Maximin and Bayesian optimal designs for linear and non-linear regression models

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
holger.dette@ruhr-uni-bochum.de

Linda M. Haines
School of Mathematics,
Statistics and Information Technology,
University of Natal,
Pietermaritzburg, South Africa
haines@stat.unp.ac.za

Lorens A. Imhof
Aachen University
Institut für Statistik
und Wirtschaftsmathematik
52056 Aachen, Germany
imhof@stochastik.rwth-aachen.de

August 6, 2002

Abstract

For many problems of statistical inference in regression modelling, the Fisher information matrix depends on certain nuisance parameters which are unknown and which enter the model nonlinearly. A common strategy to deal with this problem within the context of design is to construct maximin optimal designs as those designs which maximize the minimum value of a real valued (standardized) function of the Fisher information matrix, where the minimum is taken over a specified range of the unknown parameters. The maximin criterion is not differentiable and the construction of the associated optimal designs is therefore difficult to achieve in practice. In the present paper the relationship between maximin optimal designs and a class of Bayesian optimal designs for which the associated criteria are differentiable is explored. In particular, a general methodology for determining maximin optimal designs is introduced based on the fact that in many cases these designs can be obtained as weak limits of appropriate Bayesian optimal designs. The approach is illustrated by means of a broad range of examples for which the Bayesian optimal and hence the maximin optimal designs can be found explicitly.

AMS Subject Classification: Primary 62K05 ; Secondary 62F15

Keywords and Phrases: maximin optimal designs, Bayesian optimal designs, nonlinear regression models, parameter estimation, least favourable prior

1 Introduction

In many practical problems in regression modelling the Fisher information for the parameters of interest depends on certain unknown nuisance parameters. Within the context of design this problem translates into that of maximizing a concave function of the information matrix, which depends on the unknown parameters and clearly this cannot be achieved directly. Over the last forty years a number of strategies have been developed to address this design problem. Specifically, in 1953, Chernoff suggested the simple but elegant expedient of adopting a best guess for the unknown parameters and termed the resultant designs locally optimal. The main disadvantage to such an approach is that if the unknown parameters are misspecified the resulting optimal designs can be highly inefficient within the true model setting. An intuitively appealing extension to this notion is therefore to construct designs sequentially, using a best guess to initiate the process and then updating this guess with parameter estimates obtained from the observed responses after each step of the procedure [see e.g. Ford and Silvey (1980), Wu (1985)]. There are unfortunately a number of disadvantages to this idea. Thus in many experimental situations, particularly in agriculture, observations cannot be taken sequentially. Furthermore, statistical inference based on observations from sequentially designed experiments is not a straightforward matter.
A broad and attractive approach to the problem of design when the Fisher information matrix involves unknown parameters is to, in some sense, quantify the uncertainty in those parameters and to incorporate this additional information into the formulation of suitable optimality criteria. This has been achieved in practice through the introduction of the concepts of Bayesian and of maximin optimality. To be specific, for Bayesian optimality a prior distribution is placed on the unknown parameters and criteria involving the average of a concave function of the Fisher information matrix over that distribution are then maximized. The resultant designs, termed robust or Bayesian optimal designs, have been extensively studied in the literature and their properties are well-understood [see for example Pronzato and Walter (1985), Chaloner and Larntz (1989), Chaloner (1993), Chaloner and Verdinelli (1995)]. In particular Bayesian optimality criteria are based on criteria in classical design theory and many of the results from that theory, such as those relating to equivalence theorems and numerical procedures, can immediately be transferred into the Bayesian context. For maximin optimality, designs which maximize the minimum of a function of the Fisher information matrix over a range of parameter values are sought [see e.g. Pronzato and Walter (1985), Müller (1995), Dette (1997), Müller and Pázman (1998)]. The resultant designs, termed maximin optimal designs, are particularly attractive from a practical point of view in that the experimenter is only required to specify an appropriate range for the unknown parameters. The major problem lies in the construction of these designs in the sense that the maximin optimality criterion is not differentiable and that as a consequence results, both algebraic and numeric, are elusive. Indeed there have been few reports of maximin optimal designs in the literature and strategies for their construction are somewhat ad hoc [see e.g. Wong (1992), Haines (1995), Dette and Sahm (1998), Imhof (2001)].
In the present study a general approach to obtaining maximin optimal designs as the limits of a particular class of Bayesian optimal designs is introduced and explored. Roughly speaking the powerful tools for constructing Bayesian optimal designs for which the associated criteria are differentiable can be used to obtain maximin optimal designs for which the correspond-

ing criteria are not differentiable. Several applications of this methodology are presented and maximin optimal designs obtained explicitly. Although interest is centered primarily on the construction of maximin optimal designs for nonlinear regression models the approach is quite general and can be applied to optimal designs problems with a similar structure such as model robust design problems.

The paper is organized in the following way. In Section 2 some preliminary definitions are given and Bayesian optimality criteria analogous to Kiefer's (1974) Φ_p-criteria are introduced. The main results of the study are then presented in Section 3. In particular the relationship between Bayesian and maximin optimal designs is explored and powerful equivalence theorems and other associated results are presented. Furthermore, it is shown that under fairly general conditions the Bayesian optimal designs converge weakly to maximin optimal designs, a result which mirrors the limiting relationship of the corresponding optimality criteria. A broad range of applications are considered in Sections 4 and 5. Specifically, examples of models for which the Bayesian optimal designs are available from the literature are presented in Section 4 and it is shown that maximin optimal designs can be obtained as the limits of these designs. In Section 5 two weighted polynomial models for which the Bayesian and maximin optimal designs are not known are considered. Bayesian optimal designs for these models are obtained in analytic form using a method based on the theory of differential equations and the results of Section 3 are then invoked in order to find the maximin optimal designs explicitly. Some broad conclusions and pointers for future research are given in Section 6. For ease of reading the proofs of all lemmas and theorems in the paper are included in an appendix.

2 Preliminaries

Consider a regression model which depends, possibly nonlinearly, on the parameters θ from a parameter space $\Theta \subset \mathbb{R}^k$ and on explanatory variables x varying in a compact design space $\mathcal{X} \subset \mathbb{R}^\ell$ equipped with a σ-field, which contains all one point sets. An approximate design ξ for this model is a probability measure on the design space $\mathcal{X}$ with finite support $x_1, \ldots, x_n$ and weights $w_1, \ldots, w_n$ representing the relative proportion of total observations taken at the corresponding design points [see e.g. Kiefer (1974)]. Let Ξ denote the class of all approximate designs and $\Delta \subset \Xi$ some subset of that class. Then, very broadly, an optimality criterion can be specified as

$$\psi : \Delta \times \Theta \to [0, \infty) ,$$

where the function $\psi(\xi, \theta)$ is continuous in the sense that, if a sequence of designs $\xi_n \in \Delta$ converges weakly to a design $\xi \in \Delta$ as $n \to \infty$, then

$$\lim_{n\to\infty} \psi(\xi_n, \theta) = \psi(\xi, \theta)$$

for all $\theta \in \Theta$. Additionally, for fixed $\xi \in \Delta$, the function $\psi(\xi, \theta)$ is assumed to be continuous in θ. Examples of such a criterion include, *inter alia*, D- and c-optimality [Pukelsheim (1993)]. In the present study attention is focussed on optimality criteria which accommodate uncertainty in the unknown parameters and, specifically, on criteria based on functions of the form $\psi(\xi, \theta)$. To this end it is first necessary to consider a single, fixed parameter value $\theta \in \Theta$ and to introduce a locally ψ-optimal design over the class of designs Δ as a design $\xi^*_\theta \in \Delta$ for which the condition

$$\psi(\xi_\theta^*, \theta) \geq \psi(\xi, \theta)$$

holds for all $\xi \in \Delta$. A standardized maximin ψ-optimal design in the class Δ can then be defined as a design which maximizes the criterion

$$\Psi_{-\infty}(\xi) = \inf_{\theta \in \Theta} \frac{\psi(\xi, \theta)}{\psi(\xi_\theta^*, \theta)} \tag{2.1}$$

over all $\xi \in \Delta$ [see Dette (1997)] and a Bayesian ψ-optimal design with respect to a prior distribution π on the parameter space Θ as a design which maximizes

$$\Psi_0(\xi) = \exp \int_\Theta \log \psi(\xi, \theta) d\pi(\theta) \tag{2.2}$$

over the set Δ [see e.g. Pronzato and Walter (1985) or Chaloner and Larntz (1989)]. More generally, for fixed q such that $-\infty < q < 0$ a Bayesian Ψ_q-optimal design for a prior distribution π on Θ can be defined as a design $\xi \in \Delta$ maximizing the criterion

$$\Psi_q(\xi) = \left[\int_\Theta \left\{ \frac{\psi(\xi, \theta)}{\psi(\xi_\theta^*, \theta)} \right\}^q d\pi(\theta) \right]^{\frac{1}{q}} \tag{2.3}$$

over the subclass of designs Δ [see Dette and Wong (1996)]. Note that the Bayesian ψ-optimality criterion (2.2) is obtained from (2.3) in the limit as $q \to 0$ and that the standardized maximin criterion (2.1) is recovered as $q \to -\infty$ provided the support of the prior π coincides with the parameter space Θ, i.e. $\text{supp}(\pi) = \Theta$.

3 Bayesian and standardized maximin optimal designs

3.1 A general Equivalence Theorem

The results of this subsection relate to a particular form of optimality criterion. Specifically, suppose that the Fisher information matrix for the parameter $\theta \in \Theta$ of a design $\xi \in \Delta$ can be expressed as

$$M(\xi, \theta) = \int_\mathcal{X} f(x, \theta) f^T(x, \theta) d\xi(x) \in \mathbb{R}^{\ell_\theta \times \ell_\theta}$$

where $f(x, \theta) \in \mathbb{R}^{\ell_\theta}$ is a vector-valued function appropriate to the specified regression model and the dimension l_θ may depend on θ. Then the criterion of interest has the form

$$\psi(\xi, \theta) = \phi_\theta\{C_\theta(\xi)\} \tag{3.1}$$

where $\phi_\theta(\cdot)$ is an information function in the sense defined by Pukelsheim (1993, page 119) and

$$C_\theta(\xi) = C_{K_\theta}(\xi, \theta) = \left\{ K_\theta^T M^-(\xi, \theta) K_\theta \right\}^{-1}.$$

Here $K_\theta \in \mathbb{R}^{\ell_\theta \times s_\theta}$ represents a matrix of full column rank $s_\theta \leq \ell_\theta$ and $M^-(\xi, \theta)$ denotes a generalized inverse of $M(\xi, \theta)$. Here we assume that $\xi \in \Delta$ is feasible, that is, $\mathcal{R}(K_\theta) \subset \mathcal{R}(M(\xi, \theta))$

for all $\theta \in \Theta$. Note that this formulation is in fact very general and encompasses a broad range of model specifications and optimality criteria. Numerous examples are presented later in the paper.
An Equivalence Theorem for Bayesian Ψ_q-optimal and standardized maximin ψ-optimal designs based on criteria of the form (3.1) is now introduced and holds strictly for classes of designs Δ which are convex. The formulation adopted here is that of Pukelsheim (1993) and relies on the definition of the polar function of $\phi_\theta(\cdot)$ given by

$$\phi_\theta^\infty(D) = \inf_C \Big\{ \frac{\mathrm{tr}(CD)}{\phi_\theta(C)} \Big| C > 0 \Big\}$$

where C and D are nonnegative definite matrices. The proof of the next theorem follows essentially the same arguments as those presented in Pukelsheim (1993), Chapter 11, and is therefore omitted.

THEOREM 3.1. *Assume that the criterion $\psi(\xi, \theta)$ has the form (3.1) and that the class of designs Δ is convex. Assume also that a design denoted $\xi^* \in \Delta$ satisfies the condition $\mathcal{R}(K_\theta) \subset \mathcal{R}(M(\xi^*, \theta))$ for all $\theta \in \Theta$.*

(a) The design ξ^* is Bayesian Ψ_q-optimal in the class Δ with respect to a prior π on Θ if and only if for each $\theta \in \Theta$ there exists a nonnegative definite matrix D_θ which solves the polarity equation

$$\phi_\theta\{C_{K_\theta}(\xi^*)\}\phi_\theta^\infty(D_\theta) = \mathrm{tr}\{\mathrm{C}_{\mathrm{K}_\theta}(\xi^*)\mathrm{D}_\theta\} = 1 \tag{3.2}$$

and a generalized inverse of $M(\xi^*, \theta)$, say G_θ, such that the inequality

$$\int_\Theta \left\{ \frac{\psi(\xi^*, \theta)}{\psi(\xi_\theta^*, \theta)} \right\}^q \mathrm{tr}\{\mathrm{M}(\eta, \theta)\mathrm{B}(\xi^*, \theta)\} \, \mathrm{d}\pi(\theta) \;-\; \int_\Theta \left\{ \frac{\psi(\xi^*, \theta)}{\psi(\xi_\theta^*, \theta)} \right\}^{\mathrm{q}} \mathrm{d}\pi(\theta) \;\leq\; 0 \tag{3.3}$$

holds for all $\eta \in \Delta$, where

$$B(\xi^*, \theta) = G_\theta K_\theta C_\theta(\xi^*) D_\theta C_\theta(\xi^*) K_\theta^T G_\theta$$

(b) Let

$$\mathcal{N}(\xi^*) := \Big\{ \theta \in \Theta \mid \Psi_{-\infty}(\xi^*) = \frac{\psi(\xi^*, \theta)}{\psi(\xi_\theta^*, \theta)} \Big\} \tag{3.4}$$

denote the set of all parameter values in Θ, for which the minimum in (2.1) is attained. Then the design ξ^* is standardized maximin ψ-optimal in the class Δ if and only if there exists a prior π_ω on the set $\mathcal{N}(\xi^*)$, for each $\theta \in \mathrm{supp}(\pi_\omega)$ a nonnegative definite matrix D_θ satisfying (3.2) and a generalized inverse of $M(\xi^*, \theta)$, say G_θ, such that the inequality

$$\int_{\mathcal{N}(\xi^*)} \mathrm{tr}\{\mathrm{M}(\eta, \theta)\mathrm{B}(\xi^*, \theta)\} \, \mathrm{d}\pi_\omega(\theta) \;-\; 1 \;\leq\; 0$$

holds for all $\eta \in \Delta$.

Note that in the case of differentiability the left hand side of the inequality (3.3) is the directional derivative of the optimality criterion at the point ξ^* in the direction of η [see Silvey (1980)]. The more general formulation of Theorem 3.1 is required for non-differentiable criteria. Morover, the second part of this theorem in effect states that the standardized maximin ψ-optimal design ξ^* coincides with the Bayesian Ψ_0-optimal design for the prior distribution π_w defined on the set $\mathcal{N}(\xi^*)$. The prior π_w is usually referred to as the least favourable or "worst" prior, a term which is borrowed from Bayesian decision theory [see Berger (1985), page 360]. At this point it is useful to introduce an example in order to fix ideas.

Example 3.2. Suppose that $K_\theta = I_\theta$ for all $\theta \in \Theta$ and that Δ represents the class of all approximate designs defined on the space $\mathcal{X}$, namely Ξ. Then $C_\theta(\xi) = M(\xi, \theta)$ and, for any optimal design ξ^*, the information matrix $M(\xi^*, \theta)$ has full rank for all $\theta \in \Theta$. Suppose further that the optimality criteria of interest are based on the information function

$$\phi_\theta\{M(\xi, \theta)\} = |M(\xi, \theta)|^{1/\ell_\theta}.$$

Then $G_\theta = M^{-1}(\xi, \theta)$ and $D_\theta = \frac{1}{\ell_\theta} M^{-1}(\xi, \theta)$ for all $\theta \in \Theta$ [see Pukelsheim (1993), page 154] and it follows immediately from Theorem 3.1 and from a standard argument in design theory that the design ξ^* is Bayesian Ψ_q-optimal, i.e. ξ^* maximizes the criterion

$$\Psi_q(\xi) = \Big\{ \int_\Theta \Big(\frac{|M(\xi, \theta)|}{|M(\xi^*_\theta, \theta)|} \Big)^{q/\ell_\theta} d\pi(\theta) \Big\}^{1/q}$$

if and only if the inequality

$$\int_\Theta \Big(\frac{|\, M(\xi^*, \theta) \,|}{|\, M(\xi^*_\theta, \theta) \,|} \Big)^{q/\ell_\theta} \frac{f^T(x, \theta) M^{-1}(\xi^*, \theta) f(x, \theta)}{\ell_\theta} d\pi(\theta) \; - \; \int_\Theta \Big(\frac{|\, M(\xi^*, \theta) \,|}{|\, M(\xi^*_\theta, \theta) \,|} \Big)^{q/\ell_\theta} d\pi(\theta) \; \le \; 0 \quad (3.5)$$

holds for all $x \in \mathcal{X}$. Similarly, a design ξ^* maximizes the criterion

$$\Psi_{-\infty}(\xi) = \min_{\theta \in \Theta} \Big\{ \Big(\frac{|M(\xi, \theta)|}{|M(\xi^*_\theta, \theta)|} \Big)^{1/\ell_\theta} \Big\}$$

and is thus standardized maximin D-optimal if and only if there exists a prior π_ω on the set $\mathcal{N}(\xi^*)$ such that the inequality

$$\int_{\mathcal{N}(\xi^*)} \frac{f(x, \theta)^T M^{-1}(\xi^*, \theta) f(x, \theta)}{\ell_\theta} \, d\pi_w(\theta) \; - \; 1 \; \le \; 0 \qquad (3.6)$$

holds for all $x \in \mathcal{X}$. Furthermore equality in the conditions (3.5) and (3.6) is attained at the support points of the Bayesian Ψ_q-optimal and of the standardized maximin D-optimal designs respectively. Note that the Equivalence Theorem for Bayesian Ψ_q-optimality relating to inequality (3.5) is given in Dette and Wong (1996) under the additional assumption that the dimension ℓ_θ is a constant. □

The next result follows immediately from Theorem 3.1 and provides insight into the nature of the set $\mathcal{N}(\xi^*)$ defined in (3.4). The proof is given in the Appendix.

Lemma 3.3. *Suppose that the parameter space Θ comprises at least two points and that the class of designs Δ is convex. Then, for the standardized maximin ψ-optimal design $\xi^* \in \Delta$, the cardinality of the set $\mathcal{N}(\xi^*)$ defined in (3.4) is at least 2.*

In summary therefore, suppose that a candidate standardized maximin D-optimal design, say ξ_c^*, is available. Then the global optimality or otherwise of this design over a class of designs Δ which is convex can be confirmed by invoking Theorem 3.1 together with Lemma 3.3. Note immediately that a necessary condition for this design to be optimal over the class Δ is that the cardinality of the set $\mathcal{N}(\xi_c^*)$ is at least 2. Note also that it is not straightforward to invoke the second part of Theorem 3.1 in practice and specifically that it is not easy to construct the least favourable prior π_w.
The next two results follow directly from Theorem 3.1 and Lemma 3.3. The proofs are straightforward and are therefore omitted.

LEMMA 3.4. *The Bayesian Ψ_q-optimal design ξ^* with respect to the prior π is Bayesian $\Psi_{q'}$-optimal with respect to the prior $\tilde{\pi}'$, where*

$$d\tilde{\pi}'(\theta) = \left(\frac{\mid M(\xi^*, \theta) \mid}{\mid M(\xi_\theta^*, \theta) \mid} \right)^{q-q'} d\pi(\theta)$$

and q and q' are such that $-\infty < q, q' \leq 0$.

THEOREM 3.5. *The standardized maximin ψ-optimal design ξ^* is Bayesian Ψ_q-optimal with respect to the least favourable prior π_w on the set $\mathcal{N}(\xi^*)$ for all $q \leq 0$. Conversely, if the design ξ^* is Bayesian Ψ_q-optimal for all q such that $-\infty < q \leq 0$, then it is standardized maximin ψ-optimal.*

Thus the global optimality or otherwise of a candidate standardized maximin ψ-optimal design can be examined by checking whether or not the design is Bayesian Ψ_q-optimal for a range of values of $q \leq 0$. Moreover, if a prior which is close to the least favourable prior π_w is adopted, then the Bayesian Ψ_q-optimal designs for that prior is not expected to change greatly as the value of q changes from $-\infty$ to 0. This latter observation in turn relates to and explains certain of the numerical results given in Imhof (2001).

3.2 Convergence of Bayesian to standardized maximin designs

The results presented so far, in particular those relating to Theorems 3.1 and 3.5, indicate that a close relationship exists between Bayesian Ψ_q-optimal designs and maximin ψ-optimal designs. Furthermore, since the criterion $\Psi_q(\xi)$ converges to the maximin criterion $\Psi_{-\infty}(\xi)$ as $q \to -\infty$, it is tempting to surmise that this convergence is mirrored in the corresponding designs themselves. In fact the following theorem shows that, under fairly general conditions, standardized maximin ψ-optimal designs can be obtained as weak limits of Bayesian Ψ_q-optimal designs as $q \to -\infty$. The proof is given in the Appendix, while numerous applications can be found in the next two sections. Note that the result does not require the special structure $\psi(\xi, \theta) = \phi_\theta\{C_\theta(\xi)\}$ nor the convexity of the set $\Delta \subset \Xi$ as was the case for Theorem 3.1.

THEOREM 3.6. *Let Θ be compact and let π denote a prior distribution on Θ with* $\text{supp}(\pi) = \Theta$. *Suppose that the optimality criterion $\psi : \Delta \times \Theta \to (0, \infty)$ is continuous in each argument. Suppose that for every $q < 0$, ζ_q is a Bayesian Ψ_q-optimal design in the class of designs Δ with respect to the prior π and suppose also that the designs ζ_q converge weakly to some design $\zeta^* \in \Delta$ as $q \to -\infty$. Then the design ζ^* is standardized maximin ψ-optimal.*

It should be emphasized that no conditions need be placed on the class of designs Δ in order for the above theorem to hold, other than that the optimal designs of interest, $\{\xi_q\}$ and ξ^*, should all belong to the same class. This is in contrast to Theorem 3.1 for which the requirement that the class of designs Δ be convex is imposed.

REMARK 3.7. In practice it may well be possible to use Theorem 3.6 to construct a maximin ψ-optimal design over a class of designs Δ which is not necessarily convex such as, for example, a class of designs based on a fixed number of support points. Then the global optimality or otherwise of this design over a class of designs which is convex and which contains Δ, such as the class of all approximate designs Ξ, can be confirmed by invoking Theorem 3.1 and Lemma 3.3.

4 General applications

Several broad applications of the results developed in Sections 2 and 3 are now presented. In many cases the form of the Bayesian Ψ_q-optimal designs over a particular subclass of designs is known and the standardized maximin ψ-optimal design can be identified by invoking Theorem 3.6 and introducing some additional algebra.

4.1 Nonlinear models

Consider a nonlinear model for which the response variable y follows a distribution from an exponential family with

$$E(y|x) = \eta(x, \theta) \qquad \text{and} \qquad \mathrm{Var}(y|x) = \sigma^2(x) \ , \tag{4.1}$$

where x represents an explanatory variable in the design space $\mathcal{X} \subset \mathbb{R}^\ell$ and θ is a vector of unknown parameters in the space $\Theta \subset \mathbb{R}^k$. If $\eta(x, \theta)$ is continuously differentiable with respect to θ, then the Fisher information matrix for θ at a single point x is given by

$$I(x, \theta) = \frac{1}{\sigma^2(x)} \left\{ \frac{\partial \eta(x, \theta)}{\partial \theta} \right\} \left\{ \frac{\partial \eta(x, \theta)}{\partial \theta} \right\}^T$$

and the information matrix for a design ξ belonging to a specified class of designs Δ can be expressed as

$$M(\xi, \theta) = \int_{\mathcal{X}} I(x, \theta) \, d\xi(x)$$

[see e.g. Silvey (1980)]. For this model setting it is usual to consider criteria $\phi\{M(\xi, \theta)\}$, which are concave functions of the Fisher information matrix. Then the Bayesian Ψ_q-optimality criterion with respect to a prior π on Θ can be formulated as

$$\Psi_q(\xi) = \left[\int_\Theta \left\{ \frac{\phi\{M(\xi, \theta)\}}{\phi\{M(\xi_\theta^*, \theta)\}} \right\}^q d\pi(\theta) \right]^{\frac{1}{q}}$$

and, following Dette (1997), the standardized maximin ϕ-optimality criterion can be written as

$$\Psi_{-\infty}(\xi) = \min_{\theta \in \Theta} \left\{ \frac{\phi\{M(\xi, \theta)\}}{\phi\{M(\xi_\theta^*, \theta)\}} \right\} \ ,$$

where ξ_θ^* represents the locally ϕ-optimal design maximizing the function $\phi\{M(\xi,\theta)\}$ [see Chernoff (1953)] and where all criteria are maximized over the given class of designs Δ. Note that these criteria are obtained as a special case from the general theory with $K_\theta = I_k$ and $\psi(\xi,\theta) = \phi(M(\xi,\theta))$. The following example involves a finite parameter space Θ and helps to fix ideas.

EXAMPLE 4.1. Consider the simple exponential model for which the response y is normally distributed with mean $\exp(-\theta x)$ and variance σ^2, where $\theta > 0$ and $x \in [0,\infty)$. Suppose that a parameter space comprising two values θ_1 and θ_2 with $0 < \theta_1 < \theta_2 < \infty$ is of interest, i.e. $\Theta = \{\theta_1, \theta_2\}$, and that a single-point standardized maximin D-optimal design over that parameter space is sought. Now $I(x,\theta) = x^2\exp(-2\theta x)$ and the single-point locally D-optimal design is given by $x_\theta^* = 1/\theta$ with $I(x_\theta^*,\theta) = 1/(e^2\theta^2)$. Consequently, for a prior on Θ which puts weights β and $1-\beta$ on the values θ_1 and θ_2 respectively, a one-point Bayesian Ψ_q-optimal design with $q<0$, say x_q, maximizes the criterion

$$\Psi_q(x) = \left\{\beta\left[e^2\theta_1^2x^2\exp(-2\theta_1 x)\right]^q + (1-\beta)\left[e^2\theta_2^2x^2\exp(-2\theta_2 x)\right]^q\right\}^{\frac{1}{q}}.$$

By differentiating $\Psi_q(x)$ with respect to x and setting the result to zero, it then follows that x_q satisfies the transcendental equation

$$\frac{\log\beta + \log(1-\theta_1 x)}{2q} + (\log\theta_1 - \theta_1 x) = \frac{\log(1-\beta) + \log(\theta_2 x - 1)}{2q} + (\log\theta_2 - \theta_2 x)$$

with $1/\theta_2 < x < 1/\theta_1$. Thus, as $q \to -\infty$, the design point x_q approaches

$$x^* = \frac{\log(\theta_2/\theta_1)}{(\theta_2-\theta_1)},$$

which, by Theorem 3.6, is a maximin D-optimal one-point design. This result is in agreement with that derived by Haines (1995) and, independently, by Imhof (2001).
It now follows from Theorem 3.1 and Lemma 3.3 that the design point x^* is globally maximin D-optimal if and only if there exists a prior on the two parameters θ_1 and θ_2 for which that point is globally Bayesian D-optimal. A candidate least favourable prior can be formulated by setting x^* equal to the one-point Bayesian D-optimal design $x_0 = 1/\{\beta\theta_1 + (1-\beta)\theta_2\}$ and is thus specified by

$$\beta = \frac{\theta_1}{\theta_2-\theta_1} - \frac{1}{\log(\theta_2/\theta_1)}.$$

By examining the directional derivative for Bayesian D-optimality with respect to this prior [see formula (3.6) with $\ell_\theta = 1$ and $M(\xi^*,\theta) = I(x^*,\theta)$], it can then be shown numerically that the design point x^* is globally standardized maximin D-optimal provided $\theta_2/\theta_1 \leq 3.891$, a result in accord with the findings of Haines (1995). □

The following corollary specifies the fairly general conditions under which Theorem 3.6 holds for the nonlinear models considered in this section. The proof is outlined in the Appendix. Note that, in the statement of the theorem, the set of all nonnegative definite matrices of order $k \times k$ is denoted NND(k)

COROLLARY 4.2. *Consider the nonlinear model specified by (4.1) and a local optimality criterion of the form* $\psi(\xi, \theta) = \phi\{M(\xi, \theta)\}$, *where* $\phi(\cdot)$ *is a continuous function from* NND(k) *to* $[0, \infty)$. *Let* Θ *be compact and let* π *represent a prior distribution on* Θ *for which* $\mathrm{supp}(\pi) = \Theta$. *Suppose that* $\psi(\xi, \theta) > 0$ *on* $\Delta \times \Theta$ *and that* $I(x, \theta)$ *is bounded and continuous on* $\mathcal{X} \times \Theta$. *Then, as* $q \to -\infty$, *the weak limit of the Bayesian* Ψ_q*-optimal designs with respect to the prior* π *in the class of designs* Δ *is a standardized maximin* ϕ*-optimal design, provided it belongs to* Δ.

The next example illustrates the use of the above corollary in the construction of standardized maximin D-optimal designs for a nonlinear model with an associated parameter space which is not finite.

EXAMPLE 4.3. Consider the one-parameter logistic regression model with probability of success $1/\{1 + \exp(-(x - \theta))\}$ and $x \in I\!R$. Note that the information on θ at an observation x is given by

$$I(x, \theta) = \frac{\exp(-(x - \theta))}{\{1 + \exp(-(x - \theta))\}^2}$$

and is bounded and continuous. Note also that the locally D-optimal one-point design is located at $x^*_\theta = \theta$ with $I(x^*_\theta, \theta) = 1/4$. Suppose now that a parameter space of the form $\Theta = [-a, a]$ with $a > 0$ is of interest and that single-point standardized maximin D-optimal designs over that space are to be constructed. For a uniform prior on Θ, the one-point Bayesian Ψ_q-optimal design, say x_q, maximizes the criterion

$$\Psi_q(x) = \left\{ \frac{1}{2a} \int_{-a}^{a} \left[\frac{4 \exp(-(x - \theta))}{\{1 + \exp(-(x - \theta))\}^2} \right]^q d\theta \right\}^{\frac{1}{q}} \quad \text{for } -\infty < q < 0$$

and it is straightforward to show, either algebraically or by symmetry arguments, that $x_q = 0$ for all such q. Thus, since the conditions specified in Corollary 4.2 are satisfied for this example, it follows trivially that the one-point standardized maximin D-optimal design is given by $x^* = 0$.

Consider now invoking Theorem 3.1 in order to determine whether or not taking all observations at the point $x^* = 0$ is globally maximin D-optimal. On the basis of the results presented in Theorem 3.5, the uniform prior on $[-a, a]$ is adopted as a candidate least favourable prior. The directional derivative for Bayesian D-optimality from $x^* = 0$ to a design point x is then given by

$$\frac{1}{2a} \int_{-a}^{a} \exp(-x) \left(\frac{1 + \exp(\theta)}{1 + \exp(-(x - \theta))} \right)^2 d\theta - 1$$

[see formula (3.6) with $\ell_\theta = 1$ and $M(\xi^*, \theta) = I(x^*, \theta)$] and it can be shown numerically that this derivative is less than or equal to zero for all $x \in I\!R$ provided

$$3 + a - 3\exp(a) + a\exp(a) \leq 0.$$

Thus it follows that the single-point design $x^* = 0$ is globally standardized maximin D-optimal on the parameter space $[-a, a]$ provided a satisfies this inequality and hence provided $a \leq 2.5757$.

4.2 Linear heteroscedastic models

An interesting and widely encountered form of nonlinear design problem relates to linear models with an heteroscedastic error structure. To be precise, assume that the response y follows a distribution from an exponential family and that

$$E(y|x) = \beta_0 f_0(x) + \cdots + \beta_d f_d(x) \quad \text{and} \quad \text{var}(y|x) = \sigma^2/\lambda(x,\theta) \tag{4.2}$$

where $x \in \mathcal{X}$, $f_0(x), \ldots, f_d(x)$ are continuous linearly independent regression functions on the design space $\mathcal{X}$ and $\lambda(x,\theta)$ is an efficiency function assumed to be known up to the value of the parameter $\theta \in \Theta$ [see e.g. Fedorov (1972)]. The Fisher information matrix for the parameter β in model (4.2) is given by

$$M(\xi,\theta) = \int_{\mathcal{X}} f(x) f^T(x) \lambda(x,\theta)\, d\xi(x)$$

where $f(x) = (f_0(x), \ldots, f_d(x))^T$ denotes the vector of regression functions. Criteria of interest for this model setting are usually of the form $\psi(\xi,\theta) = \phi\{M(\xi,\theta)\}$ where $\phi(\cdot)$ is a concave function of the information matrix, such as the $(d+1)$th root of the determinant. The definitions of Bayesian Ψ_q-optimal and standardized maximin ψ-optimal designs are essentially the same as those given for the nonlinear models specified by (4.1) and are not repeated here.
In the present setting the mean of the random variable y defines a linear model and nonlinearity is introduced through the dependence of the variance function on the parameter θ. In fact, by including an additional parameter in the variance function of the nonlinear model specified in (4.1), model (4.2) can be treated as a special case of that nonlinear model [see Atkinson and Cook (1995)]. The model settings are however fundamentally different and are treated as such in the present paper. Specifically, for the nonlinear model (4.1) the parameter θ is the parameter of interest, while for the linear heteroscedastic model (4.2) the parameter $\beta = (\beta_0, \ldots, \beta_d)$ is of importance and θ appears simply as a nuisance parameter. The following corollary to Theorem 3.6 provides conditions for that theorem to hold in the present setting. The proof is given in the Appendix.

Corollary 4.4. *Suppose that the regression functions $f_0(x), \ldots, f_d(x)$ and the efficiency function $\lambda(x,\theta)$ in the linear heteroscedastic model (4.2) are continuous and that $\psi(\xi,\theta) = \phi\{M(\xi,\theta)\}$ where $\phi(\cdot)$ is a continuous function from $NND(d+1)$ to $[0,\infty)$. Let the parameter space Θ be compact and let π represent a prior on Θ with* $\text{supp}(\pi) = \Theta$. *Suppose that $\psi(\xi,\theta) > 0$ on $\Delta \times \Theta$ and that at least one of the following conditions is met.*

(i) The design space $\mathcal{X}$ is compact.

(ii) $\mathcal{X} = [a,\infty)$, $a \in \mathbb{R}$, and

$$\lim_{x\to\infty} \left\{ f_j(x) \max_{\theta\in\Theta} \sqrt{\lambda(x,\theta)} \right\} = 0, \qquad j = 0, \ldots, d.$$

(iii) $\mathcal{X} = \mathbb{R}$ and

$$\lim_{x\to-\infty} \left\{ f_j(x) \max_{\theta\in\Theta} \sqrt{\lambda(x,\theta)} \right\} = \lim_{x\to\infty} \left\{ f_j(x) \max_{\theta\in\Theta} \sqrt{\lambda(x,\theta)} \right\} = 0, \qquad j = 0, \ldots, d.$$

Then the weak limit of the Bayesian Ψ_q-optimal designs with respect to the prior π in the class of designs Δ as $q \to -\infty$ is a standardized maximin ψ-optimal design, provided it belongs to Δ.

Finally, if the class of designs Δ is convex, Theorem 3.1 can be invoked to confirm the global optimality or otherwise of a candidate design. Specifically, if $\phi(M(\xi,\theta)) = |M(\xi,\theta)|^{\frac{1}{d+1}}$, a design ξ^* is Bayesian Ψ_q-optimal in the class of all designs if and only if the inequality

$$\int_\Theta \Big(\frac{|M(\xi^*,\theta)|}{|M(\xi^*_\theta,\theta)|}\Big)^{\frac{q}{d+1}} \lambda(x,\theta) f^T(x) M^{-1}(\xi^*,\theta) f(x)\ d\pi(\theta) \ \le\ (d+1) \int_\Theta \Big(\frac{|M(\xi^*,\theta)|}{|M(\xi^*_\theta,\theta)|}\Big)^{\frac{q}{d+1}}\ d\pi(\theta)$$

holds for all $x \in \mathcal{X}$ and standardized maximin D-optimal if and only if there exists a least favourable prior π_w on the set $\mathcal{N}(\xi^*)$ defined in (3.4) such that

$$\int_{\mathcal{N}(\xi^*)} \lambda(x,\theta) f^T(x) M^{-1}(\xi^*,\theta) f(x)\ d\pi_w(\theta) \ \le\ (d+1)$$

is satisfied, again for all $x \in \mathcal{X}$. The application of the above results to the construction of standardized maximin D-optimal designs for a particular class of weighted polynomial regression models is now explored in the following example.

Example 4.5. Consider the weighted polynomial model defined by

$$y = \beta_0 + \beta_1 x + \ldots + \beta_d x^d + \epsilon, \qquad x \in [0,\infty), \tag{4.3}$$

where the regression parameters $\beta = (\beta_0, \ldots, \beta_d)$ are of interest and the error term ϵ is assumed to be normally distributed with mean 0 and variance $\sigma^2/\lambda(x,\theta)$. The efficiency function is taken to be of the form $\lambda(x,\theta) = x^v \exp(-\theta x)$, with the parameter θ deemed to be a nuisance parameter and the power v either positive or zero and assumed known. Then the Fisher information matrix for a design ξ belonging to a specified class of designs Δ can be written as

$$M(\xi,\theta) = \int_0^\infty x^v \exp(-\theta x) f(x) f^T(x)\, d\xi(x)$$

where $f(x) = (1, x, \ldots, x^d)^T$. Suppose that the criterion of interest is D-optimality and thus that $\psi(\xi,\theta) = |M(\xi,\theta)|^{1/(d+1)}$. Suppose also that θ belongs to a compact parameter space $\Theta = [\theta_{\min}, \theta_{\max}]$ with $0 \le \theta_{\min} < \theta_{max} < \infty$ and that attention is restricted to prior distributions π on Θ for which $\text{supp}(\pi) = \Theta$. Then all the broad assumptions of Corollary 4.4 hold and, in addition, since $x \in [0,\infty)$ and

$$\lim_{x\to\infty} \Big\{ f_j(x) \max_{\theta\in\Theta} \sqrt{\lambda(x,\theta)} \Big\} = \lim_{x\to\infty} \Big\{ \frac{x^{j+v/2}}{\exp(\theta_{\max} x/2)} \Big\} = 0$$

condition (ii) of that corollary is satisfied. Thus, in summary, the design problem presented here is precisely that for a linear heteroscedastic model described above, Corollary 4.4 can be invoked and standardized maximin D-optimal designs obtained as the limits of the appropriate Bayesian Ψ_q-optimal designs.

Consider now optimal designs which are restricted to belong to the class of all $(d+1)$-point designs. Then it follows immediately from the results of Dette and Wong (1996) that designs which are Bayesian Ψ_q-optimal over this class for $-\infty < q < 0$ put equal masses at the zeros of the polynomials

$$xL_d^{(1)}\{-F_q(qz_q)x\} \qquad \text{for } v = 0$$

and

$$L_{d+1}^{(v-1)}\{-F_q(qz_q)x\} \qquad \text{for } v > 0.$$

Here $L_n^{(\alpha)}(u)$ denotes the generalized Laguerre polynomial of degree n orthogonal with respect to the measure $u^\alpha \exp(-u)\,du$ for $u \geq 0$ [see Szegö (1975)] and $z = z_q$ is the unique solution of the equation

$$z = -\frac{(d+1)(d+v)}{F_q(qz)},$$

where

$$F_q(qz) = -\frac{\int_\Theta \theta e^{-\theta qz}\theta^{(d+1)(d+v)q}\,d\pi(\theta)}{\int_\Theta e^{-\theta qz}\theta^{(d+1)(d+v)q}\,d\pi(\theta)}.$$

In order to determine the weak limit of these Bayesian Ψ_q-optimal designs, that is to determine the limit of $F_q(qz)$ as $q \to -\infty$, the following lemma is introduced. The proof is intricate and is given in the Appendix.

LEMMA 4.6. *Let π be a prior distribution with support $\Theta = [\theta_{\min}, \theta_{\max}]$ where $0 \leq \theta_{\min} < \theta_{\max}$. Let*

$$F_q(x) = -\frac{\int \theta e^{-\theta x} g(\theta)^{-q}\,d\pi(\theta)}{\int e^{-\theta x} g(\theta)^{-q}\,d\pi(\theta)}, \qquad -\infty < q < 0, \quad x \in \mathbb{R},$$

where g is a continuous log-convex function on Θ. For every $q < 0$, let z_q be such that

$$z_q = h\{-F_q(qz_q)\}, \tag{4.4}$$

where $h(\cdot)$ is a strictly decreasing function on Θ. If there exists a parameter value $t^* \in \Theta$ such that

$$h(t^*) = -\frac{\log\{g(\theta_{\max})/g(\theta_{\min})\}}{\theta_{\max} - \theta_{\min}}, \tag{4.5}$$

then $\lim_{q\to-\infty} F_q(qz_q) = -t^*$.

In the present case $g(\theta) = \theta^{-(d+1)(d+v)}$, which is log convex, and $h(t) = (d+1)(d+v)/t$, which is decreasing. Thus Lemma 4.6 holds and

$$\lim_{q\to-\infty} F_q(qz_q) = -t^* = -\frac{\theta_{\max} - \theta_{\min}}{\log(\theta_{\max}/\theta_{\min})}$$

provided that $\theta_{\min} \leq t^* \leq \theta_{\max}$. In fact, since $\log u \leq u - 1$ for all $u > 0$, the inequalities involving t^* do indeed hold and the following result is thus established.

THEOREM 4.7 *The standardized maximin D-optimal $(d+1)$-point design for the weighted polynomial regression model (4.3) with*

$$Var(y|x) = \frac{\sigma^2}{\lambda(x,\theta)} = \sigma^2 x^{-v}\exp(-\theta x)\ ,\ x \geq 0$$

and $\theta \in [\theta_{\min}, \theta_{\max}]$ *puts equal masses at the zeros of the polynomials*

$$xL_d^{(1)}\left\{\frac{\theta_{\max}-\theta_{\min}}{\log\left(\theta_{\max}/\theta_{\min}\right)}x\right\} \qquad \textit{for } v=0,$$

and

$$L_{d+1}^{(v-1)}\left\{\frac{\theta_{\max}-\theta_{\min}}{\log\left(\theta_{\max}/\theta_{\min}\right)}x\right\} \qquad \textit{for } v>0.$$

Note that this theorem extends the results for the case of $v = 0$ given in Theorem 5.1 of Imhof (2001).

REMARK 4.8. Consider the generalized exponential model defined by

$$y = x^v \exp(-\theta x)\{\beta_0 + \beta_1 x + \ldots \beta_{d-1}x^{d-1}\} + \epsilon, \qquad x \in [0, \infty), \tag{4.6}$$

where the error terms ϵ are normally distributed with mean zero and constant variance σ^2, the parameters $\beta = \{\beta_0, \ldots, \beta_{d-1}\}$ and θ are of interest and the parameter v is assumed to be known. Then, following Dette and Wong (1996), it can readily be shown that the problem of constructing Bayesian Ψ_q-optimal designs for this model is equivalent to that of finding such designs for the weighted polynomial regression model with efficiency function $\lambda(x,\theta) = x^{2v}\exp(-2\theta x)$. Thus it follows immediately that standardized maximin D-optimal designs for the exponential model (4.6) can be derived from the corresponding optimal designs given in Example 4.5.

4.3 Model robust and discrimination designs

It is not uncommon for a practitioner to identify a set of plausible models, rather than a single model, as being appropriate for a particular data set. In order to accommodate such model uncertainty within the context of optimal design, criteria which are robust to the choice of model have been developed [see e.g. Läuter (1974)] and certain of these are explored here. To be specific, consider a class of linear models with means

$$E(y|x) = g(x,\theta) = \beta_0 f_0(x,\theta) + \ldots + \beta_{\ell_\theta} f_{\ell_\theta}(x,\theta),$$

where x belongs to some design space $\mathcal{X}$ and the regression functions $f_i(x,\theta), i = 0,\ldots,\ell_\theta$, are known, and with constant variances σ^2, i.e. $\lambda(x,\theta) \equiv 1$. Each model is indexed by a parameter θ taken from a finite set of indices Θ and the class of such models is denoted $\mathcal{F} = \{g(x,\theta) \mid \theta \in \Theta\}$. Note that in many applications the models in the set $\mathcal{F}$ are nested but this is not necessary for the development of the robust design criteria described here.
The Fisher information matrix for the regression parameters $(\beta_0,\ldots,\beta_{\ell_\theta})$ in the model specified by $g(x,\theta)$ at a design $\xi \in \Delta$ can be expressed as

$$M(\xi,\theta) = \frac{1}{\sigma^2}\Big(\int_{\mathcal{X}} f_i(x,\theta) f_j(x,\theta) d\xi(x)\Big)_{i,j=0}^{\ell_\theta}$$

for $\theta \in \Theta$. Thus an optimal design which is robust to choice of model over the class $\mathcal{F}$ should maximize an appropriate real valued function of the matrices $\{M(\xi, \theta) \mid \theta \in \Theta\}$ over the set of designs Δ [see e.g. Läuter (1974)]. In particular, suppose that a prior π on the index set Θ puts probability $\pi(\theta)$ on the parameter θ, where $\pi(\theta) \geq 0$ and $\sum_{\theta \in \Theta} \pi(\theta) = 1$. Suppose also that for each model $g(x, \theta)$ in the class $\mathcal{F}$, a criterion of the form $\psi(\xi, \theta) = \phi_\theta\{M(\xi, \theta)\}$, where $\phi_\theta(\cdot)$ is an information function, is of interest and that ξ_θ^* is the locally ϕ_θ-optimal design associated with this criterion. Then, following Läuter (1974), a Ψ_q-optimal robust design with respect to the prior π for the class of models $\mathcal{F}$ maximizes the criterion

$$\Psi_q(\xi) = \left[\sum_{\theta \in \Theta} \pi(\theta) \left\{ \frac{\phi_\theta\{M(\xi, \theta)\}}{\phi_\theta\{M(\xi_\theta^*, \theta)\}} \right\}^q \right]^{\frac{1}{q}} \tag{4.7}$$

over the set of designs Δ. Furthermore, following Dette (1997), a standardized maximin optimal robust design for the class $\mathcal{F}$ maximizes the function

$$\min_{\theta \in \Theta} \left\{ \frac{\phi_\theta\{M(\xi, \theta)\}}{\phi_\theta\{M(\xi_\theta^*, \theta)\}} \right\}$$

again over the set Δ. In view of Theorem 3.6, the standardized maximin robust designs can be found as weak limits of Ψ_q-optimal robust designs. Furthermore, if the class of designs Δ is taken to be convex, Theorem 3.1 can be invoked in order to ascertain whether or not a candidate design is in fact globally optimal. These ideas are illustrated by means of the following example, which discusses the problem of identifying the degree of a polynomial regression.

Example 4.9. Consider the class of nested polynomial models with means

$$g(x, \theta) = \beta_0 + \beta_1 x + \ldots + \beta_\theta x^\theta \ ,$$

where $x \in \mathcal{X} = [-1, 1]$ and $\theta \in \Theta = \{1, \ldots d\}$. Note that the regression functions are given by $f_i(x, \theta) = x^i, \ i = 0, \ldots, \theta$, and that the information matrix for the model of degree θ can be expressed as

$$M(\xi, \theta) = \Big(\int_{\mathcal{X}} x^{i+j} d\xi(x) \Big)_{i,j=0}^{\theta}.$$

In order to obtain efficient designs for identifying the appropriate degree of the polynomial regression, Spruill (1990) proposed that a function of the criteria

$$\psi(\xi, \theta) = \phi_\theta\{M(\xi, \theta)\} = \frac{|M(\xi, \theta)|}{|M(\xi, \theta - 1)|}$$

for $\theta \in \{1, \ldots, d\}$ should be maximized. The rational behind this proposal is that the power of the t-test for the hypothesis that the highest coefficient in the model of degree θ vanishes, $H_0 : \beta_\theta = 0$, is an increasing function of $\psi(\xi, \theta)$ and consequently that a good discrimination design should make these quantities as large as possible [for more details see Spruill (1990)]. Suppose now that a uniform prior π is placed on the index set Θ, i.e. $\pi(\theta) = \frac{1}{d}$ for $\theta \in \{1, \ldots, d\}$. Then the Ψ_q-optimal (discrimination) design with respect to the prior π, say ξ_q^*, maximizes (4.7) and can be characterized explicitly in terms of its canonical moments [see Dette and Studden (1997)]. In particular, by using results in Dette (1994) it can be shown that the canonical moments $(p_1,\dots,p_{2d})$ of the Ψ_q-optimal (discrimination) design ξ_q^* are given by $p_{2d}=1$, $p_{2j-1}=\frac{1}{2}$ for $j=1,\dots,d$, and by the system of equations

$$2^{2(d-\ell)}\left\{\prod_{i=\ell+1}^{d-1} p_{2i}^{1+1/q}q_{2i}^{1-1/q}\right\}(1-p_{2\ell})^{1-1/q}(2p_{2\ell}-1)^{1/q} \;=\; 1\;, \quad \ell=1,\dots,d-1\;,$$

where $q_{2i}=1-p_{2i}$ and $\prod_d^{d-1}$ is interpreted as unity. As $q\to-\infty$ this latter system reduces to the recursion

$$p_{2\ell}=1-2^{-2(d-\ell)}\prod_{i=\ell+1}^{d-1}(p_{2i}q_{2i})^{-1}$$

and consequently ξ_q^* converges weakly to the design ξ^* with canonical moments $p_{2d}=1, p_{2j-1}=\frac{1}{2}$ for $j=1,\dots,d$, and

$$p_{2\ell}=\frac{d-\ell+2}{2(d-\ell)+2}$$

for $\ell=1,\dots,d-1$. It now follows immediately from Theorem 3.6, and more specifically from the general discussion of this section, that the design ξ^* is standardized maximin optimal. Moreover, by invoking Corollary 4.3.3 in Dette and Studden (1997), it is readily shown that the design ξ^* puts equal masses at the zeros of the ultraspherical polynomial $C_{d-1}^{(2)}(x)$ [see Szegö (1975)] and masses 1.5 times larger at the boundary points $+1$ and -1.
The Ψ_q-optimal and the maximin optimal discrimination designs described here are in fact globally optimal in the sense that they are optimal over the class of all approximate designs, Ξ. Thus there exists a least favorable prior π_w on the index set Θ for which ξ^* is Ψ_0-optimal. Furthermore this prior can be obtained explicitly from the canonical moments of the optimal design ξ^* by invoking Theorem 6.2.3 of Dette and Studden (1997) and puts weights

$$\pi_w(\theta)=\frac{2(d-\theta+1)}{d(d+1)}$$

on the parameters $\theta\in\{1,\dots,d\}$. For example, consider the case of $d=4$. Then the standardized maximin optimal discrimination design is given by

$$\xi^* \;=\; \left\{\begin{array}{ccccc} -1 & -\frac{\sqrt{3}}{2\sqrt{2}} & 0 & \frac{\sqrt{3}}{2\sqrt{2}} & 1 \\ \frac{1}{4} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{4} \end{array}\right\}$$

and the least favourable prior associates weights $2/5$, $3/10$, $1/5$ and $1/10$ with the polynomial models of degree $1,2,3$ and 4, respectively. $\Box$

4.4 Designs for estimating nonlinear functions

Consider the homoscedastic linear regression model with mean

$$E(y|x)=\beta_0f_0(x)+\beta_1f_1(x)+\ldots+\beta_df_d(x) \tag{4.8}$$

and variance σ^2. Suppose that the parameters β belong to a space $\mathcal{B}$ and that a nonlinear function of those parameters, denoted $h(\beta)$, is of interest. Then the approximate asymptotic variance of such a function is proportional to

$$\theta(\beta)^TM^-(\xi)\theta(\beta),$$

where $\theta(\beta)$ represents the vector of derivatives of $h(\beta)$ with respect to β, $M^-(\xi)$ is a generalized inverse of the information matrix and $\theta(\beta) \in \mathcal{R}(M(\xi))$. Optimal designs which in some sense minimize this variance are now sought. For ease of notation, consider the induced parameter space $\Theta = \{[\theta(\beta)^T\theta(\beta)]^{-\frac{1}{2}}\theta(\beta) : \beta \in \mathcal{B}\}$. Then an appropriate optimality criterion can be formulated as

$$\psi(\xi, \theta) = \begin{cases} \{\theta^T M^-(\xi)\theta\}^{-1} & \text{for } \theta \in \mathcal{R}(M(\xi)) \\ 0 & \text{otherwise} \end{cases}$$

and the locally optimal design ξ_θ^* maximizes this criterion. The design problem so described occurs, for example, when the turning point of a quadratic regression function is of interest [see e.g. Chaloner (1989)] and also in the context of constructing optimal extrapolation designs for an interval [see e.g. Spruill (1987)].
The definitions of Bayesian Ψ_q-optimal and of standardized maximin ψ-optimal designs based on the above criterion follow directly from the general formulations given in Section 2. Furthermore Theorem 3.6 holds under the conditions specified in the following corollary. The proof is given in the Appendix.

COROLLARY 4.10. *Let π denote a prior distribution on Θ with* $\text{supp}(\pi) = \Theta$. *Suppose that the functions $f_0(x), \ldots, f_d(x)$ in model (4.8) are continuous and bounded and that the locally optimal criterion value $\psi(\xi_\theta^*, \theta)$ is continuous in θ. Then the weak limit of the Bayesian Ψ_q-optimal designs in the class of designs Δ as $q \to -\infty$ is a standardized maximin ψ-optimal design, provided the limiting design belongs to Δ and is non-singular, i.e. its Fisher information matrix is non-singular.*

5 New designs for heteroscedastic polynomial models

The potential application and usefulness of the results presented in this paper are now explored and illustrated for two heteroscedastic polynomial models for which the Bayesian Ψ_q-optimal and standardized maximin ψ-optimal designs are not available in the literature. In particular the construction of standardized maximin D-optimal designs for these models is considered [see Example 3.2] and a suite of new and interesting results derived. To be specific, the models of interest are defined by

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_d x^d + \epsilon \ , \qquad (5.1)$$

where ϵ is an error term with mean 0 and variance $\sigma^2/\lambda(x, \theta)$ and the efficiency functions and design spaces are of the form

$$\lambda(x, \theta) = (1 + x^2)^{-\theta} \quad \text{for} \quad \theta > d \quad \text{and} \quad x \in I\!R$$

and

$$\lambda(x, \theta) = (1 + x)^{-\theta} \quad \text{for} \quad \theta > 2d \quad \text{and} \quad x \in [0, \infty).$$

Attention is restricted to the class of designs based on exactly $d + 1$ points. Bayesian Ψ_q-optimal designs are first constructed and standardized maximin D-optimal designs are then obtained as weak limits of these designs.

5.1 Bayesian Ψ_q-optimal (d+1)-point designs

Dette and Wong (1996) derived Bayesian Ψ_q-optimal $(d+1)$-point designs for weighted polynomial regression models with a wide range of efficiency functions using arguments based on canonical moments [see Dette and Studden (1997)]. Their approach is not in fact completely general and specifically does not hold for the models under consideration here. However it is possible to determine Bayesian Ψ_q-optimal designs based on exactly $d+1$ points explicitly for these models using arguments based on the theory of differential equations. The steps involved in this approach and the attendant results are now presented.

First observe that any D-optimal $(d+1)$-point design necessarily puts equal masses at its support points, say $x_1, \ldots, x_{d+1}$ [see Fedorov (1972)]. The determinant of the information matrix for the regression parameters of model (5.1) can thus be written as

$$| M(\xi, \theta) | \;=\; \frac{1}{(d+1)^{d+1}} \mid X_R \mid^2 \prod_{i=1}^{d+1} \lambda(x_i, \theta) \;,$$

where X_R is the Vandermonde matrix with ith row $\{1 \; x_i \; x_i^2 \; \ldots \; x_i^d\}$ $(i = 1, \ldots, d+1)$. It then follows immediately that

$$\{\Psi_{q(d+1)}(\xi)\}^{q(d+1)} \;=\; \int_\Theta \Big(\frac{|M(\xi, \theta)|}{|M(\xi_\theta^*, \theta)|}\Big)^q d\pi(\theta) = \frac{1}{(d+1)^{d+1}} \mid X_R \mid^2 \int_\Theta \prod_{i=1}^{d+1} \lambda(x_i, \theta)^q d\tilde{\pi}(\theta),$$

where

$$d\tilde{\pi}(\theta) = \frac{d\pi(\theta)}{|M(\xi_\theta^*, \theta)|^q} \tag{5.2}$$

and ξ_θ^* denotes the appropriate locally D-optimal design. The design ξ_θ^* for the efficiency functions of interest, and certain key properties of that design including the determinant, $|M(\xi_\theta^*, \theta)|$, are summarized in the following two lemmas. Proofs are given in the Appendix.

LEMMA 5.1. *Suppose that $\lambda(x, \theta) = (1 + x^2)^{-\theta}$ with $x \in I\!R$ and $\theta > d$. Then the locally D-optimal design ξ_θ^* in the class of all approximate designs Ξ puts equal weights on the roots of the ultraspherical polynomial $C_{d+1}^{(-\theta-\frac{1}{2})}(\sqrt{-x^2})$,*

$$\prod_{i=1}^{d+1}(1 + x_i^2) \;=\; \prod_{j=1}^{d} \frac{(d - 2\theta - j)^2}{(2d + 1 - 2\theta - 2j)^2}$$

and

$$| M(\xi_\theta^*, \theta) | \;=\; \prod_{j=1}^{d} j^j \prod_{j=1}^{d} \frac{(2\theta - 2j + 1)^{2\theta - 2j + 1}}{(2\theta - j + 1)^{2\theta - j + 1}} \quad .$$

LEMMA 5.2. *Suppose that $\lambda(x, \theta) = (1 + x)^{-\theta}$ with $x \in [0, \infty)$ and $\theta > 2d$. Then the locally D-optimal design ξ_θ^* in the class of all approximate designs Ξ puts equal weights on the roots of the Jacobi polynomial $xP_d^{(1,-\theta-1)}(2x + 1)$,*

$$\prod_{i=1}^{d+1}(1 + x_i) = \prod_{j=1}^{d} \frac{(\theta - j + 1)}{(\theta - d - j)}$$

and

$$| M(\xi_\theta^*, \theta) |= \prod_{j=1}^{d} j^{2j} \frac{(\theta - d - j)^{\theta-d-j}}{(\theta - j + 1)^{\theta-j+1}}.$$

The derivation of the required Bayesian Ψ_q-optimal $(d+1)$-point designs, based on the theory of differential equations can be found in the Appendix. The results are summarized in the next two theorems.

THEOREM 5.3 *Consider model (5.1) with $\lambda(x, \theta) = (1+x^2)^{-\theta}$, $x \in I\!R$ and $\theta > d$. Assume that the condition $\int_\Theta a^{-q\theta} d\tilde{\pi}(\theta) < \infty$ holds for all $a > 1$. Then the Bayesian $\Psi_{q(d+1)}$-optimal $(d+1)$-point design with respect to the prior π puts equal weights on the roots of the ultraspherical polynomial*

$$C_{d+1}^{(F(qz)-\frac{1}{2})}(\sqrt{-x^2}) \ ,$$

where

$$F(qz) = -\frac{\int_\Theta \theta e^{-\theta q z} d\tilde{\pi}(\theta)}{\int_\Theta e^{-\theta q z} d\tilde{\pi}(\theta)} \ ,$$

the prior $\tilde{\pi}$ is given by

$$d\tilde{\pi}(\theta) = \Big(\prod_{j=1}^{d} \frac{(2\theta - j + 1)^{2\theta-j+1}}{(2\theta - 2j + 1)^{2\theta-2j+1}}\Big)^q d\pi(\theta),$$

and z is a solution to the equation

$$z = 2\sum_{j=1}^{d} \log\Big(\frac{d + 2F(qz) - j}{2d + 1 + 2F(qz) - 2j}\Big) \ .$$

THEOREM 5.4 *Consider model (5.1) with $\lambda(x, \theta) = (1+x)^{-\theta}$, $x \in [0, \infty)$ and $\theta > 2d$. Assume that the condition $\int_\Theta a^{-q\theta} d\tilde{\pi}(\theta) < \infty$ holds for all $a > 1$. The Bayesian $\Psi_{q(d+1)}$-optimal $(d+1)$-point design with respect to the prior π puts equal weights on the roots of the Jacobi polynomial*

$$xP_d^{(1,F(qz)-1)}(2x+1) \ ,$$

where the function $F(\cdot)$ is defined in Theorem 5.3, the prior $\tilde{\pi}$ is given by

$$d\tilde{\pi}(\theta) = \Big(\prod_{j=1}^{d} \frac{(\theta - j + 1)^{\theta-j+1}}{(\theta - d - j)^{\theta-d-j}}\Big)^q d\pi(\theta) \ ,$$

and z satisfies the equation

$$z = \sum_{j=1}^{d} \log\Big(\frac{-F(qz) - j + 1}{-F(qz) - d - j}\Big).$$

REMARK 5.5. Note that the solutions to the equations involving z in the above theorems are unique, a result which is based on standard arguments. Morover, it follows immediately from Lemmas 5.1 and 5.2 that the Bayesian $\Psi_{q(d+1)}$-optimal $(d+1)$-point designs given in Theorems 5.3 and 5.4 respectively coincide with the locally D-optimal designs for a best guess of the parameter $\theta = -F(qz)$. Furthermore for Bayesian D-optimality with $q = 0$ this best guess is the mean of the parameter θ over the prior π, i.e. $-F(qz) = -F(0) = E_\pi(\theta)$.

REMARK 5.6. Bayesian Ψ_q-optimal $(d+1)$-point designs for weighted polynomial models with efficiency functions $\lambda(x,\theta) = \exp(-\theta x)$ and $\lambda(x,\theta) = \exp(-\theta x^2)$ are presented in Theorem 3.2 and 3.5 of Dette and Wong (1996). These designs can also be derived using the approach based on differential equation discussed here and the attendant proofs follow closely those for Theorems 5.3 and 5.4 given in the Appendix.

5.2 Standardized maximin D-optimal designs

Consider the standardized maximin D-optimality criterion for the weighted polynomial model (5.1) formulated as

$$\min_{\theta\in\Theta}\Big(\frac{|\ M(\xi,\theta)\ |}{|\ M(\xi_\theta^*,\theta)\ |}\Big)^{1/(d+1)}$$

where the parameter θ relates to the efficiency functions $(1+x^2)^{-\theta}$ and $(1+x)^{-\theta}$ and is assumed to belong to a specified parameter space of the form $\Theta = [\theta_{\min},\theta_{\max}]$. Then, following arguments similar to those used in Example 4.5, the $(d+1)$-point designs which maximize this criterion can be obtained as weak limits of the Bayesian Ψ_q-optimal $(d+1)$-point designs derived above. The relevant results are presented formally in the next two theorems and the attendant proofs are given in the Appendix. Note that for these examples it is straightforward to show that Corollary 4.4 holds but the derivation of the required limiting designs is more intricate.

THEOREM 5.7. *The standardized maximin D-optimal $(d+1)$-point design for model (5.1) with $\lambda(x,\theta) = (1+x^2)^{-\theta}$, $x \in I\!R$, $\theta \in \Theta = [\theta_{\min},\theta_{\max}]$ and $\theta_{\min} > d$ puts equal weights on the roots of the ultraspherical polynomial*

$$C_{d+1}^{(-\theta_0-\frac{1}{2})}(\sqrt{-x^2}),$$

where θ_0 falls in the interior of Θ and satisfies the equation

$$2\sum_{j=1}^{d}\log\Big(\frac{d-2\theta_0-j}{2d+1-2\theta_0-2j}\Big) \;=\; \frac{-\log\{m(\theta_{\max})/m(\theta_{\min})\}}{\theta_{\max}-\theta_{\min}}$$

with

$$m(\theta) = \prod_{j=1}^{d}\frac{(2\theta-2j+1)^{2\theta-2j+1}}{(2\theta-j+1)^{2\theta-j+1}}.$$

Theorem 5.8. *The standardized maximin D-optimal $(d+1)$-point design for model (5.1) with $\lambda(x,\theta) = (1+x)^{-\theta}$, $x \in [0,\infty)$, $\theta \in \Theta = [\theta_{\min}, \theta_{\max}]$ and $\theta_{\min} > 2d$ puts equal weights on the roots of the Jacobi polynomial*

$$xP_d^{(1,-\theta_0-1)}(2x+1)\ ,$$

where θ_0 falls in the range Θ and satisfies the equation

$$\sum_{i=1}^{d} \log\Big(\frac{\theta_0 - j + 1}{\theta_0 - j - d}\Big) \;=\; \frac{-\log(m(\theta_{\max})/m(\theta_{\min})))}{\theta_{\max} - \theta_{\min}}$$

with

$$m(\theta) = \prod_{j=1}^{d} \frac{(\theta - d - j)^{\theta - d - j}}{(\theta - j + 1)^{\theta - j + 1}}.$$

Remark 5.9.

(a) Note that the equations involving θ_0 in the two theorems presented above have unique solutions in the range Θ. Note also however that these equations must be solved numerically, at least in general.

(b) The standardized maximin D-optimal $(d+1)$-point designs derived in Theorems 5.7 and 5.8 clearly coincide with the corresponding locally D-optimal designs with a best guess of the parameter θ equal to θ_0 and thus, following Remark 5.5, with Bayesian D-optimal designs with $E_\pi(\theta) = \theta_0$. This observation can be helpful in constructing candidate least favourable priors π_w on the set $\mathcal{N}(\xi^*)$ which can in turn be used together with Theorem 3.1 to explore the global optimality or otherwise of the standardized maximin D-optimal $(d+1)$-point designs.

Example 5.10. Consider the weighted quadratic regression model with efficiency function $\lambda(x,\theta) = (1+x)^{-\theta}$ and $x \in [0,\infty)$. Note that $d = 2$ and thus that the parameter θ must be strictly greater than 4. Suppose now that θ belongs to a specified parameter space $\Theta = [\theta_{\min}, \theta_{\max}]$. Then the standardized maximin D-optimal design based on exactly 3 points, say ξ^*, puts equal weights on the zeros of the polynomial $xP_2^{(1,-\theta_m-1)}(2x+1)$, where θ_m belongs to the parameter space Θ and satisfies

$$\frac{\theta_m(\theta_m - 1)}{(\theta_m - 3)(\theta_m - 4)} = \left[\frac{m(\theta_{\min})}{m(\theta_{\max})}\right]^{\frac{1}{\theta_{\max} - \theta_{\min}}} = c$$

with

$$m(\theta) = \frac{(\theta - 3)^{\theta - 3}(\theta - 4)^{\theta - 4}}{\theta^\theta (\theta - 1)^{\theta - 1}}.$$

In fact it can be shown by a tedious but straightforward calculation that θ_m is unique and is given explicitly by

$$\theta_m = \frac{7c - 1 + \sqrt{1 + 34c + c^2}}{2(c-1)}\ ,$$

and that the support points of the standardized maximin D-optimal 3-point design ξ^* are located at 0,

$$\frac{3(\theta_m - 3) - \sqrt{3(\theta_m - 1)(\theta_m - 3)}}{(\theta_m - 3)(\theta_m - 4)} \quad \text{and} \quad \frac{3(\theta_m - 3) + \sqrt{3(\theta_m - 1)(\theta_m - 3)}}{(\theta_m - 3)(\theta_m - 4)}.$$

Suppose now that the efficiency of the standardized maximin D-optimal design relative to the locally D-optimal design, defined by

$$R(\xi^*, \theta) = \left\{ \frac{|M(\xi^*, \theta)|}{|M(\xi^*_\theta, \theta)|} \right\}^{1/3},$$

attains its minimum over the parameter space at the end-points of that space, $\theta_{\min}$ and $\theta_{\max}$. Then $\mathcal{N}(\xi^*) = \{\theta_{\min}, \theta_{\max}\}$ and a candidate worst prior based on the two points $\theta_{\min}$ and $\theta_{\max}$ can be constructed. Furthermore, following Remark 5.8, the maximin optimal design ξ^* coincides with the three-point Bayesian D-optimal design associated with a prior distribution on θ for which $E(\theta) = \theta_m$. Thus the candidate worst prior puts weights α and $(1-\alpha)$ on the parameter values $\theta_{\min}$ and $\theta_{\max}$ where α satisfies

$$\alpha\theta_{\min} + (1-\alpha)\theta_{\max} = \theta_m.$$

It then follows from Theorem 3.1 that the 3-point maximin D-optimal design, ξ^*, is globally optimal provided it is Bayesian D-optimal for the candidate worst prior and thus provided the directional derivative

$$d(x, \xi^*) = \alpha \mathrm{tr}\{M(x, \theta_{\min})M^{-1}(\xi^*, \theta_{\min})\} + (1-\alpha)\mathrm{tr}\{M(x, \theta_{\max})M^{-1}(\xi^*, \theta_{\max})\} - 1$$

is less than or equal 0 for all $x \in [0, \infty)$. For the parameter space $\Theta = [5, 6]$, the value of θ_m is 5.4665 and the standardized maximin D-optimal 3-point design, ξ^*, puts equal weights on the points $0, 0.4563$ and 3.6350. The efficiency $R(\xi^*, \theta)$ is a minimum at the end points of Θ, i.e. at 5 and 6, and the worst prior associates weights 0.5335 and 0.4665 respectively with those values. A plot of the directional derivative $d(x, \xi^*)$ against x for this setting shows, at least numerically, that the design ξ^* is globally optimal. For the parameter space $\Theta = [5, 10]$, the maximin D-optimal 3-point design, ξ^*, has support at the points $0, 0.2909$ and 1.6893, the efficiency $R(\xi^*, \theta)$ is again a minimum at the end-points of Θ and the candidate worst prior puts weights of 0.5940 and 0.4060 on the parameter values 5 and 10 respectively. A plot of $d(x, \xi^*)$ against x indicates that, in contrast to the case with $\Theta = [5, 6]$, the standardized maximin D-optimal 3-point design ξ^* is not globally optimal.

6 Conclusions

This study provides a cohesive approach to the construction of standardized maximin optimal designs for a broad range of nonlinear model settings. It is demonstrated that under fairly general conditions Bayesian Ψ_q-optimal designs converge to standardized maximin optimal designs. In Section 4 it is shown that these conditions hold for a number of design problems, including those involving nonlinear models, weighted polynomial models and nonlinear functions of the parameters of a linear model. Before the results can be implemented, Bayesian Ψ_q-optimal designs for the model, the optimality criterion and the class of designs of interest must necessarily be constructed and in the study of this paper the emphasis is on obtaining these designs explicitly. In some cases, as for example those in Section 4, such designs are available in the literature. In other modelling situations, new methods for the construction of

Bayesian Ψ_q-optimal designs must be devised, as is done for the weighted polynomial models of Section 5. Finally, in implementing Theorem 3.6, it is necessary to find the requisite standardized maximin optimal design as the limit of the appropriate Bayesian Ψ_q-optimal designs and, as indicated in Sections 4 and 5, the mathematics involved in deriving such a limiting design analytically can be somewhat intricate.

On the other hand for many nonlinear model settings it is possible that Bayesian Ψ_q-optimal designs cannot be obtained in an explicit algebraic form. In such cases these Bayesian optimal designs can usually be calculated numerically for a range of increasingly negative q values and the limiting and hence the standardized maximin optimal design identified, at least approximately. This strategy and its dependence on the choice of model, criterion and prior are clearly of interest and comprise an attractive area for further research.

A secondary but nevertheless important feature of the present study is the suite of results for convex classes of designs presented in Section 3.1 and based on the Equivalence Theorem, Theorem 3.1. These results provide considerable insight into the nature of standardized maximin optimal designs and their relation to the Bayesian Ψ_q-optimal designs and in addition provide tools for confirming the global optimality or otherwise of candidate designs. However it should immediately be emphasized that, while a standardized maximin optimal design is globally optimal provided it is Bayesian Ψ_0-optimal for some least favourable prior, the identification of that prior is not straightforward. The results of the present study provide further inside in the problem of constructing standardized maximin optimal designs (either analytically or numerically) and also, in cases where the class of designs is convex, for finding the associated least favourable priors.

Acknowledgements. The authors are grateful to Isolde Gottschlich who typed most parts of this paper with considerable technical expertise. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, Reduction for complexity in multivariate data structures), and of the University of Natal and the National Research Foundation of South Africa is gratefully acknowledged.

References

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

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer, New York.

Chaloner, K. (1989). Bayesian design for estimating the turning point of a quadratic regression. *Commun. Statist. Theor. & Meth.* **18**, 1385-1400.

Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear problems. *J. Statist. Plann. Inf.* **37**, 229-235.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inf.* **21**, 191-208.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science* **10**, 273-304.

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

Dette, H. (1994). Discrimination designs for polynomial regression on compact intervals. *Ann.*

Statist. **22**, 890-903.

Dette, H. (1997). Designing experiments with respect to 'standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59**, 97-110.

Dette, H., Haines, L.M. and Imhof, L. (1999). Optimal designs for rational models and weighted polynomial regression. *Ann. Statist.* **27**, 1272-1293.

Dette, H. and Sahm, M. (1998). Minimax optimal designs in nonlinear regression models. *Statist. Sinica* **8**, 1249-1264.

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

Dette, H. and Wong, W.K. (1996). Optimal Bayesian designs for models with partially specified heteroscedastic structure. *Ann. Statist.* **24**, 2108-2127.

Fedorov, V.V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

Ford, I. and Silvey, S.D. (1980). A sequentially constructed design for estimating a nonlinear parametric function. *Biometrika* **67**, 381-388.

Haines, L.M. (1995). A geometric approach to optimal design for one-parameter non-linear models. *J. Roy. Statist. Soc. Ser. B* **57**, 575-598.

Hewitt, E. and Stromberg, K. (1965). *Real and Abstract Analysis.* Springer, New York.

Imhof, L.A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Ann. Statist.* **29**, 561-576.

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

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

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

Müller, C.H. and Pázman, A. (1998). Applications of necessary and sufficient conditions for maximin efficient designs. *Metrika* **48**, 1-19.

Pronzato, L. and Walter, E. (1985). Robust experiment design via stochastic approximation. *Math. Biosci.* **75**, 103-120.

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

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

Spruill, M.C. (1987). Optimal designs for interpolation. *J. Statist. Plann. Inf.* **16**, 219-229.

Spruill, M.C. (1990). Good designs for polynomial extrapolation. *J. Statist. Plann. Inf.* **26**, 149-159.

Szegö, G. (1975). *Orthogonal Polynomials.* American Mathematical Society, Providence, Rhode Island.

Wong, W. K. (1992). A unified approach to the construction of minimax designs. *Biometrika* **79**, 611-619.

Wu, C.F.J. (1985). Asymptotic inference from sequential design in a nonlinear situation. *Biometrika* **72**, 553-558.

Proof of Lemma 3.3. Let ξ^* denote the standardized maximin optimal design and assume that $\mathcal{N}(\xi^*) = \{\theta_0\}$ is a singleton. Then the Equivalence Theorem 3.1 for standardized maximin optimality shows that ξ^* is locally D-optimal for the parameter θ_0 in the class Δ. Therefore

$$1 = \frac{\psi(M(\xi^*, \theta_0))}{\psi(M(\xi^*_{\theta_0}, \theta_0))} = \min\Big\{ \frac{\psi(M(\xi^*, \theta))}{\psi(M(\xi^*_\theta, \theta))} \;\Big|\; \theta \in \Theta \Big\} \;\le\; 1,$$

which implies $\mathcal{N}(\xi^*) = \Theta$ contradicting the hypothesis that $\#\mathcal{N}(\xi^*) = 1$. $\Box$

Proof of Theorem 3.6. Note first that the continuity of ψ implies that the normalizing function

$$\nu(\theta) := \psi(\xi^*_\theta, \theta)\ , \quad \theta \in \Theta,$$

is lower semicontinuous. For if $\theta \in \Theta$ and $\{\theta_j\}_{j=1}^\infty \subset \Theta$ is a sequence that converges to θ, then

$$\liminf_{j\to\infty} \nu(\theta_j) = \liminf_{j\to\infty} \psi(\xi^*_{\theta_j}, \theta_j) \ge \liminf_{j\to\infty} \psi(\xi^*_\theta, \theta_j) = \psi(\xi^*_\theta, \theta) = \nu(\theta).$$

Let $\epsilon > 0$, and let $\theta_0 \in \Theta$ be such that

$$\frac{\psi(\zeta^*, \theta_0)}{\nu(\theta_0)} \;\le\; \Psi_{-\infty}(\zeta^*) + \epsilon.$$

Then, since ν is lower semicontinuous and ψ is continuous, there is a relatively open neighborhood $U \subset \Theta$ of θ_0 such that

$$\frac{\psi(\zeta^*, \theta)}{\nu(\theta)} \le \Psi_{-\infty}(\zeta^*) + 2\epsilon \qquad \text{for all } \theta \in U.$$

As $\mathrm{supp}(\pi) = \Theta$, $\pi(U) > 0$. Since ζ_q converges weakly to ζ^*,

$$\frac{\psi(\zeta_q, \theta)}{\nu(\theta)} \;\to\; \frac{\psi(\zeta^*, \theta)}{\nu(\theta)}$$

for every $\theta \in \Theta$. It therefore follows from Egorov's theorem [see e.g. Hewitt and Stromberg (1965), page 158] that there exist a measurable set $V \subset \Theta$ with $\pi(V) > 1 - \frac{1}{2}\pi(U)$ and a number $q_0 < 0$ such that

$$\left| \frac{\psi(\zeta_q, \theta)}{\nu(\theta)} - \frac{\psi(\zeta^*, \theta)}{\nu(\theta)} \right| \le \epsilon \qquad \text{for all } \theta \in V \text{ and all } -\infty < q \le q_0.$$

Thus for all $-\infty < q \le q_0$,

$$\{\Psi_q(\zeta_q)\}^q \ge \int_{U\cap V} \left\{ \frac{\psi(\zeta_q, \theta)}{\nu(\theta)} \right\}^q d\pi(\theta) \ge \{\Psi_{-\infty}(\zeta^*) + 3\epsilon\}^q \, \pi(U \cap V).$$

Obviously, $\pi(U \cap V) > 0$, and it follows that

$$\limsup_{q\to-\infty} \Psi_q(\zeta_q) \le \{\Psi_{-\infty}(\zeta^*) + 3\epsilon\} \limsup_{q\to-\infty} \{\pi(U \cap V)\}^{\frac{1}{q}} = \Psi_{-\infty}(\zeta^*) + 3\epsilon.$$

As $\epsilon > 0$ was arbitrary, one has

$$\limsup_{q\to-\infty} \Psi_q(\zeta_q) \le \Psi_{-\infty}(\zeta^*).$$

Consequently, if $\xi \in \Delta$ is any competing design, then

$$\Psi_{-\infty}(\xi) = \lim_{q\to-\infty} \Psi_q(\xi) \le \limsup_{q\to-\infty} \Psi_q(\zeta_q) \le \Psi_{-\infty}(\zeta^*).$$

This proves that ζ^* is indeed a standardized maximin optimal design in the class Δ. □

Proof of Corollary 4.2. The assumption that $I(x,\theta)$ is continuous and bounded implies that for every fixed θ, the criterion $\psi(\xi,\theta) = \phi\{M(\xi,\theta)\}$ is continuous in ξ. The assumption also implies by Lebesgue's convergence theorem that for every ξ, $\psi(\xi,\theta)$ is continuous in θ. The assertion now follows from Theorem 3.6. □

Proof of Corollary 4.4. Under any one of the three conditions stated in the theorem, $f_i(x)f_j(x)\lambda(x,\theta)$ is bounded and continuous on $\mathcal{X} \times \Theta$ for $i,j = 0,\ldots,d$. The assertion now follows along the lines of the proof of Corollary 4.2. □

Proof of Lemma 4.6. Set

$$G_q(t) = t + F_q\{qh(t)\}, \quad t \in \Theta,$$

and $t_q = -F_q(qz_q)$. It follows from (4.4) that $G_q(t_q) = 0$. As $F_q'(x) \ge 0$ for all x, G_q is strictly increasing. Thus t_q is the only zero of the function G_q. It has to be shown that $\lim_{q\to-\infty} t_q = t^*$. Assume first that $t^* < \theta_{\max}$. Let $\varepsilon > 0$ be such that $t^* + \varepsilon \le \theta_{\max}$. Setting

$$\phi(\theta) = \exp\{\theta h(t^* + \varepsilon)\}g(\theta),$$

one has

$$F_q\{qh(t^* + \varepsilon)\} = -\frac{\int_\Theta \theta\phi(\theta)^{-q}\, d\pi(\theta)}{\int_\Theta \phi(\theta)^{-q}\, d\pi(\theta)}. \tag{A.1}$$

As $\theta h(t^* + \varepsilon) + \log g(\theta)$ is convex, so is $\phi(\theta)$. Since h is strictly decreasing, it follows from (4.5),

$$\begin{aligned}\frac{\phi(\theta_{\max})}{\phi(\theta_{\min})} &= \exp\{(\theta_{\max} - \theta_{\min})h(t^* + \varepsilon)\}\frac{g(\theta_{\max})}{g(\theta_{\min})} \\ &< \exp\{(\theta_{\max} - \theta_{\min})h(t^*)\}\frac{g(\theta_{\max})}{g(\theta_{\min})} = 1.\end{aligned}$$

Thus $\phi(\theta_{\max}) < \phi(\theta_{\min})$, and so $\phi(\theta) < \phi(\theta_{\min})$ for all $\theta > \theta_{\min}$. Consequently, for every $\theta_0 \in \mathrm{int}\Theta$,

$$\lim_{q\to-\infty} \frac{\left\{\int_{[\theta_0,\theta_{\max}]} \phi(\theta)^{-q}\, d\pi(\theta)\right\}^{-\frac{1}{q}}}{\left\{\int_\Theta \phi(\theta)^{-q}\, d\pi(\theta)\right\}^{-\frac{1}{q}}} = \frac{\max_{\theta\in[\theta_0,\theta_{\max}]}\phi(\theta)}{\max_{\theta\in\Theta}\phi(\theta)} < 1,$$

and so

$$\lim_{q\to-\infty}\frac{\int_{[\theta_0,\theta_{\max}]}\phi(\theta)^{-q}\ d\pi(\theta)}{\int_\Theta \phi(\theta)^{-q}\ d\pi(\theta)}=0,\qquad \lim_{q\to-\infty}\frac{\int_{[\theta_{\min},\theta_0)}\phi(\theta)^{-q}\ d\pi(\theta)}{\int_\Theta \phi(\theta)^{-q}\ d\pi(\theta)}=1.$$

In view of (A.1),

$$\begin{aligned} G_q(t^*+\varepsilon) &= t^*+\varepsilon-\frac{\int_\Theta \theta\phi(\theta)^{-q}\ d\pi(\theta)}{\int_\Theta \phi(\theta)^{-q}\ d\pi(\theta)} \\ &\geq t^*+\varepsilon-\frac{\theta_0\int_{[\theta_{\min},\theta_0)}\phi(\theta)^{-q}\ d\pi(\theta)+\theta_{\max}\int_{[\theta_0,\theta_{\max}]}\phi(\theta)^{-q}\ d\pi(\theta)}{\int_\Theta \phi(\theta)^{-q}\ d\pi(\theta)}. \end{aligned}$$

It follows that

$$\liminf_{q\to-\infty} G_q(t^*+\varepsilon)\geq t^*+\varepsilon-\theta_0$$

for all $\theta_0\in \mathrm{int}\Theta$. Thus

$$\liminf_{q\to-\infty} G_q(t^*+\varepsilon)\geq \varepsilon,$$

so that $G_q(t^*+\varepsilon)>0$ for $q\leq q_0=q_0(\varepsilon)$, say. Since G_q is increasing, this implies that $t_q<t^*+\varepsilon$ for $q\leq q_0$. As $\varepsilon>0$ was arbitrarily small, $\limsup_{q\to-\infty} t_q\leq t^*$, which is trivially true if $t^*=\theta_{\max}$. A similar argument shows that $\liminf_{q\to-\infty} t_q\geq t^*$, completing the proof of Lemma 4.6. □

Proof of Corollary 4.10. Let ζ_q be the Ψ_q-optimal designs in the class Δ, so that as $q\to-\infty$, ζ_q converges weakly to a non-singular design ζ^*. Then $\lim_{q\to-\infty} M(\zeta_q)=M(\zeta^*)$, in particular $M(\zeta_q)$ is non-singular for $q\leq q_0$, say. Hence $\psi(\zeta_q,\theta)$ converges to $\psi(\zeta^*,\theta)$ for each θ. Thus for $q\leq q_0$, $\psi(\zeta_q,\theta)/\psi(\xi^*_\theta,\theta)$ is continuous and for $q\to-\infty$, converges to $\psi(\zeta^*,\theta)/\psi(\xi^*_\theta,\theta)$. An inspection of the proof of Theorem 3.6 shows that this is sufficient to ensure that ζ^* is a standardized maximin optimal design. □

Proof Lemma 5.1. It was shown in Theorem 3.1 of Dette, Haines and Imhof (1999) that for $\theta>d$ the locally D-optimal design has equal masses at the roots $-1<x_1<\ldots<x_{d+1}<1$ of the polynomial

$$C_{d+1}^{(-\theta-1/2)}(\sqrt{-x^2}).$$

For a proof of the representation of $\prod_{\ell=1}^{d+1}(1+x_\ell^2)$ put $\lambda=-\theta-1/2$ and note that

$$C_{d+1}^{(\lambda)}(\sqrt{-x^2})=C_{d+1}^{(\lambda)}(ix)=c_{d+1}\prod_{\ell=1}^{d+1}(x-x_\ell)\ , \tag{A.2}$$

where c_{d+1} denotes the leading coefficient of the ultraspherical polynomial, i.e.

$$c_{d+1}=(2i)^{d+1}\binom{d+\lambda}{d+1} \tag{A.3}$$

[see e.g. Szegö (1975), formula (4.7.9)]. Therefore the identity (A.2) implies

$$\prod_{\ell=1}^{d+1}(1+x_\ell^2)=\prod_{\ell=1}^{d+1}(i-x_\ell)(-i-x_\ell)=\frac{C_{d+1}^{(\lambda)}(1)C_{d+1}^{(\lambda)}(-1)}{c_{d+1}^2}=(-1)^{d+1}\frac{\left\{C_{d+1}^{(\lambda)}(1)\right\}^2}{c_{d+1}^2},$$

where the last identity follows from the symmetry of ultraspherical polynomials [see Szegö (1975), formula (4.7.4)]. From formula (4.7.3) in the same reference and (A.3) it therefore follows that

$$\prod_{\ell=1}^{d+1}(1+x_\ell^2) = \Big\{\frac{1}{2^{d+1}}\prod_{j=1}^{d+1}\frac{j}{d+1+\lambda-j}\cdot\frac{d+2\lambda+1-j}{j}\Big\}^2 \quad (A.4)$$

$$= \Big\{\prod_{j=1}^{d}\frac{d-2\theta-j}{2d+1-2\theta-2j}\Big\}^2,$$

which proves the first equation asserted in Lemma 5.1.

For a proof of the second one note that

$$|M(\xi_\theta,\theta)| = \Big(\frac{1}{d+1}\Big)^{d+1}\prod_{\ell=1}^{d+1}(1+x_i^2)^{-\theta}\prod_{1\le\ell<k\le d+1}(x_\ell-x_k)^2$$

and it is therefore sufficient to calculate the value of the last factor. To this end again put $\lambda=-\theta-1/2$ and define

$$P_{d+1}(x)=\frac{C_{d+1}^{(\lambda)}(ix)}{c_{d+1}i^{d+1}}$$

as the ultraspherical polynomial with parameter λ, argument ix and leading coefficient 1. Then it follows by a straightforward calculation

$$\prod_{1\le\ell<k\le d+1}(x_\ell-x_k)^2 = (-1)^{d(d+1)/2}\prod_{\ell=1}^{d+1}\prod_{k\neq\ell}(x_\ell-x_k) = (-1)^{d(d+1)/2}\prod_{\ell=1}^{d+1}P'_{d+1}(x_\ell).$$

From formula (4.7.27) in Szegö (1975) we obtain that for any $\ell\in\{1,\dots,d+1\}$

$$(1+x^2)\frac{d}{dx}P_{d+1}(x)\Big|_{x=x_\ell} = \frac{1+x^2}{i^{d+1}c_{d+1}}\frac{d}{dx}C_{d+1}^{(\lambda)}(ix)\Big|_{x=x_\ell}$$

$$= \frac{i(d+2\lambda)}{i^d c_{d+1}}C_d^{(\lambda)}(ix_\ell) = \frac{(d+2\lambda)}{c_{d+1}}c_d P_d(x_\ell) = \frac{d+1}{2}\frac{d+2\lambda}{d+\lambda}P_d(x_\ell),$$

where the notation (A.3) is used in the last equality. Observing that the recursive relation for the polynomial $P_j(x)$ is given by

$$P_{k+1}(x) = xP_k(x) + \frac{(k-1+2\lambda)k}{4(k+\lambda-1)(k+\lambda)}P_{k-1}(x)$$

($P_{-1}(x)=0, P_0(x)=1$) it now follows from formula (6.71.2) in Szegö (1975) that [$a_n=1, c_n=$

$-(n-1)(n-2+2\lambda)/4(n-2+\lambda)(n-1+\lambda)]$

$$\prod_{\ell=1}^{d+1}(1+x_\ell^2)\prod_{1\le\ell<k\le d+1}(x_\ell-x_k)^2 = (-1)^{d(d+1)/2}\prod_{\ell=1}^{d+1}\Big(\frac{d+1}{2}\Big)\Big(\frac{d+2\lambda}{d+\lambda}\Big)P_d(x_\ell)$$

$$= (-1)^{d(d+1)/2}\Big(\frac{d+1}{2}\Big)^{d+1}\Big(\frac{d+2\lambda}{d+\lambda}\Big)^{d+1}\prod_{j=1}^{d+1}\Big\{\frac{(j-1)(j-2+2\lambda)}{4(j-2+\lambda)(j-1+\lambda)}\Big\}^{j-1}$$

$$= \prod_{j=1}^{d+1} j^j\cdot\prod_{j=1}^{d}\frac{(2\theta-j+1)^{j+1}}{(2\theta-2j+1)^{2j+1}}.$$

Combining this identity with (A.4) yields

$$\prod_{\ell=1}^{d+1}(1+x_\ell^2)^{-\theta}\prod_{1\le\ell<k\le d+1}(x_\ell-x_k)^2=\prod_{j=1}^{d+1} j^j\cdot\prod_{j=1}^{d}\frac{(2\theta-2j+1)^{2\theta-2j+1}}{(2\theta-j+1)^{2\theta-j+1}}$$

and the second equation asserted in Lemma 5.1 follows. □

Proof of Lemma 5.2. It follows by similar arguments to those given in Dette, Haines and Imhof (1999) that the locally D-optimal design is supported at $d+1$ points including the point 0, say $0=x_1<x_2<\ldots<x_{d+1}$, and that the supporting polynomial $f(x)=\Pi_{i=1}^{d+1}(x-x_i)$ is a solution of the differential equation

$$x(1+x)y''(x)-\theta xy'(x)+(d+1)(\theta-d)y(x)=0.$$

The polynomial solution of this equation is given by the hypergeometric series

$$xF(-d,d+1-\theta,2,-x)\ ,$$

which is proportional to the Jacobi polynomial

$$f(x)=xP_d^{(1,-\theta-1)}(2x+1)$$

[see formula (4.21.2) in Szegö (1975)]. The remaining assertions of Lemma 5.2 now follow by similar arguments to those given in the proof of Lemma 5.1 but are omitted for the sake of brevity. □

Proof of Theorem 5.3. The determinant of the information matrix for the parameters β of model (5.1) from a $(d+1)$-point design ξ which puts equal masses at the support points $x_1,\ldots,x_{d+1}$ can be written as

$$\mid M(\theta,\xi)\mid \;=\; \Big(\frac{1}{d+1}\Big)^{d+1}\mid X_R\mid^2\prod_{i=1}^{d+1}(1+x_i^2)^{-\theta}$$

where X_R is the Vandermonde matrix with ith row $\{1\ x_i\ x_i^2\ \ldots\ x_i^d\}$ $(i=1,\ldots,d+1)$. Thus observing the definition of $\tilde\pi$ in (5.2)

$$(\Psi_{q(d+1)}(\xi))^{d+1} \;=\; \frac{1}{(d+1)^{d+1}}\mid X_R\mid^2\Big\{\int_\Theta\prod_{i=1}^{d+1}(1+x_i^2)^{-q\theta}d\tilde\pi(\theta)\Big\}^{1/q}$$

and differentiating $\log \Psi_{q(d+1)}(\xi)$ with respect to x_j and setting the result to 0 in turn gives

$$\frac{2}{\mid X_R \mid} \frac{\partial \mid X_R \mid}{\partial x_j} - \frac{2x_j}{1+x_j^2} \frac{\displaystyle\int_\Theta \theta \prod_{i=1}^{d+1} (1+x_i^2)^{-q\theta} d\tilde{\pi}(\theta)}{\displaystyle\int_\Theta \prod_{i=1}^{d+1} (1+x_i^2)^{-q\theta} d\tilde{\pi}(\theta)} = 0.$$

It then follows by arguments similar to those used in Dette, Haines and Imhof (1999) that the required points are the roots of the polynomial $\prod_{j=1}^{d+1}(x-x_j)$, which satisfies the differential equation

$$(1+x^2)f''(x) + 2xF(qz)f'(x) - (d+1)(d+2F(qz))f(x) = 0,$$

where $f(x)$ is a polynomial of degree $d+1$ in x, $z = \sum_{i=1}^{d+1} \log(1+x_i^2)$ and

$$F(qz) = -\frac{\displaystyle\int_\Theta \theta e^{-\theta q z} d\tilde{\pi}(\theta)}{\displaystyle\int_\Theta e^{-\theta q z} d\tilde{\pi}(\theta)}$$

with $-F(qz) > d$. The support points of the Bayesian $\Psi_{q(d+1)}$-optimal design are thus the roots of the ultraspherical polynomial

$$C_{d+1}^{(F(qz)-\frac{1}{2})}(\sqrt{-x^2})$$

and $z = \sum_{i=1}^{d+1} \log(1+x_i^2)$ is obtained by invoking expression for $\prod_{i=1}^{d+1}(1+x_i^2)$ given in Lemma 5.1. Note that for $q=0$ the criterion corresponds to Bayesian D-optimality and that for a one point prior the locally D-optimal design is recovered. □

Proof of Theorem 5.4. This follows along the lines given in the proof of Theorem 5.3. □

Proof of Theorem 5.7 and 5.8. Both theorems can be proved by similar arguments and attention is restricted to the proof of Theorem 5.7. For $j = 0, \ldots, d$,

$$\lim_{x \to \pm\infty} x^j \max_{\theta \in \Theta} \sqrt{\lambda(x,\theta)} = \lim_{x \to \pm\infty} \frac{x^j}{(1+x^2)^{\frac{1}{2}\theta_{\min}}} = 0.$$

Thus condition (iii) of Corollary 4.4 is satisfied and the maximin design can therefore be obtained as the limit of Ψ_q-optimal designs. Suppose d is even; the case where d is odd is similar. Let π be any prior distribution with support Θ and write

$$F_q(x) = -\frac{\int \theta e^{-\theta x} m(\theta)^{-q}\, d\pi(\theta)}{\int e^{-\theta x} m(\theta)^{-q}\, d\pi(\theta)}, \qquad q < 0,\ x \in \mathbb{R},$$

$$h(\theta) = 2 \sum_{j=1}^{\frac{d}{2}} \log \frac{2\theta - 2j + 2}{2\theta - d - 2j + 1} \ = \ 2 \sum_{j=1}^{d} \log \frac{d - 2\theta - j}{2d + 1 - 2\theta - 2j}, \qquad \theta \in \Theta.$$

If z_q denotes the solution of the equation $z_q = h(-F_q(qz_q))$, then, by Theorem 5.3, the Bayesian $\Psi_{q(d+1)}$-optimal $(d+1)$-point design puts equal weights on the roots of the polynomial

$$C_{d+1}^{\left(F_q(qz_q)-\frac{1}{2}\right)}(\sqrt{-x^2}).$$

It remains to show that $\lim_{q\to-\infty} F_q(qz_q) = -\theta_0$. To see that there indeed exists $\theta_0 \in \text{int}\Theta$ as defined in the theorem, set

$$H(x,\theta) = \sum_{j=1}^{\frac{d}{2}} \log \frac{(2x-d-2j+1)^{2\theta-d-2j+1}}{(2x-2j+2)^{2\theta-2j+2}}, \qquad x,\theta \in \Theta.$$

Then

$$\frac{\partial H(x,\theta)}{\partial x} = 2(d+1)\sum_{j=1}^{\frac{d}{2}} \frac{\theta-x}{(2x-d-2j+1)(x-j+1)},$$

so that $H(\cdot,\theta_{\max})$ is strictly increasing and $H(\cdot,\theta_{\min})$ is strictly decreasing on Θ. Hence, in view of Lemma 5.1,

$$\begin{aligned}\frac{-\log\{m(\theta_{\max})/m(\theta_{\min})\}}{\theta_{\max}-\theta_{\min}} &= \frac{H(\theta_{\min},\theta_{\min})-H(\theta_{\max},\theta_{\max})}{\theta_{\max}-\theta_{\min}} \\ &< \frac{H(\theta_{\min},\theta_{\min})-H(\theta_{\min},\theta_{\max})}{\theta_{\max}-\theta_{\min}} = h(\theta_{\min})\end{aligned}$$

and, similarly,

$$\frac{-\log\{m(\theta_{\max})/m(\theta_{\min})\}}{\theta_{\max}-\theta_{\min}} > h(\theta_{\max}).$$

This ensures the existence of θ_0. It is easily verified that $h(t)$ is strictly decreasing and that $\log m(\theta)$ is convex. It now follows by Lemma 4.6 that $\lim_{q\to-\infty} F_q(qz_q) = -\theta_0$, which completes the proof of Theorem 5.7. □