A note on maximin and Bayesian D-optimal designs in weighted polynomial regression

Stefanie Biedermann
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: stefanie.biedermann@ruhr-uni-bochum.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 70 94 559

February 21, 2003

Abstract

We consider the problem of finding D-optimal designs for estimating the coefficients in a weighted polynomial regression model with a certain efficiency function depending on two unknown parameters, which models the heteroscedastic error structure. This problem is tackled by adopting a Bayesian and a maximin approach, and optimal designs supported on a minimal number of support points are determined explicitly.

Keywords and Phrases: maximin optimality, Bayesian optimal designs, efficiency function, parameter estimation, Jacobi polynomials

AMS Subject Classification: 62K05, 33C45

1 Introduction

Consider the problem of designing a regression experiment with polynomial mean and heteroscedastic error structure. In most circumstances, the exact variance function will not be available in advance but it is possible to specify its general structure apart from a finite number of parameters. This leads to the following model setup for the response Y at a certain level of the independent variable x assumed to lie in a given design space χ,

$$(1.1) \qquad Y(x) = \sum_{k=0}^{n} \gamma_k x^k + \frac{\sigma\,\varepsilon}{\sqrt{\lambda(x,\vartheta)}}, \quad \mathrm{E}[Y(x)] = \sum_{k=0}^{n} \gamma_k x^k, \quad \mathrm{Var}[Y(x)] = \frac{\sigma^2}{\lambda(x,\vartheta)},$$

where $\gamma^T = (\gamma_0, \ldots, \gamma_n)$ denotes the vector of the parameters of interest and σ is a positive constant. For different observations the errors ε are assumed i.i.d. and standard normal. The function $\lambda(x, \vartheta)$, which is proportional to the inverse of the variance function, is commonly called efficiency function in the design literature [see Fedorov (1972)]. So there are two sets of parameters in the model, the parameters, which are summarized in the vector γ and are of primary interest and furthermore the nuisance parameters in the vector ϑ, which model the uncertainty of the experimenter about the heteroscedastic error structure.
The efficiency function to be considered in the following was first discussed in Antille, Dette and Weinberg (2002). These authors considered the design space $\chi = I\!R$ and the efficiency function

$$\lambda(x, \vartheta) = (1 + x^2)^{\alpha+1} \exp(2\beta \arctan x), \quad \alpha < -n - 1, \ \beta \in I\!R, \tag{1.2}$$

which gives some kind of U-shaped dependency of the model variance on x, and determined the locally D-optimal design $\xi_{\alpha,\beta}$ in the sense of Chernoff (1953). The specific case $\beta = 0, \ \alpha < -n - 1$ corresponds to a rational regression model and locally D-optimal designs for this model were determined by Dette, Haines and Imhof (1999). Locally optimal designs have been critized by numerous authors because of their lack of robustness against misspecifications of the initial parameter $\vartheta = (\alpha, \beta)$ [see e.g. Chaloner and Verdinelli (1995)].
It is the purpose of the present paper to design an efficient experiment for the estimation of the parameter vector γ in the model (1.1) with efficiency function (1.2) that is robust against misspecifications of the unknown value of ϑ, because such misspecifications can result in an inefficient parameter estimation and therefore severe inaccuracies in the subsequent data analysis. We define robust optimality criteria in the Bayesian and maximin sense based on the D-optimality criterion more formally in the next section. Our methology enables us to find such designs supported on a minimal number of support points. In most cases the designs can even be derived analytically, sometimes they must be determined numerically though. The main results of this study are given in section 3 while section 4 contains some of the corresponding proofs.

2 Standardized maximin and Bayesian optimal designs

Throughout this paper we consider approximate designs ξ, i.e. a design ξ is treated as a probability measure with a finite number of support points in the design space χ, where the relative number of measurements in each design point is approximately equal to the corresponding mass. Given a design ξ and a fixed value of ϑ, the i, j th component of the information matrix of ξ in the heteroscedastic polynomial regression model for the parameter γ can be specified by

$$M(\xi, \vartheta)_{i,j} = \int_\chi \lambda(x, \vartheta) x^{i+j} d\xi(x), \quad i, j = 0, \ldots, n,$$

[see Atkinson and Cook (1995)]. This matrix describes a measure for the information contained in the design ξ to estimate the parameter vector γ in the model (1.1). The widely used D-optimality criterion, which maximizes the determinant of the information matrix, yields a design depending on the unknown parameter ϑ, given below in Lemma 2.1. This means that the D-optimal design is local in the sense of Chernoff (1953) and particularly not robust against misspecifications of the nuisance parameters.

There are two (in some sense related) non-sequential approaches to construct robust optimality criteria, the Bayesian and the maximin procedures, which will be discussed in the following. On the one hand, it can be reasonable to assume that some prior knowledge about the parameter ϑ is available in advance, which can be specified by a probability distribution. In such cases it makes sense to choose a design that maximizes a function of the determinant of $M(\xi, \vartheta)$ after averaging out the plausible values of ϑ by a prior distribution π. Since the use of standardized criteria is recommended to avoid different scaling [see Dette (1997)] this leads to a Bayesian design criterion based on the D-efficiencies

$$\text{eff}_D(\xi, \vartheta) = \left(\frac{|M(\xi, \vartheta)|}{|M(\xi_\vartheta, \vartheta)|} \right)^{1/n+1},$$

where ξ_ϑ denotes the locally D-optimal design for the parameter ϑ. The criterion function $\Psi_p(\xi)$ is then defined as

$$\Psi_p(\xi) = \left[\int_\Theta \left(\frac{|M(\xi, \vartheta)|}{|M(\xi_\vartheta, \vartheta)|} \right)^{p/n+1} d\pi(\vartheta) \right]^{\frac{1}{p}}, \quad -\infty < p < 0, \tag{2.1}$$

and a Bayesian-Ψ_p-optimal design ξ with respect to the prior π maximizes this expression over the set of all approximate designs on the design space χ [see Dette and Wong (1996)].

If – in contrast – the aim is to protect the design of the experiment against the worst possible case it is reasonable to maximize the minimal D-efficiency, which yields the standardized maximin D-optimality criterion

$$\Psi_{-\infty}(\xi) = \inf_{\vartheta \in \Theta} \left[\left(\frac{|M(\xi, \vartheta)|}{|M(\xi_\vartheta, \vartheta)|} \right)^{1/(n+1)} \right] \tag{2.2}$$

[see Müller (1995) or Imhof (2001)]. Throughout this paper we call a design maximizing the above function $\Psi_{-\infty}$-optimal (with respect to Θ), where Θ denotes the region in which ϑ can vary. Note that the standardized maximin criterion is obtained in the limit as $p \to -\infty$ from the Bayesian criterion so the notation $\Psi_{-\infty}$ is consistent with the above definition. Moreover, the classical Bayesian D-optimality criterion [see e.g. Chaloner and Larntz (1989)]

$$\Psi_0(\xi) = \exp \int_\Theta \log |M(\xi, \vartheta)| \, d\pi(\vartheta)$$

is received in the limit as p tends to zero.

In order to simplify our statements of the main results it is helpful to identify each prior distribution π with an associated prior $\tilde{\pi}$ defined by

$$d\tilde{\pi} = |M(\xi_\vartheta, \vartheta)|^{-q} \, d\pi,$$

where we replaced $p/(n+1)$ by q. Hence an equivalent formulation for (2.1) is to find a design ξ maximizing the function

$$\tilde{\Psi}_q(\xi) = \left[\int_{\Theta} |M(\xi, \vartheta)|^q \, d\tilde{\pi}(\vartheta) \right]^{\frac{1}{q}}, \quad -\infty \le q < 0, \tag{2.3}$$

where $\tilde{\Psi}_0(\xi) = \Psi_0(\xi)$. To compute $d\tilde{\pi}$ for a given prior $d\pi$ it is necessary to calculate the value of the determinant $|M(\xi_\vartheta, \vartheta)|$ for the currently considered efficiency function (1.2). The locally D-optimal design ξ_ϑ and the corresponding value of the determinant are given in the following Lemma. Throughout this paper $P_{n+1}^{(\mu,\nu)}(x)$ denote the $(n+1)$th Jacobi polynomial with parameters (μ, ν).

Lemma 2.1.

(i) The locally D-optimal design $\xi_{\alpha,\beta}$ for the weighted polynomial regression model of degree n with design space $\chi = I\!R$, efficiency function

$$(1+x^2)^{\alpha+1} \exp(2\beta \arctan x)$$

and $\alpha < -n-1$ is the uniform distribution on the $(n+1)$ zeros of the Jacobi polynomial

$$P_{n+1}^{(\alpha+i\beta,\, \alpha-i\beta)}(ix).$$

(ii) The value of the determinant of the corresponding information matrix is given by

$$|M(\xi_{\alpha,\beta}, \alpha, \beta)| = 2^{(n+1)(2\alpha+n+2)} \prod_{j=1}^{n} j^j \prod_{j=1}^{n+1} \frac{((\alpha+j)^2+\beta^2)^{\alpha+j} \exp(2\beta \arctan(\frac{-\beta}{j+\alpha}))}{(-2\alpha-(n+j+1))^{2\alpha+n+j+1}}.$$

The proof of part (i) is given in Antille, Dette and Weinberg (2002), whereas the proof of part (ii) is deferred to the appendix.
Some Jacobi-polynomials of lower order degree are given in the following table for the sake of completeness.

n	$P_{n+1}^{(\alpha+i\beta,\alpha-i\beta)}(ix)$
0	$i\,[\beta + (\alpha+1)x]$
1	$\frac{1}{4}\,[-2-2\alpha-2\beta^2-2(3+2\alpha)\beta x-(2+\alpha)(3+2\alpha)x^2]$
2	$\frac{-i}{12}\,[\beta(3\alpha+2(4+\beta^2))+3(2+\alpha)(3+\alpha+2\beta^2)x+3(2+\alpha)(5+2\alpha)\beta x^2$ $+(2+\alpha)(3+\alpha)(5+2\alpha)x^3]$

Table 2.1. The Jacobi polynomials in Lemma 2.1.

3 Main results

We are now ready to state our main results. The following Theorem gives the Bayesian-Ψ_p-optimal $(n+1)$-point design in the settings of (1.1) and (1.2) for all $-\infty < p \leq 0$.

Theorem 3.1. *Let $\chi = \mathbb{R}$, $\Theta \subset (-\infty, -n-1) \times \mathbb{R}$ and the efficiency function $\lambda(x, \alpha, \beta) = (1+x^2)^{\alpha+1} \exp(2\beta \arctan x)$. Let furthermore π be a given prior on Θ with finite first moment, $\tilde{\pi}$ its associated prior and $q = p/(n+1)$, $-\infty < q \leq 0$, with the assumption that*

$$\int_\Theta e^{(\alpha+1)qy+2\beta qz} d\tilde{\pi}(\alpha,\beta) < \infty \quad \forall\, y, z \in \mathbb{R}.$$

Define the functions

$$\begin{aligned} F_1(qy, qz) &= \frac{\int_\Theta (\alpha+1)\, e^{(\alpha+1)qy+2\beta qz} d\tilde{\pi}(\alpha,\beta)}{\int_\Theta e^{(\alpha+1)qy+2\beta qz} d\tilde{\pi}(\alpha,\beta)} \\ F_2(qy, qz) &= \frac{\int_\Theta \beta\, e^{(\alpha+1)qy+2\beta qz} d\tilde{\pi}(\alpha,\beta)}{\int_\Theta e^{(\alpha+1)qy+2\beta qz} d\tilde{\pi}(\alpha,\beta)}. \end{aligned}$$

Then the Bayesian-Ψ_p-optimal $(n+1)$-point design with respect to the prior π has equal mass at the zeros of the Jacobi polynomial

$$P_{n+1}^{(F_1(qy,qz)-1+i\,F_2(qy,qz),F_1(qy,qz)-1-i\,F_2(qy,qz))}(ix),$$

where (y, z) is a solution of the simultaneous equations

$$\begin{aligned} y &= 2(n+1)\log 2 + \sum_{j=1}^{n+1} \Big\{ \log\Big((F_1(qy,qz)-1+j)^2 + F_2^2(qy,qz) \Big) \\ &\qquad -2\log\Big(2F_1(qy,qz)-1+n+j\Big) \Big\} \\ z &= \sum_{j=1}^{n+1} \arctan\Big(\frac{-F_2(qy,qz)}{F_1(qy,qz)-1+j} \Big). \end{aligned} \tag{3.1}$$

Example 3.2.

(i) The case $p = 0$ corresponding to the classical Bayesian-D-optimality criterion is of particular interest. In this case we have $F_1(qy, qz) = E_\pi[\alpha] + 1$, $F_2(qy, qz) = E_\pi[\beta]$ and the Bayesian D-optimal design is given by the uniform distribution on the roots of the polynomial

$$P_{n+1}^{(E_\pi[\alpha]+iE_\pi[\beta],E_\pi[\alpha]+iE_\pi[\beta])}(ix).$$

(ii) If one of the parameters can be assumed as known the representation for the solution in Theorem 3.1 simplifies substantially. For example, if $\tilde{\pi} = \tilde{\pi}_1 \otimes \tilde{\pi}_2$ is a product measure where $\tilde{\pi}_2$ is a Dirac measure at the point β_0, then $F_2(qy, qz) = \beta_0$,

$$F_1(qy) = F_1(qy, qz) = \frac{\int_\Theta (\alpha+1) e^{(\alpha+1)qy} d\tilde{\pi}_1(\alpha)}{\int_\Theta e^{(\alpha+1)qy} d\tilde{\pi}_1(\alpha)}$$

does not depend on z, and y is the solution of the equation

$$y = 2(n+1)\log 2 + \sum_{j=1}^{n+1} \Big\{ \log((F_1(qy) - 1 + j)^2 + \beta_0^2) - 2\log(2F_1(qy) - 1 + n + j) \Big\}.$$

The Ψ_p-optimal design is a uniform distribution at the roots of the polynomial

$$P_{n+1}^{(F_1(qy)-1+i\beta_0, F_1(qy)-1-i\beta_0)}(ix).$$

In particular, if $\beta_0 = 0$ we obtain robust optimal designs for the heteroscedastic polynomial regression model considered in Dette, Haines and Imhof (1999).

The corresponding maximin result is given in the Theorem below.

Theorem 3.3. *Assume that $\chi = I\!R$, Θ a compact and convex subset of $(-\infty, -n-1) \times I\!R$ with boundary $\Gamma = \partial\Theta$ and consider the efficiency function $\lambda(x, \alpha, \beta) = (1+x^2)^{\alpha+1} \exp(2\beta \arctan x)$. Then the $\Psi_{-\infty}$-optimal $(n+1)$-point design with respect to Θ is unique and given by the uniform distribution on the components of the vector $(x_1^*, \ldots, x_{n+1}^*)^T \in I\!R^{n+1}$, where $x_1^* < \ldots < x_{n+1}^*$ maximize the function*

$$h(x_1, \ldots, x_n) = \min_{(\alpha,\beta)\in\Gamma} \Bigg[\Bigg\{ 2^{-(n+1)(2\alpha+n+2)} \prod_{j=1}^{n+1} \frac{(1+x_j^2)^{\alpha+1}(-2\alpha-(n+j+1))^{2\alpha+n+j+1}}{j^j\ ((\alpha+j)^2+\beta^2)^{\alpha+j}}$$

$$\times \prod_{1\le j<k\le n+1} (x_k - x_j)^2 \prod_{j=1}^{n+1} \exp\Big(2\beta\Big(\arctan(x_j) - \arctan\Big(\frac{-\beta}{\alpha+j}\Big)\Big)\Big) \Bigg\}^{1/(n+1)} \Bigg].$$

There are two special cases of the settings in Theorem 3.3 with more explicit results. If one of the variables at a time can be assumed to be known we get the following.

Theorem 3.4. *Assume that the assumptions of Theorem 3.3. are satisfied.*

(i) If the value of α is known and β lies in some interval $[\beta_1, \beta_2] \subset I\!R$, $\beta_1 < \beta_2$, then the $\Psi_{-\infty}$-optimal $(n+1)$-point design with respect to Θ is given by the uniform distribution on the $(n+1)$ zeros of the Jacobi polynomial

$$P_{n+1}^{(\alpha+i\beta_0, \alpha-i\beta_0)}(i\,x),$$

where β_0 is the unique solution of the equation

$$\prod_{j=1}^{n+1} e^{\arctan(\frac{-\beta_0}{\alpha+j})} = \Big\{ \prod_{j=1}^{n+1} \Big(\frac{(\alpha+j)^2+\beta_2^2}{(\alpha+j)^2+\beta_1^2}\Big)^{\alpha+j} \ \frac{e^{2\beta_2 \arctan(\frac{-\beta_2}{\alpha+j})}}{e^{2\beta_1 \arctan(\frac{-\beta_1}{\alpha+j})}} \Big\}^{1/2(\beta_2-\beta_1)}.$$

(ii) If the value of the parameter β is known and α lies in some interval $[\alpha_1, \alpha_2]$, $\alpha_1 < \alpha_2 < -n-1$, then the $\Psi_{-\infty}$-optimal $(n+1)$-point design with respect to Θ is given by the uniform distribution on the $(n+1)$ zeros of the Jacobi polynomial

$$P_{n+1}^{(\alpha_0+i\,\beta,\alpha_0-i\,\beta)}(i\,x),$$

where α_0 is the unique solution of the equation

$$\prod_{j=1}^{n+1} \frac{(2\alpha_0+2j)^2+4\beta^2}{(2\alpha_0+n+j+1)^2} = d_{n+1}(\alpha_1,\alpha_2),$$

and the constant $d_{n+1}(\alpha_1,\alpha_2)$ is given by the expression

$$\Big\{ \prod_{j=1}^{n+1} \frac{((2\alpha_1+2j)^2+4\beta^2)^{-\alpha_1-j} \ (-2\alpha_2-(n+j+1))^{-2\alpha_2-(n+j+1)} \ e^{2\beta \arctan(\frac{-\beta}{\alpha_2+j})}}{((2\alpha_2+2j)^2+4\beta^2)^{-\alpha_2-j} \ (-2\alpha_1-(n+j+1))^{-2\alpha_1-(n+j+1)} \ e^{2\beta \arctan(\frac{-\beta}{\alpha_1+j})}} \Big\}^{\frac{1}{\alpha_2-\alpha_1}}.$$

Example 3.5. If $\beta = 0$ the situation in Theorem 3.4 (ii) simplifies further. In this case the standardized maximin D-optimal design is a (symmetric) uniform distribution at the roots of the polynomial $P_{n+1}^{(\alpha_0,\alpha_0)}(ix)$, where α_0 is the unique solution of the equation

$$\prod_{j=1}^{n+1} \Big(\frac{\alpha_0+j}{2\alpha_0+n+j+1}\Big)^{2(\alpha_2-\alpha_1)} = \prod_{j=1}^{n+1} \frac{(\alpha_2+j)^{\alpha_2+j}(-2\alpha_1-(n+j+1))^{2\alpha_1+n+j+1}}{(\alpha_1+j)^{\alpha_1+j}(-2\alpha_2-(n+j+1))^{2\alpha_2+n+j+1}}.$$

Note that this result provides robust designs (based on the maximin approach) in the situation discussed by Dette, Haines and Imhof (1999).

4 Appendix: proofs

Proof of part (ii) of Lemma 2.1. The determinant of the information matrix of the locally D-optimal design $\xi_{\alpha,\beta}$ is given by

$$|M(\xi_{\alpha,\beta},\alpha,\beta)| = \Big(\frac{1}{n+1}\Big)^{n+1} \prod_{j=1}^{n+1} (1+x_{j,\alpha,\beta}^2)^{\alpha+1} \prod_{j=1}^{n+1} e^{2\beta \arctan(x_{j,\alpha,\beta})} \prod_{1\le j<k\le n+1} (x_{k,\alpha,\beta}-x_{j,\alpha,\beta})^2,$$

where $x_{j,\alpha,\beta}$, $j = 1, \ldots, n+1$, denote the support points of $\xi_{\alpha,\beta}$. Note that the supporting polynomial can be written as

$$P_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}(ix) = \lambda_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)} \prod_{j=1}^{n+1} (x - x_{j,\alpha,\beta})$$

where $\lambda_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}$ denotes the leading coefficient of the Jacobi polynomial. This gives the representation

$$\prod_{j=1}^{n+1} (1 + x_{j,\alpha,\beta}^2) = \prod_{j=1}^{n+1} (i - x_{j,\alpha,\beta}) \prod_{j=1}^{n+1} (-i - x_{j,\alpha,\beta}) = \frac{P_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}(-1)\; P_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}(1)}{(\lambda_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)})^2}$$

$$= 2^{2(n+1)} \prod_{j=1}^{n+1} \frac{(\alpha+j)^2 + \beta^2}{(2\alpha + n + j + 1)^2}, \tag{4.1}$$

where we applied formulae (4.1.1), (4.1.4) and (4.21.6) from Szegö (1975). Using the above formulae again as well as the representation

$$\arctan z = -\frac{i}{2} \log\Big(\frac{1+iz}{1-iz}\Big),$$

we obtain

$$\prod_{j=1}^{n+1} \exp(\arctan(x_{j,\alpha,\beta})) = \prod_{j=1}^{n+1} \exp\Big(\arctan\Big(\frac{-\beta}{j+\alpha}\Big)\Big). \tag{4.2}$$

To calculate the value of the remaining factor in the determinant we consider the corresponding Jacobi polynomials in the variable x instead of $i\,x$, i.e.

$$P_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}(x) = P_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}(i\,(-ix)) = \frac{\lambda_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}}{i^{n+1}} \prod_{j=1}^{n+1} (x - \tilde{x}_{j,\alpha,\beta}),$$

where the $\tilde{x}_{j,\alpha,\beta} = ix_{j,\alpha,\beta}$, $j = 1, \ldots, n+1$, denote the zeros of these functions. The discriminant formula (6.71.5) in Szegö (1975), Theorem 6.71, gives

$$D_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)} = 2^{-n(n+1)} \prod_{j=1}^{n+1} j^{j-2n} (j+\alpha+i\beta)^{j-1} (j+\alpha-i\beta)^{j-1} (n+j+1+2\alpha)^{n+1-j}$$

with the discriminant $D_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}$ defined as

$$D_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)} = \Big(\frac{\lambda_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)}}{i^{n+1}}\Big)^{2n} \prod_{1\le j<k\le n+1} (\tilde{x}_{k,\alpha,\beta} - \tilde{x}_{j,\alpha,\beta})^2$$

$$= (\lambda_{n+1}^{(\alpha+i\beta,\,\alpha-i\beta)})^{2n} (-1)^{n(n+1)/2} \prod_{1\le j<k\le n+1} (x_{k,\alpha,\beta} - x_{j,\alpha,\beta})^2.$$

Therefore we end up with the identity

$$(4.3)\qquad \prod_{1\le j<k\le n+1}(x_{k,\alpha,\beta}-x_{j,\alpha,\beta})^2 = 2^{n(n+1)}\prod_{j=1}^{n+1}\frac{j^j(-j-\alpha-i\beta)^{j-1}(-j-\alpha+i\beta)^{j-1}}{(-2\alpha-(n+j+1))^{n+j-1}}.$$

Summarizing the results in (4.1), (4.2), (4.3) and adding the exponents we obtain

$$|M^n(\xi_{\alpha,\beta},\alpha,\beta)| = 2^{(n+1)(2\alpha+n+2)}\prod_{j=1}^{n} j^j \prod_{j=1}^{n+1}\frac{((\alpha+j)^2+\beta^2)^{\alpha+j}\exp(2\beta\arctan(\frac{-\beta}{j+\alpha}))}{(-2\alpha-(n+j+1))^{2\alpha+n+j+1}},$$

which ends the proof of Lemma 2.1 □

Proof of Theorem 3.1. By a standard argument in design theory [see Silvey (1980)] it can be shown that the Bayesian-Ψ_p-optimal $(n+1)$-point design with respect to the prior π is a uniform distribution. Thus, by an application of the Vandermonde determinant formula the induced optimality criterion $\tilde{\Psi}_q$ in (2.3) is proportional to

$$(4.4)\qquad \tilde{\Psi}_q(\xi) \propto \prod_{1\le j<k\le n+1}(x_k-x_j)^2\left[\int_\Theta\Big(\prod_{j=1}^{n+1}(1+x_j^2)^{\alpha+1}\,e^{2\beta\arctan x_j}\Big)^q d\tilde{\pi}(\alpha,\beta)\right]^{\frac{1}{q}},$$

where the points $x_j,\ j=1,\dots,n+1$, denote the support points of the design ξ and the integral exists by the assumption in the Theorem. If $x_j\to-\infty$, we have $\tilde{\Psi}_q(\xi)\to 0,\ j=1,\dots,n+1$, and the same holds for the other direction $x_j\to\infty$. Hence the criterion function is maximized by some interior point of $I\!R^{n+1}$. Since additionally $\tilde{\Psi}_q(\xi)$ is differentiable with respect to the x_j's, a necessary condition for a maximum in a certain point $x^*\in I\!R^{n+1}$ is a vanishing gradient of the function $\tilde{\Psi}_q(\xi)$ or $\log\tilde{\Psi}_q(\xi)$ in x^*, respectively. Differentiating $\log\tilde{\Psi}_q(\xi)$ with respect to $x_l,\ l=1,\dots,n+1$, yields

$$(4.5)\qquad \frac{\partial\log\tilde{\Psi}_q(\xi)}{\partial x_l} = 2\sum_{j=1,j\ne l}^{n+1}\frac{1}{x_l-x_j}+\frac{2x_l}{1+x_l^2}F_1(qy,qz)+\frac{2}{1+x_l^2}F_2(qy,qz),$$

where the functions $F_1,\ F_2$ are defined in Theorem 3.1 and y, z are given by the expressions

$$(4.6)\qquad y=\sum_{j=1}^{n+1}\log(1+x_j^2),\qquad z=\sum_{j=1}^{n+1}\arctan(x_j).$$

Equating the right hand side of (4.5) to zero for $l=1,\dots,n+1$, gives a system of equations, which can be transformed into a differential equation for the polynomial $f(x)=\prod_{j=1}^{n+1}(x-x_j^*)$ [see e.g. Szegö (1975), p. 141], i.e.

$$(4.7)\quad (1+x^2)f''(x)+(2xF_1(qy,qz)+2F_2(qy,qz))f'(x)-(n+1)(n+2F_1(qy,qz))\equiv 0.$$

¿From Szegö (1975) it follows that the polynomial solution of the differential equation (4.7) is the Jacobi polynomial $P_{n+1}^{(\mu,\nu)}(ix)$ with parameters

$$\mu = F_1(qy, qz) - 1 + i\, F_2(qy, qz), \quad \nu = F_1(qy, qz) - 1 - i\, F_2(qy, qz).$$

The conditions on y and z are finally obtained by using the results of (4.1) and (4.2) in the defining equations (4.6), respectively. □

Proof of Theorem 3.3. By a standard argument [see Silvey (1980)] it can be shown that a $\Psi_{-\infty}$-optimal $(n+1)$-point design has equal weights on its support points. Hence the criterion function is proportional to

$$\inf_{(\alpha,\beta)\in\Theta} \Bigg[\Bigg\{ 2^{-(n+1)(2\alpha+n+2)} \prod_{j=1}^{n+1} \frac{(1+x_j^2)^{\alpha+1}(-2\alpha-(n+j+1))^{2\alpha+n+j+1}}{j^j\ ((\alpha+j)^2+\beta^2)^{\alpha+j}}$$

(4.8)

$$\times \prod_{1\le j<k\le n+1} (x_k - x_j)^2 \prod_{j=1}^{n+1} \exp\Big(2\beta \Big(\arctan(x_j) - \arctan\Big(\frac{-\beta}{\alpha+j}\Big)\Big)\Big) \Bigg\}^{1/(n+1)} \Bigg].$$

Since the parameter space Θ is assumed to be compact, the notation "infimum" can be replaced by "minimum". So it remains to prove that the minimum of the D-efficiencies lies on the boundary of Θ and, additionally, that the solution of the optimal design problems is unique. To show that the minimum lies on Γ (the boundary of Θ) for every $(n+1)$-point design ξ we demonstrate that the function over which the minimum is taken is log-concave in (α, β) by proving the negative (semi)definiteness of its Hessian with respect to these parameters. The entries of the Hessian are given by

(4.9)

$$X := \frac{\partial^2 \log \Phi}{\partial \alpha^2} = \sum_{j=1}^{n+1} \Big[\frac{2(\alpha+j)(n+1-j) - 4\beta^2}{(-2\alpha-(n+j+1))((\alpha+j)^2+\beta^2)} \Big]$$

$$Y := \frac{\partial^2 \log \Phi}{\partial \alpha\, \partial \beta} = \sum_{j=1}^{n+1} \frac{-2\beta}{(\alpha+j)^2+\beta^2}$$

$$Z := \frac{\partial^2 \log \Phi}{\partial \beta^2} = \sum_{j=1}^{n+1} \frac{2(\alpha+j)}{(\alpha+j)^2+\beta^2},$$

which yields for the corresponding eigenvalues λ_1, λ_2

$$\lambda_{1,2} = \frac{X+Z}{2} \mp \sqrt{\frac{(X+Z)^2}{4} - XZ + Y^2}.$$

¿From $\alpha + j < 0$ we conclude $X < 0$ and $Z < 0$. Hence it can immediately be seen that the smaller eigenvalue is negative. To show that the larger eigenvalue is non-positive we consider the product $\lambda_1\lambda_2$ and prove its non-negativity.

$$
\begin{aligned}
\lambda_1\,\lambda_2 \;&=\; \sum_{j=1}^{n+1}\sum_{k=1}^{n+1} \frac{4(\alpha+j)(\alpha+k)(n+1-j)}{(-2\alpha-(n+j+1))((\alpha+j)^2+\beta^2)((\alpha+k)^2+\beta^2)} \\
&+\; 4\beta^2 \sum_{j=1}^{n+1}\sum_{k=1}^{n+1} \frac{n+1+j-2k}{(-2\alpha-(n+j+1))((\alpha+j)^2+\beta^2)((\alpha+k)^2+\beta^2)}.
\end{aligned}
\tag{4.10}
$$

The first term on the right hand side of (4.10) is obviously non-negative, for $n > 0$ even strictly positive. To show that the same claim holds for the second one we split it up into

$$
\sum_{j=1}^{n+1}\sum_{k=1}^{n+1} r_{n+1}^{\alpha,\beta}(j,k) = \sum_{j=1}^{n+1} r_{n+1}^{\alpha,\beta}(j,j) + \sum_{j=1}^{n+1}\sum_{k=1,\,k<j}^{n+1} [r_{n+1}^{\alpha,\beta}(j,k) + r_{n+1}^{\alpha,\beta}(k,j)],
$$

where $r_{n+1}^{\alpha,\beta}(j,k),\ j,\ k = 1,\ldots,n+1$, denotes the expression

$$
r_{n+1}^{\alpha,\beta}(j,k) := \frac{n+1+j-2k}{(-2\alpha-(n+j+1))((\alpha+j)^2+\beta^2)((\alpha+k)^2+\beta^2)}.
$$

A straightforward but tedious calculation yields that both $r_{n+1}^{\alpha,\beta}(j,j)$ and $r_{n+1}^{\alpha,\beta}(j,k) + r_{n+1}^{\alpha,\beta}(k,j)$ are non-negative, which implies the log-concavity of the function under consideration.
The uniqueness of the solution in the class of $(n+1)$-point designs with uniform mass can be shown as follows. Define the function of the D-efficiencies as $\Phi(\xi,\vartheta) = \mathrm{eff}_D(\xi,\vartheta)$, a function of the support points of ξ and the unknown parameter vector ϑ. Obviously, the relation between $\Phi(\xi,\vartheta)$ and the maximin optimality criterion is given by

$$
\Psi_{-\infty}(\xi) = \min_{\vartheta\in\Theta} \Phi(\xi,\vartheta).
$$

Let $\xi^{(1)}$, $\xi^{(2)}$ denote two different $(n{+}1)$-point uniform designs with support points $x_j^{(1)}$, $x_j^{(2)}$, $j = 1,\ldots,n+1$, respectively, and construct another design $\xi^{(1,2)}$ in this class as a uniform distribution on the support points

$$
x_j^{(1,2)} = \frac{x_j^{(1)} + x_j^{(2)}}{2}, \qquad j = 1,\ldots,n+1.
$$

For fixed $\vartheta \in \Theta$ the function $\Phi(\xi,\vartheta)$ is strictly unimodal on the set

$$
\Big\{ (x_1,\ldots,x_{n+1})^T \subset I\!R^{n+1} \;\Big|\; -\infty < x_1 < \ldots < x_{n+1} < \infty \Big\}
$$

[see Antille, Dette and Weinberg (2002)], which implies that the following holds

$$
\Phi(\xi^{(1,2)},\vartheta) > \min\{\Phi(\xi^{(1)},\vartheta), \Phi(\xi^{(2)},\vartheta)\} \qquad \forall\,\vartheta\in\Theta. \tag{4.11}
$$

Now let $\xi^{(1)}$, $\xi^{(2)}$ be different$\Psi_{-\infty}$-optimal in the class of $(n+1)$-point designs with optimal value $\Psi^*_{-\infty}$, i.e.

$$\min_{\vartheta\in\Theta} \Phi(\xi^{(1)}, \vartheta) = \min_{\vartheta\in\Theta} \Phi(\xi^{(2)}, \vartheta) = \Psi^*_{-\infty}.$$

In particular, we have

$$\Phi(\xi^{(1)}, \vartheta) \geq \Psi^*_{-\infty}, \;\; \Phi(\xi^{(2)}, \vartheta) \geq \Psi^*_{-\infty} \qquad \forall\, \vartheta \in \Theta. \tag{4.12}$$

The above minima exist because the parameter space Θ is assumed to be compact. We define $\mathcal{N}(\xi)$ as the set of parameter values, for which the function $\Phi(\xi, \vartheta)$ becomes minimal, i.e.

$$\mathcal{N}(\xi) = \Big\{ \vartheta \in \Theta \;\Big|\; \Phi(\xi, \vartheta) = \min_{\vartheta\in\Theta} \Phi(\xi, \vartheta) \Big\}.$$

The compactness of Θ implies that the set $\mathcal{N}(\xi^{(1,2)})$ is not empty. ¿From the optimality of the value $\Psi^*_{-\infty}$ we conclude that for all $\vartheta^* \in \mathcal{N}(\xi^{(1,2)})$ the inequality

$$\Phi(\xi^{(1,2)}, \vartheta^*) \leq \Psi^*_{-\infty} \tag{4.13}$$

holds. But from (4.11) and (4.12) it follows that

$$\Phi(\xi^{(1,2)}, \vartheta^*) > \min\{\Phi(\xi^{(1)}, \vartheta^*), \Phi(\xi^{(2)}, \vartheta^*)\} \geq \Psi^*_{-\infty}, \tag{4.14}$$

which is a contradiction. Thus there can not be more than one $\Psi_{-\infty}$-optimal $(n+1)$-point design and the proof of Theorem 3.2 is complete. □

Proof of part (i) of Theorem 3.4. ¿From the last equation in (4.9) it follows that for fixed α the function $\log \Phi(\xi, \vartheta)$ defined in the proof of Theorem 3.2. is strictly concave in β. Hence the minimum of $\Phi(\xi, \vartheta)$ occurs on the boundary of $\Theta = \{\alpha\} \times [\beta_1, \beta_2]$, i.e.

$$\min_{\vartheta\in\Theta} \Phi(\xi, \vartheta) = \min\{\Phi(\xi, \alpha, \beta_1), \Phi(\xi, \alpha, \beta_2)\}.$$

In the next step we show that for the $\Psi_{-\infty}$-optimal $(n+1)$-point design ξ^* the following holds

$$\Phi(\xi^*, \alpha, \beta_1) = \Phi(\xi^*, \alpha, \beta_2). \tag{4.15}$$

Having proven the above equation we use it as a side condition on the optimal design ξ^*, which enables us to apply a constrained optimization technique to our problem. We then apply the differential equation approach already pointed out in the proof of Theorem 3.1 and end up with the result of Theorem 3.4. It is straightforward to show the uniqueness of the solution and those calculations are therefore omitted.
Thus it remains to prove equation (4.15). We only provide a sketch of the proof since the calculations are straightforward but tedious. For more details see Biedermann (2003). Let the sets $\mathcal{M}_{\beta_1}$, $\mathcal{M}_{\beta_2}$ be defined by

$$\mathcal{M}_{\beta_1} = \{\xi \mid \Phi(\xi, \alpha, \beta_1) < \Phi(\xi, \alpha, \beta_2)\}, \quad \mathcal{M}_{\beta_2} = \{\xi \mid \Phi(\xi, \alpha, \beta_1) < \Phi(\xi, \alpha, \beta_2)\},$$

i.e. $\mathcal{M}_{\beta_i}$ denotes the set of uniformly weighted $(n+1)$-points designs, for which the minimum of $\Phi(\xi, \alpha, \beta)$ is attained at the point β_i, $(i = 1, 2)$. The inequality in the definition of the set $\mathcal{M}_{\beta_1}$ is equivalent to

$$\prod_{j=1}^{n+1} e^{\arctan(x_j)} > \left\{ \prod_{j=1}^{n+1} \Big(\frac{(\alpha+j)^2 + \beta_2^2}{(\alpha+j)^2 + \beta_1^2} \Big)^{\alpha+j} \; \frac{e^{2\beta_2 \arctan(\frac{-\beta_2}{\alpha+j})}}{e^{2\beta_1 \arctan(\frac{-\beta_1}{\alpha+j})}} \right\}^{1/2(\beta_2-\beta_1)} . \tag{4.16}$$

If the optimal $(n+1)$-point design ξ^* is in the interior of $\mathcal{M}_{\beta_1}$ it must be obtained as the zero of the gradient of $\Phi(\xi, \alpha, \beta_1)$ and hence must coincide with the locally D-optimal design ξ_{α,β_1}. Plugging this into inequality (4.16) and using (4.2) yields the inequality

$$\prod_{j=1}^{n+1} e^{\arctan(\frac{-\beta_1}{\alpha+j})} > \left\{ \prod_{j=1}^{n+1} \Big(\frac{(\alpha+j)^2 + \beta_2^2}{(\alpha+j)^2 + \beta_1^2} \Big)^{\alpha+j} \; \frac{e^{2\beta_2 \arctan(\frac{-\beta_2}{\alpha+j})}}{e^{2\beta_1 \arctan(\frac{-\beta_1}{\alpha+j})}} \right\}^{1/2(\beta_2-\beta_1)} .$$

It can easily be shown that this inequality is not valid and therefore ξ^* is not an element of the interior of $\mathcal{M}_{\beta_1}$. Analogous reasoning leads to the same result for $\mathcal{M}_{\beta_2}$, hence ξ^* lies on the common boundary of $\mathcal{M}_{\beta_1}$ and $\mathcal{M}_{\beta_2}$ and the equation (4.15) holds. □

Proof of part (ii) of Theorem 3.4. The proof is is omitted since we can proceed analogously to the proof of part (i).

Acknowledgements: This paper contains results from the first author's PhD thesis written at the Department of Mathematics, Ruhr-Universität Bochum, Germany. The support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt A2) is gratefully acknowledged.

References

[1] **Antille, G., Dette, H., and Weinberg, A.** (2002). *A note on optimal designs in weighted polynomial regression for the classical efficiency functions.* To appear in: Journal of Statistical Planning and Inference.

[2] **Atkinson, A.C. und Cook, R.D.** (1995). *D-optimum designs for heteroscedastic linear models.* J. Am. Statist. Assoc., **90**, 204-212.

[3] **Biedermann, S. (2003).** *Klassische orthogonale Polynome und ihre Anwendung in der optimalen Versuchsplanung* (in German). Dissertation, Fakultät für Mathematik, Ruhr-Universität Bochum, Germany.

[4] **Chaloner, K., Verdinelli, I. (1995).** *Bayesian experimental design: A review.* Statistical Science, **10**, 273-304.

[5] **Chaloner, K. and Larntz, K.** (1989). *Optimal Bayesian design applied to logistic regression experiments.* J. Stat. Plann. Inference, **21**, No. 2, 191-208.

[6] **Chernoff, H.** (1953). *Locally optimal designs for estimating parameters.* Ann. Math. Statist., **24**, 586-602.

[7] **Dette, H., Haines, L., Imhof, L. (1999).** *Optimal designs for rational models and weighted polynomial regression.* Annals of Statistics, **27**, 1272-1293.

[8] **Dette, H.** (1997). *Designing experiments with respect to standardized optimality criteria.* J. Roy. Statist. Soc., Ser. B, **59**, No. 1, 97-110.

[9] **Dette, H. and Wong, W.K.** (1996). *Optimal Bayesian designs for models with partially specified heteroscedastic structure.* Annals of Statistics, **24**, No. 5, 2108-2127.

[10] **Fedorov, V.V.** (1972). *Theory of optimal experiments* (translated by W.J. Studden and E.M. Klimko). Academic Press, New York.

[11] **Imhof, L.** (2001). *Maximin designs for exponential growth models and heteroscedastic polynomial models.* Annals of Statistics, **29**, No. 2, 561-576.

[12] **Müller, C. H. (1995).** *Maximin efficient designs for estimating nonlinear aspects in linear models.* J. Statist. Plann. Inf. **44**, 117-132.

[13] **Silvey, S.D.** (1980). *Optimal Design.* Chapman and Hall, London.

[14] **Szegö, G.** (1975). *Orthogonal Polynomials.* 4th edition. American Mathematical Society Colloquium Publications, **23**, Providence.