2n-2}, \frac{1}{2}, 1)$ where $(p_2, \ldots, p_{2n-2})$ is the unique solution of the system of equations*

$$\frac{\beta_{\ell+1}}{\ell+2}\left(1 - 2\frac{q_{2\ell}}{p_{2\ell}} + \frac{q_{2\ell}q_{2\ell+2}}{p_{2\ell}p_{2\ell+2}}\right)\left(\prod_{j=1}^{\ell+1}(q_{2j-2}p_{2j})^j\right)^{p/(\ell+1)(\ell+2)}$$

$$= \frac{\beta_\ell}{\ell+1}\frac{q_{2\ell}}{p_{2\ell}}\left(1 - 2\frac{q_{2\ell+2}}{p_{2\ell+2}} + \frac{q_{2\ell+2}q_{2\ell+4}}{p_{2\ell+2}p_{2\ell+4}}\right)C_\ell^p, \qquad \ell = 1, \ldots, n-1, \quad (3.1)$$

which satisfies (2.3). Here

$$C_\ell = \frac{|M_{\ell+1}(\xi_{\ell+1}^D)|^{\frac{1}{\ell+2}}}{|M_\ell(\xi_\ell^D)|^{\frac{1}{\ell+1}}}$$

$$= \left[\frac{\ell^{\ell^2}(\ell+1)^{(\ell+1)^2}(2\ell-1)^\ell}{(2\ell+1)^{(\ell+1)(2\ell+1)}}\prod_{j=2}^{\ell}\left\{\frac{(\ell+1-j)^2}{(2(\ell-j)+1)(2(\ell-j)+3)}\right\}^{-(\ell+1-j)}\right]^{1/(\ell+1)(\ell+2)}, \quad (3.2)$$

$\ell = 1, \ldots, n-1$, and the ℓth equation in (3.1) has to be replaced by the equation

$$1 - 2\frac{q_{2\ell}}{p_{2\ell}} + \frac{q_{2\ell}q_{2\ell+2}}{p_{2\ell}p_{2\ell+2}} = 0 \qquad (3.3)$$

whenever $\beta_\ell = 0$, $\ell = 1, \ldots, n-1$.

Proof. Let $p_2, \ldots, p_{2n-2}$ denote the canonical moments (of even order) of the $\Phi_{p,\beta}$-optimal design $\xi_{p,\beta}$. From Lemma 2.1 it follows that $\xi_{p,\beta}$ is $\Phi_{0,\beta'}$-optimal where the prior distribution $\beta' = (\beta_1', \ldots, \beta_n')$ is given by $\beta_\ell' = \beta_\ell(\text{eff}_\ell(\xi_{p,\beta}))^p / \sum_{j=1}^n \beta_j(\text{eff}_j(\xi_{p,\beta}))^p$. Because the map ψ in Theorem 2.2 is one to one we have

$$(\beta_1', \ldots, \beta_n') = \psi^{-1}(\xi_{p,\beta}) = (\beta_1^*, \ldots, \beta_n^*), \qquad (3.4)$$

where β_ℓ^* is defined in (2.4) and consequently the canonical moments of $\xi_{p,\beta}$ satisfy (2.3). On the other hand, if $\beta_\ell \neq 0$, we obtain from (3.4) and (2.4)

$$\frac{\beta_{\ell+1}(\text{eff}_{\ell+1}(\xi_{p,\beta}))^p}{\beta_\ell(\text{eff}_\ell(\xi_{p,\beta}))^p} = \frac{\beta_{\ell+1}'}{\beta_\ell'} = \frac{(\ell+2)\prod_{j=1}^{\ell}\frac{q_{2j}}{p_{2j}}\left(1 - 2\frac{q_{2\ell+2}}{p_{2\ell+2}} + \frac{q_{2\ell+2}q_{2\ell+4}}{p_{2\ell+2}p_{2\ell+4}}\right)}{(\ell+1)\prod_{j=1}^{\ell-1}\frac{q_{2j}}{p_{2j}}\left(1 - 2\frac{q_{2\ell}}{p_{2\ell}} + \frac{q_{2\ell}q_{2\ell+2}}{p_{2\ell}p_{2\ell+2}}\right)},$$

which is equivalent to (3.1). If $\beta_\ell = 0$, (3.3) follows directly from (3.4) and (2.4). This shows that the canonical moments (of even order) of $\xi_{p,\beta}$ form a solution of the system of equations defined in Theorem 3.1.

Finally, let $(p_2^*, \ldots, p_{2n-2}^*)$ denote a second solution of the system of equations in Theorem 3.1 that satisfies (2.3) and let $\xi^* \in \Xi^{(n)}$ denote the corresponding design. By Theorem 2.2 it follows that ξ^* is Φ_{0,β^*}-optimal for the prior $\beta^* = (\beta_1^*, \ldots, \beta_n^*)$ where $(p_{2n}^* = 1, q_{2n+2}^* = 0)$

$$\beta_\ell^* \;=\; \frac{\ell+1}{1 - q_2^*/p_2^*} \prod_{j=1}^{\ell-1} \frac{q_{2j}^*}{p_{2j}^*} \left(1 - 2\frac{q_{2\ell}^*}{p_{2\ell}^*} + \frac{q_{2\ell}^* q_{2\ell+2}^*}{p_{2\ell}^* p_{2\ell+2}^*} \right), \qquad \ell = 1, \ldots, n. \qquad (3.5)$$

An application of Lemma 2.1 shows that ξ^* is $\Phi_{p,\tilde\beta}$-optimal with respect to the prior $\tilde\beta = (\tilde\beta_1, \ldots, \tilde\beta_n)$ where

$$\tilde\beta_\ell \;=\; \frac{\beta_\ell^* \left(\mathrm{eff}_\ell(\xi^*)\right)^{-p}}{\sum_{\ell=1}^n \beta_\ell^* \left(\mathrm{eff}_\ell(\xi^*)\right)^{-p}} \;=\; \beta_\ell, \qquad \ell = 1, \ldots, n.$$

Here the last identity is a consequence of (2.2), (3.5) and the fact that $(p_2^*, \ldots, p_{2n-2}^*)$ is a solution of the system of equations in Theorem 3.1. It follows from standard arguments of optimal design theory, that the $\Phi_{p,\beta}$-optimal design is unique and consequently we conclude that $\xi^* = \xi_{p,\beta}$. But this is equivalent to the fact $p_{2j}^* = p_{2j}$, $j = 1, \ldots, n$, and proves the assertion of the theorem.

Remark 3.2. It is worthwhile to mention that a more complicated proof of Theorem 3.1 can be obtained from Theorem 3.3 in Dette (1994) by observing that a $\Phi_{p,\beta}$-optimal design ξ^* is also a $\Phi_{p,\gamma}^c$-optimal discriminating design (in the sense of Dette (1994)) where γ_ℓ is proportional to

$$\frac{|M_{\ell-1}(\xi^*)|^p}{|M_\ell(\xi^*)|^p} \sum_{j=l}^n \frac{\beta_j \left(\mathrm{eff}(\xi^*)\right)^p}{j+1} \;.$$

It should also be noted that, in general, every $\Phi_{p,\gamma}^c$-optimal discriminating design is also $\Phi_{0,\beta}$-optimal for an appropriate prior β but not necessarily $\Phi_{p,\beta'}$-optimal for $p \neq 0$ (in the case of negative weights β_l Lemma 2.1 is not applicable).

In general, the system of equations in Theorem 3.1 has to be solved numerically except in the case $p = 0$ where it can be shown that the solution of (3.1) and (3.2) is given by

$$p_{2j} = \frac{\sum_{\ell=j}^n \frac{\ell+1-j}{\ell+1} \beta_\ell}{\sum_{\ell=j}^n \frac{\ell+1-j}{\ell+1} \beta_\ell + \sum_{\ell=j+1}^n \frac{\ell-j}{\ell+1} \beta_\ell}, \qquad j = 1, \ldots, n-1,$$

which is the result of Dette (1990, p. 1789). A further simplification occurs if $\beta_{n-1} = \cdots = \beta_k = 0$, $k \leq n-1$. In this case the canonical moments of the $\Phi_{p,\beta}$-optimal design have a similar behavior as in the D_s-optimal design problem (see

Studden (1980)). The proof of the following result is an immediate consequence of Theorem 3.1 and is therefore omitted.

Theorem 3.3. *Let* $\beta = (\beta_1, \ldots, \beta_{k-1}, 0, \ldots, 0, \beta_n)$, $1 \leq k \leq n$, *and* $p \in (-\infty, 1]$, *then the* $\Phi_{p,\beta}$-*optimal design is uniquely determined by its canonical moments* $p_2, \ldots, p_{2n-2}, p_{2n}$ *where* $p_{2k}, p_{2k+2}, \ldots, p_{2n}$ *are given by*

$$p_{2j} = \frac{n-j+1}{2(n-j)+1}, \qquad j = n, n-1, \ldots, k, \qquad (3.6)$$

while $(p_2, \ldots, p_{2k-2})$ *is the unique solution of the first* $k-1$ *equations in Theorem 3.1.*

4. Optimal Designs Which Maximize the Minimum Efficiency

In this section we consider the criterion

$$\Phi_{-\infty}(\xi) = \min_{j=1}^{n}\{\mathrm{eff}_j(\xi)\} \qquad (4.1)$$

for which the solution of the design problem is more transparent than in the general case. Throughout this section we make frequent use of the quantities

$$a_\ell = \begin{cases} \dfrac{|M_{\ell+1}(\xi_{\ell+1}^D)||M_{\ell-1}(\xi_{\ell-1}^D)|}{|M_\ell(\xi_\ell^D)|^2}, & \text{if } \ell = 2, 3, \ldots, n-1, \\ |M_2(\xi_2^D)|^2, & \text{if } \ell = 1, \end{cases} \qquad (4.2)$$

which can be rewritten as (using (2.2))

$$a_\ell = \begin{cases} \dfrac{(\ell+1)^{\ell+1}(2\ell-1)^{2\ell-1}}{(\ell-1)^{\ell-1}(2\ell+1)^{2\ell+1}}, & \text{if } \ell = 2, \ldots, n, \\ \dfrac{2^4}{3^6}, & \text{if } \ell = 1. \end{cases} \qquad (4.3)$$

In the following Lemma we collect some of the properties of the sequence $(a_\ell)_{\ell \in \mathbb{N}}$. Its proof is straightforward and therefore omitted.

Lemma 4.1. *The sequence* $(a_\ell)_{\ell \in \mathbb{N}}$ *is increasing, bounded,* $(1/5 < a_\ell < 1/4$, *for all* $\ell \geq 2)$, *and has the limit* $1/4$.

Theorem 4.2. *The* $\Phi_{-\infty}$-*optimal design* $\xi_{-\infty}$ *is uniquely determined by its canonical moments* $(1/2, p_2, 1/2, \ldots, 1/2, p_{2n-2}, 1/2, 1)$, *where the canonical moments (of even order)* $p_2, \ldots, p_{2n-2}$ *are given by the continued fractions*

$$p_{2\ell} = 1 - \frac{a_\ell|}{|\ 1} - \frac{a_{\ell+1}|}{|\ 1} - \ldots - \frac{a_{n-1}|}{|\ 1}, \qquad \ell = 2, \ldots, n-1, \qquad (4.4)$$

and p_2 is the largest root in the interval $[0, 1]$ of the equation

$$p_2(1 - p_2)^2 = \frac{16}{729 p_4^2}. \tag{4.5}$$

Proof. The proof consists of three steps. In Step 1 we show that the set $\mathcal{N}(\xi_{-\infty})$ defined in the second part of Lemma 2.1 is precisely $\{1, \dots, n\}$, in Step 2 we prove that the quantities in (2.3) are all nonnegative for the design $\xi_{-\infty}$ and finally in Step 3 we apply the results of Section 2 in order to establish the assertion of the theorem.

Step 1 ($\mathcal{N}(\xi_{-\infty}) = \{1, \dots, n\}$). For $\ell = 2, \dots, n - 1$ consider the equations

$$\mathrm{eff}_\ell(\xi) = \mathrm{eff}_{\ell+1}(\xi), \qquad \ell = 2, \dots, n - 1, \tag{4.6}$$

where ξ is a symmetric design supported at $n + 1$ points including -1 and 1 (that is ξ has canonical moments $(1/2, p_2, 1/2, \dots, 1/2, p_{2n-2}, 1/2, 1)$). By an application of (2.1) and straightforward algebra we find that (4.6) is equivalent to

$$q_{2\ell} p_{2\ell+2} = a_\ell, \qquad \ell = 2, \dots, n - 1, \tag{4.7}$$

$(p_{2n} = 1)$ which can easily be rewritten as (4.4). Similarly, it follows that the equation $\mathrm{eff}_1(\xi) = \mathrm{eff}_2(\xi)$ is equivalent to (4.5). This implies that the design $\xi_{-\infty}$ defined in (4.4) and (4.5) has equal efficiency in all models up to degree n, that is $\mathcal{N}(\xi_{-\infty}) = \{1, \dots, n\}$.

Step 2 ($\xi_{-\infty} \in \Xi^{(n)}$). Obviously $\xi_{-\infty}$ is symmetric and supported at $n + 1$ points including -1 and 1. In order to show that $\xi_{-\infty}$ satisfies (2.3) we consider, first, the case $\ell = 2, \dots, n - 1$ and rewrite (2.3) as

$$1 - 3q_{2\ell} p_{2\ell+2} - p_{2\ell} q_{2\ell+2} \geq 0,$$

where we have used $p_{2\ell} = 1 - q_{2\ell}$. Observing (4.7) we obtain

$$1 - 3a_\ell - p_{2\ell}\left(1 - \frac{a_\ell}{1 - p_{2\ell}}\right) \geq 0,$$

which is equivalent to the inequality

$$p_{2\ell}^2 - 2p_{2\ell}(1 - 2a_\ell) + (1 - 3a_\ell) \geq 0. \tag{4.8}$$

The minimum of the left hand side in (4.8) (as a function of $p_{2\ell}$) is attained at $p_{2\ell} = 1 - 2a_\ell$ and given by $a_\ell(1 - 4a_\ell)$ which is positive because of Lemma 4.1. This proves that the canonical moments of $\xi_{-\infty}$ satisfy (2.3) for $\ell = 2, \dots, n - 1$.

In order to show the remaining case $\ell = 1$ we remark that it is easy to see that the canonical moments of $\xi_{-\infty}$ are all greater than $1/2$ (here we use (4.7), Lemma 4.1 and the assumption that p_2 is defined as the largest root of (4.5) in the interval $[0,1]$). By a procedure similar to the above (using (4.5) instead of (4.7)) we find that (2.3) for $\ell = 1$ is equivalent to the inequality

$$f(p_2) \;=\; 27p_2^{\frac{5}{2}} - 54p_2^{\frac{3}{2}} + 16p_2 + 27p_2^{\frac{1}{2}} - 12 \;\geq\; 0, \qquad p_2 > \frac{1}{2}.$$

It is easy to see that f is an increasing function of $p_2 \in [1/2, 1]$ and consequently it follows that $f(p_2) > f(1/2) > 0$, which shows that the canonical moments of $\xi_{-\infty}$ satisfy (2.3).

Step 3 (Proof of Theorem 4.2). From Step 2 we have $\xi_{-\infty} \in \Xi^{(n)}$ and by Theorem 2.2 we find that $\xi_{-\infty}$ is Φ_{0,β^*}-optimal, where $\beta^* = \psi^{-1}(\xi_{-\infty}) = (\beta_1^*, \ldots, \beta_n^*)$ is defined in (2.4). In Step 1 we showed that $\mathcal{N}(\xi_{-\infty}) = \{1, \ldots, n\}$ and consequently $\xi_{-\infty}$ satisfies the condition in the second part of Lemma 2.1 with $\alpha = \beta^*$. This proves the $\Phi_{-\infty}$-optimality of $\xi_{-\infty}$.

In the following sections we see that the $\Phi_{-\infty}$-optimal design serves as an appropriate approximation for the $\Phi_{p,\beta}$-optimal design when p is sufficiently small. For this reason we state some properties of the canonical moments of the $\Phi_{-\infty}$-optimal design in the following Lemma. The proof is omitted for the sake of brevity.

Lemma 4.3. *Let $p_{2j}^{(n)}$ denote the canonical moments (of even order) of the $\Phi_{-\infty}$-optimal design for polynomial regression models up to degree n. The following statements hold true*
(a) *$p_{2j}^{(n)} \geq 1/2$ for all $j = 1, \ldots, n$ and $n \in I\!\!N$*
(b) *$p_{2j}^{(n)} < p_{2j}^{(n-1)}$ for all $n \in I\!\!N$*
(c) *If $n > 2$, there exists an index j_0 such that*
$$1 = p_{2n}^{(n)} > p_{2n-2}^{(n)} > \cdots > p_{2j_0}^{(n)} < p_{2j_0+2}^{(n)} < \cdots < p_4^{(n)} < p_2^{(n)}.$$

The following result gives the limit distribution of the $\Phi_{-\infty}$-optimal design as $n \to \infty$. It shows that the limit is NOT the arcsin-distribution in contrast with the case $p = 0$ and the uniform prior (see Dette (1990, p. 1797)).

Theorem 4.4. *If $n \to \infty$, then the $\Phi_{-\infty}$-optimal design converges weakly to a symmetric distribution ξ^* with canonical moments (of even order) $p_2, p_4, \ldots$ where for $\ell \geq 2$, $p_{2\ell}$ is given by the (infinite) continued fraction*

$$p_{2\ell} = 1 - \frac{a_\ell|}{|\; 1} - \frac{a_{\ell+1}|}{|\; 1} - \cdots, \qquad \ell \geq 2,$$

and p_2 is the largest root in $[0,1]$ of equation (4.5).

Proof. For fixed n the canonical moments of the $\Phi_{-\infty}$-optimal design for polynomials up to degree n are given in (4.4) and (4.5). By Lemma 4.1 the quantities a_ℓ in (4.4) satisfy $a_\ell < 1/4, \ \ell \geq 2$, and by Worpitzky's Theorem (see Wall (1948, p. 42)) the continued fraction in (4.4) converges. This proves the assertion.

Remark 4.5. Numerical calculations yield for the first two canonical moments of the limiting design ξ^*, $p_2 = 0.68563939$, $p_4 = 0.56914133$ while the canonical moments of higher order can be calculated recursively from $p_{2\ell+2} = a_\ell/q_{2\ell}, \ell \geq 2$. For example we obtain $p_6 = 0.5414, p_8 = 0.5296, p_{10} = 0.5230, \ldots$ (note that $\lim_{\ell \to \infty} p_{2\ell} = 1/2$). It is also worthwhile to mention that the sequence of the canonical moments of the limiting design ξ^* is strictly decreasing in contrast to the sequence of canonical moments of the $\Phi_{-\infty}$-optimal design for polynomials up to degree n. Figure 1 shows the density of the limiting distribution ξ^* (solid line) together with the arc-sin density $1/\pi\sqrt{1-x^2}$ (dashed line). The arc-sin density is well known to be the limiting density of similar sequences of designs. For example if η_n denotes the D-optimal design for n^{th} degree polynomial regression then η_n converges weakly to the arc-sin law. Note that the limiting density of ξ^* has less mass near the center and more near the end points ± 1 than the arc-sin law.
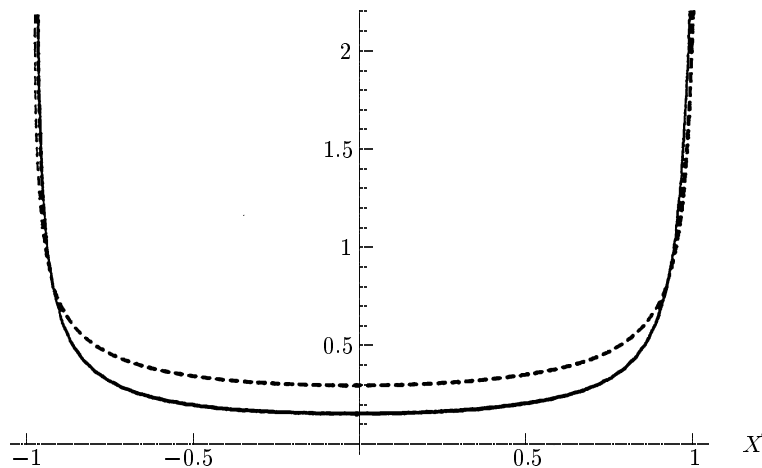

Figure 1. Solid line = density of ξ^*; dashed line = arc-sin density

If the minimum in the optimality criterion (4.1) is not taken over the full index set $\{1, \ldots, n\}$ then the solution of the optimal design problem becomes more complicated. For the sake of brevity we restrict ourselves to the following two special cases which can be proved by similar arguments as presented in the proof of Theorem 4.2.

Theorem 4.6. *The design which maximizes*

$$\min\{\operatorname{eff}_\ell(\xi) \mid \ell = 1, \ldots, k-1, n\}$$

$(2 \le k \le n-1)$ *is uniquely determined by its canonical moments* $p_2, \ldots, p_{2n}$ *where* $p_{2k}, p_{2k+2}, \ldots, p_{2n}$ *are given by* (3.6), $p_4, \ldots, p_{2k-2}$ *are given by the continued fractions*

$$p_{2\ell} = 1 - \frac{a_\ell|}{|~1} - \cdots - \frac{a_{k-2}|}{|~1} - \frac{a_{k-1}^*|}{|~1}, \qquad \ell = 2, \ldots, k-1,$$

with a_l *defined in* (4.2) $(l = 2, \ldots, k-2)$,

$$a_{k-1}^* = \left[\prod_{j=1}^{k-1} \left(\frac{n-j}{2(n-j)+1} \right)^{n-j} \left(\frac{n-j+1}{2(n-j)+1} \right)^{n-j+1} \right]^{\frac{1}{n+1-k}} \frac{|M_{k-2}(\xi_{k-2}^D)|}{|M_{k-1}(\xi_{k-1}^D)|^{\frac{n+2-k}{n+1-k}}}$$

and p_2 *is the largest root in* $[0, 1]$ *of the equation*

$$p_2(1 - p_2)^2 = \frac{16}{729 p_4^2}$$

if $k \ge 3$, *and the largest root of the equation*

$$p_2(1 - p_2)^2 = \left(\frac{n}{2n-1} \right)^{\frac{2n}{n-1}} \left(\frac{2n-3}{2n-1} \right)^2$$

if $k = 2$.

Theorem 4.7. *The design which maximizes*

$$\min\{\operatorname{eff}_\ell(\xi) \mid \ell = k, \ldots, n\}$$

$(1 \le k \le n)$ *is uniquely determined by its canonical moments* $p_2, \ldots, p_{2n}$ *where* $p_{2k+2}, \ldots, p_{2n}$ *are given by the continued fractions*

$$p_{2\ell} = 1 - \frac{a_\ell|}{|~1} - \cdots - \frac{a_{n-2}|}{|~1} - \frac{a_{n-1}|}{|~1}, \qquad \ell = k+1, \ldots, n,$$

(with a_l *defined in* (4.2)*) and* $p_2, \ldots, p_{2k}$ *are the unique solution of the system of equations*

$$p_{2l} = \frac{3p_{2l+2} - 1}{4p_{2l+2} - 1}, \qquad l = k-1, \ldots, 1,$$

$$\frac{C_k^{(k+1)(k+2)}}{(p_{2k+2})^{(k+1)}} = \prod_{j=1}^{k} (q_{2j-2} p_{2j})^j (q_{2k})^{k+1}$$

(*with* C_k *defined in* (3.2)).

5. Examples

5.1. Optimal designs with respect to various $\Phi_{p,\beta}$-criteria

Consider the case $n = 2$ (linear or quadratic regression) and a uniform prior $\beta_1^u = \beta_2^u = 1/2$. In this case there is one equation for the determination of p_2 in Theorem 3.1, namely

$$\left(1 - \frac{2q_2}{p_2}\right)(p_2 q_2^2)^{p/6} = \frac{3}{2}\frac{q_2}{p_2} \cdot \left(\frac{16}{729}\right)^{p/6} \tag{5.1}$$

and the optimal $\Phi_{p,\beta}$-optimal design has canonical moments $(1/2, p_2, 1/2, 1)$ where p_2 is the unique root of (5.1) such that (2.3) is satisfied, i.e. $p_2 \geq 2/3$. There is a considerable amount of literature concerning the relationship between the sequence of canonical moments and the corresponding design (see e.g. Lau (1983)). Throughout this chapter we use Lemma 4.4 in Lim and Studden (1988) which is applicable for polynomial regression up to degree 4. Table 5.1 gives the weights of the Φ_{p,β^u}-optimal design and the D-efficiencies in the linear and quadratic model for different values of $p \in [-\infty, 1]$. The case $p = -\infty$ can be directly obtained from the equation (4.5) in Theorem 4.2 which can be interpreted as the limit of (5.1) when $p \to -\infty$. Note that all designs are supported at $-1, 0, 1$.

Table 5.1. Weights of the Φ_{p,β^u}-optimal design for linear and quadratic regression using a uniform prior β^u

p	$\xi_{p,\beta^u}(\{\pm 1\})$	$\xi_{p,\beta^u}(\{0\})$	$\mathrm{eff}_1(\xi_{p,\beta^u})$	$\mathrm{eff}_2(\xi_{p,\beta^u})$
1	0.38515	0.22970	0.8776	0.9725
0	0.38889	0.22222	0.8819	0.9681
−1	0.39208	0.21584	0.8855	0.9641
−2	0.39478	0.21044	0.8886	0.9603
−3	0.39707	0.20586	0.8911	0.9570
−∞	0.41910	0.16180	0.9155	0.9155

The results in Table 5.1 demonstrate that there do not exist essential differences between the Φ_{p,β^u}-optimal designs for polynomials up to degree 2, with respect to different values of p. We observe a similar behavior in the cases $n = 3$ (linear, quadratic and cubic regression). Here Theorem 3.1 gives two equations for (p_2, p_4)

$$\left(1 - 2\frac{q_2}{p_2} + \frac{q_2 q_4}{p_2 p_4}\right) (p_2(q_2 p_4)^2)^{p/6} = \frac{3}{2}\frac{q_2}{p_2}\left(1 - 2\frac{q_4}{p_4}\right) C_1^p$$
$$\left(1 - \frac{2q_4}{p_4}\right) \{p_2(q_2 p_4)^2 q_4^3\}^{p/12} = \frac{4}{3}\frac{q_4}{p_4} C_2^p \tag{5.2}$$

and p_2, p_4 have to satisfy (2.3). The optimal design puts masses $\alpha, (1/2) - \alpha, (1/2) - \alpha, \alpha$ at the points $-1, -t, t$ and 1 where $t = p_2 q_4$ and $\alpha = p_2 p_4/(2(q_2 + p_2 p_4))$ (see Lim and Studden (1988, p. 1233)). The solution of (5.2) was determined using the Newton Raphson algorithm and the corresponding designs and efficiencies are given in Table 5.2. Again we observe some robustness of the design with respect to different optimality criteria Φ_{p,β^u}.

Table 5.2. Φ_{p,β^u}-optimal designs for polynomials up to degree 3 and a uniform prior. First two columns: weights, third column: interior positive support point

p	$\xi_{p,\beta^u}(\pm 1)$	$\xi_{p,\beta^u}(\pm t)$	t	$\mathrm{eff}_1(\xi_{p,\beta^u})$	$\mathrm{eff}_2(\xi_{p,\beta^u})$	$\mathrm{eff}_3(\xi_{p,\beta^u})$
1	0.31501	0.18499	0.40193	0.8305	0.9138	0.9594
0	0.31944	0.18056	0.40105	0.8348	0.9143	0.9542
-1	0.32345	0.17655	0.40059	0.8388	0.9134	0.9494
-2	0.32703	0.17297	0.40047	0.8423	0.9141	0.9448
-3	0.33021	0.16979	0.40059	0.8455	0.9137	0.9407
$-\infty$	0.36634	0.13366	0.42695	0.8840	0.8840	0.8840

Obviously the robustness of the $\Phi_{p,\beta}$-optimal design with respect to different values of p will also depend on the prior β. As an example for a stronger dependence of the design $\xi_{p,\beta}$ on the parameter p we consider the case $n = 3$ (linear, quadratic or cubic regression) and the prior $\tilde{\beta}_1 = 3/16, \tilde{\beta}_2 = 12/16, \tilde{\beta}_3 = 1/16$ (more weight on the linear and quadratic model). The results are listed in Table 5.3.

Table 5.3. $\Phi_{p,\tilde{\beta}}$-optimal designs for polynomial regression up to degree 3 and the prior $\tilde{\beta} = (3/16, 12/16, 1/16)$. First two columns: weights, third column: interior positive support point

p	$\xi_{p,\tilde{\beta}}(\pm 1)$	$\xi_{p,\tilde{\beta}}(\pm t)$	t	$\mathrm{eff}_1(\xi_{p,\tilde{\beta}})$	$\mathrm{eff}_2(\xi_{p,\tilde{\beta}})$	$\mathrm{eff}_3(\xi_{p,\tilde{\beta}})$
1	0.34203	0.15797	0.16290	0.8321	0.9855	0.6828
0	0.34167	0.15833	0.19124	0.8336	0.9833	0.7327
-1	0.34178	0.15822	0.21194	0.8353	0.9758	0.7645
-2	0.34228	0.15772	0.22807	0.8372	0.9719	0.7864
-3	0.34304	0.15696	0.24122	0.8392	0.9684	0.8025
$-\infty$	0.36634	0.13366	0.42695	0.8840	0.8840	0.8840

5.2. Robustness of the $\Phi_{0,\beta}$-optimal design

The results of Example 5.1 indicate that a given Φ_{p,β^u}-optimal design for the uniform prior β^u is quite robust with respect to different Φ_{p',β^u}-criteria. Because the Φ_{0,β^u}-optimal designs are very easy to calculate (Dette (1990)) it might be of interest how these designs behave with respect to the other $\Phi_{p,\beta}$-criteria. As a representative example we consider the case $n = 4$, $\beta_1^u = \cdots = \beta_4^u = \frac{1}{4}$. It follows from the results of Dette (1990) that the Φ_{0,β^u}-optimal design puts masses $0.27167, 0.10354, 0.24958, 0.10354, 0.27167$ at the points $-1, -0.60508, 0, 0.60508$ and 1 respectively. The performance of the design ξ_{0,β^u} with respect to the other Φ_{p,β^u} criteria is evaluated through the Φ_{p,β^u}-efficiency

$$R_p(\xi) = \frac{\Phi_{p,\beta^u}(\xi)}{\Phi_{p,\beta^u}(\xi_{p,\beta^u})}, \qquad p \in [-\infty, 1],$$

where the design ξ_{p,β^u} is determined by Theorems 3.1 and 4.2.

The results are illustrated in Table 5.4 and show a remarkable robustness of ξ_{0,β^u}-optimal design with respect to the other Φ_{p,β^u} criteria. For this reason and because of the easy computation of the $\Phi_{0,\beta}$-optimal designs we conclude with the statement that the design ξ_{0,β^u} might be a good choice in polynomial regression models when only an upper bound on the degree of the polynomial is known and a uniform prior is used to reflect the experimenters belief about the adequacy of the different models. It should also be mentioned again that this statement is not necessarily true for arbitrary prior distributions β.

Table 5.4. Φ_{p,β^u}-efficiencies of the Φ_{0,β^u}-optimal design for different values of $p \in [-\infty, 1]$

p	1	0.6	−0.6	−1	−2	−3	−∞
$R_p(\xi_{0,\beta^u})$	0.99989	0.99995	0.99996	0.99989	0.99957	0.99906	0.93220

References

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

Dette, H. (1991). A note on robust designs for polynomial regression. J. Statist. Plann. Inference **28**, 223-232.

Dette, H. (1994). Discrimination designs for polynomial regression on compact intervals. Ann. Statist. **22**, 890-903.

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

Lau, T. S. (1983). Theory of canonical moments and its application in polynomial regression. Technical Report 83-23, Department of Statistics, Purdue University.

Läuter, E. (1974). Experimental design in a class of models. Mathematische Operationsforschung und Statistik **5**, 379-398.

Lim, Y. B. and Studden, W. 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. John Wiley, New York.

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

Studden, W. J. (1982). Some robust type D-optimal designs in polynomial regression. J. Amer. Statist. Assoc. **77**, 916-921.

Wall, H. S. (1948). Analytic Theory of Continued Fractions. Van Nostrand, New York.

Wynn, H. P. (1972). Results in the theory and construction of D-optimum experimental designs. J. Roy. Statist. Soc. Ser.B **34**, 133-147.

Institut für Mathematische Stochastik, Technische Universität Dresden, Mommsenstr. 13, 01062 Dresden, Germany.

Department of Statistics, Purdue University, West Lafayette, IN 47907-1399, U.S.A.