The Annals of Statistics
1996, Vol. 24, No. 3, 1225–1234

A NOTE ON BAYESIAN *c*- AND *D*-OPTIMAL DESIGNS IN NONLINEAR REGRESSION MODELS[1]

By Holger Dette

Technische Universität Dresden

We present a version of Elfving's theorem for the Bayesian D-optimality criterion in nonlinear regression models. The Bayesian optimal design can be characterized as a design which allows a representation of a (uniquely determined) boundary point of a convex subset of L^2-integrable functions. A similar characterization is given for the Bayesian c-optimality criterion where a (possible) nonlinear function of the unknown parameters has to be estimated. The results are illustrated in the example of an exponential growth model using a gamma prior distribution.

1. Introduction. This paper is intended to serve as an addendum to a recent paper of Dette (1993). In that paper a geometric characterization for D-optimal designs in linear regression models is presented generalizing the famous result of Elfving (1952) for c-optimal designs. In nonlinear models the information matrix of an experimental design usually depends on the unknown parameter, say ϑ, and optimal designs cannot, owing to that dependence, be determined in practice. Various optimality criteria have been proposed in the literature in order to overcome the dependency of the optimality criterion on the unknown parameters. The most popular approaches are local optimality criteria and Bayesian optimality criteria. For the determination of a locally optimal design a best guess of ϑ, say ϑ_0, is needed, and a function of the information matrix evaluated at ϑ_0 has to be maximized [see, e.g., Chernoff (1953)]. Bayesian optimal designs maximize the expectation (with respect to a prior distribution) of some function of the information matrix, where the function approximates some utility function [see, e.g., Zacks (1977), Pronzato and Walter (1985), Chaloner (1987, 1989, 1993) and Chaloner and Larntz (1989, 1992)]. In this paper we are mainly interested in the Bayesian D-optimality criterion, where the function of the information matrix is the logarithm of its determinant. In Section 2 we present a geometric characterization of Elfving type for the optimal designs with respect to this criterion. The analysis is based on a convex subset of L^2-integrable functions (with respect to the prior distribution) and can be seen as an extension of recent results in Dette (1993) for optimal, model-robust designs in linear regression models. In Section 3 we give a similar characterization for a Bayesian c-optimality criterion where a (possible) nonlinear

Received March 1993; revised July 1995.
[1]Research supported in part by the Deutsche Forschungsgemeinschaft.
AMS 1991 *subject classification*. 62K05.
Key words and phrases. Nonlinear regression, Bayesian D-optimal designs, Elfving's theorem, geometric characterization.

function c of the unknown parameter has to be estimated. The famous geometric characterization of Elfving (1952) and the Elfving theorem for quadratic loss in Studden (1971) appear as special cases. Finally, some examples illustrating the results are given in Section 4.

2. Bayesian *D*-optimality. Consider the nonlinear regression model

$$Y(x) = \eta(x, \vartheta) + \varepsilon(x),$$

where $x \in \mathscr{X}$ is the explanatory variable, the design space $\mathscr{X} \subseteq \mathbb{R}^s$ is compact and $\vartheta = (\vartheta_1, \dots, \vartheta_k)' \in \Theta$ is the vector of unknown parameters. The parameter space $\Theta \subseteq \mathbb{R}^k$ is an open set and for every $x \in \mathscr{X}$, $\varepsilon(x)$ is a normally distributed random variable with mean 0 and variance $\sigma^2 > 0$ such that $\varepsilon(x)$ and $\varepsilon(y)$ are independent whenever $x \neq y$. The regression function $\eta(x, \vartheta)$ is assumed to be differentiable with respect to ϑ (for any fixed $x \in \mathscr{X}$) and the partial derivatives $\partial\eta(x, \vartheta)/\partial\vartheta_j$ are supposed to be continuous on $\mathscr{X}$ (for any fixed $\vartheta \in \Theta$). A (approximate) design ξ is a probability measure on the design space $\mathscr{X}$ and the information matrix of ξ is defined by

$$M_k(\xi, \vartheta) = \int_{\mathscr{X}} f_k(x, \vartheta) f_k(x, \vartheta)' \, d\xi(x) \subseteq \mathbb{R}^{k \times k}, \tag{2.1}$$

where

$$f_l(x, \vartheta)' = \left(\frac{\partial\eta(x, \vartheta)}{\partial\vartheta_1}, \dots, \frac{\partial\eta(x, \vartheta)}{\partial\vartheta_l} \right) \in \mathbb{R}^l, \qquad l = 1, \dots, k, \tag{2.2}$$

is the vector of the first l components of the gradient of $\eta(x, \vartheta)$ with respect to ϑ. An exact design ξ for the sample size n is a probability measure with finite support $x_1, \dots, x_l$ and masses $n_1/n, \dots, n_l/n$ which means that the experimenter takes n_j uncorrelated observations at each x_j, $j = 1, \dots, l$. In this case the inverse of the information matrix is proportional to the asymptotic covariance matrix of the maximum likelihood estimator for the parameter vector ϑ [see, e.g., Silvey (1980), page 3]. In practice, efficient exact designs can be found from optimal approximate designs by the use of an appropriate rounding procedure [see Kiefer (1971)].

Throughout this paper we will assume that the partial derivatives $\partial\eta(x, \vartheta)/\partial\vartheta_j$ are L^2-integrable with respect to a prior distribution μ on Θ. A design ξ^* is called Bayesian D-optimal with respect to the prior μ if ξ^* maximizes

$$S(\xi) := E_\mu\left[\log\left(\det(M_k(\xi, \vartheta))\right)\right] = \int_\Theta \log\left(\det(M_k(\xi, \vartheta))\right) d\mu(\vartheta) \tag{2.3}$$

among all designs for which $|S(\xi)| < \infty$. This criterion has been motivated by Bayesian arguments in Chaloner and Larntz (1989) and has been applied for a couple of models in Chaloner (1987, 1993) and Dette and Neugebauer (1996). The optimality criterion (2.3) appears also in the context of model-robust optimality criteria for linear regression [see Läuter (1974a, b)]. More precisely, assume that for every $\vartheta \in \Theta$ we have a linear model

$$Y = \alpha'_\vartheta f_k(x, \vartheta) + \varepsilon(x),$$

where α'_ϑ and $f(x, \vartheta)$ denote the vector of parameters and regression functions in the model with index $\vartheta \in \Theta$ and ε is a normally distributed error term. Consider the class of linear regression models

$$\mathscr{F}_\Theta = \left\{ \alpha'_\vartheta f_k(x, \vartheta) \middle| \vartheta \in \Theta,\ \alpha_\vartheta \in \mathbb{R}^k \right\}. \tag{2.4}$$

Then for fixed $\vartheta \in \Theta$ the information matrix of a design ξ in the linear model $\alpha'_\vartheta f_k(x, \vartheta)$ with index ϑ is precisely (2.1). For the determination of a design that allows efficient estimates of the parameters in all models of the class $\mathscr{F}_\Theta$, Läuter (1974a) proposed to maximize the function in (2.3) [note that $M_k^{-1}(\xi, \vartheta)$ is proportional to the covariance matrix of the least squares estimator for the parameter vector α_ϑ in the model $\alpha'_\vartheta f_k(x, \vartheta)$]. In the case that the set of models $\mathscr{F}_\Theta$ in (2.4) consists only of one model (or equivalently that the support of the prior distribution μ contains only one point), the criterion (2.3) reduces to the well-known D-optimality criterion. In the usual linear regression model, Dette (1993) proved a geometric characterization for the D-optimal design which generalizes the famous theorem of Elfving (1952) for the c-optimality criterion. By the preceding discussion we see that the D-optimality criterion in the common linear regression model can be seen as a special case of (2.3) (i.e., $\operatorname{supp}(\mu) = \{\vartheta_0\}$) and it is therefore reasonable to expect a similar geometric characterization for Bayesian D-optimal designs which will be discussed in the following.

In order to establish results of this type, we will need the following notation and assumptions. Let L^2_μ denote the set of all quadratic integrable (with respect to the prior distribution μ), real-valued functions and define, for $l \in \mathbb{N}$,

$$L^2_\mu(l) := \left\{ g \colon \Theta \to \mathbb{R}^l \middle| g(\vartheta) = (g_1(\vartheta), \dots, g_l(\vartheta))',\ g_j \in L^2_\mu,\ j = 1, \dots, l \right\}$$

as the set of all $\mathbb{R}^l$-valued functions with quadratic integrable components. On $L^2_\mu(l)$ we consider the usual inner product

$$\langle f, g \rangle_l := \sum_{j=1}^{l} \int_\Theta f_j(\vartheta) g_j(\vartheta)\, d\mu(\vartheta), \qquad f, g \in L^2_\mu(l). \tag{2.5}$$

Let

$$\Xi := \left\{ \xi \middle| |S(\xi)| < \infty, \det(M_k(\xi, \vartheta)) > 0\ \mu\text{-a.e.} \right\}$$

denote the set of probability measures with nonsingular information matrix (μ-a.e.) for which (2.3) is finite. Then the maximum in (2.3) is obviously attained in Ξ. Let $c_l = (0, \dots, 0, 1)' \in \mathbb{R}^l$, $l = 1, \dots, k$, and define for $x \in \mathscr{X}$ and $\xi \in \Xi$ the functions

$$\gamma_l^{\varepsilon}: \begin{cases} \Theta \to \mathbb{R}, \\ \vartheta \to \gamma_l^{\varepsilon}(\vartheta) := \left(c_l' M_l^{-1}(\xi, \vartheta) c_l\right)^{-1/2}, \end{cases} \qquad l = 1, \dots, k, \tag{2.6}$$

$$h_l^{\varepsilon}: \begin{cases} \Theta \to \mathbb{R}^l, \\ \vartheta \to h_l^{\varepsilon}(\vartheta) := \gamma_l(\vartheta) M_l^{-1}(\xi, \vartheta) c_l, \end{cases} \qquad l = 1, \dots, k, \tag{2.7}$$

$$\varepsilon_l^{\xi}(x): \begin{cases} \Theta \to \mathbb{R}, \\ \vartheta \to \varepsilon_l^{\xi}(x, \vartheta) := f_l(x, \vartheta)' h_l(\vartheta), \end{cases} \qquad l = 1, \dots, k, \tag{2.8}$$

and

$$d_{\xi}(x): \begin{cases} \Theta \to \mathbb{R}, \\ \vartheta \to d_{\xi}(x, \vartheta) := f_k(x, \vartheta)' M_k^{-1}(\xi, \vartheta) f_k(x, \vartheta). \end{cases}$$

Finally, let $\varepsilon_l \in L_{\mu}^2$, $l = 1, \dots, k$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_k)' \in L_{\mu}^2(k)$ and define a generalized Elfving set by

$$\mathscr{R}_{\Theta}^{D} := \operatorname{conv}\left(\left\{ g \in L_{\mu}^2\left(\frac{k(k+1)}{2}\right) \middle| g(\vartheta) = \begin{pmatrix} \varepsilon_1(\vartheta) f_1(x, \vartheta) \\ \vdots \\ \varepsilon_k(\vartheta) f_k(x, \vartheta) \end{pmatrix}, \right.\right.$$
$$\left.\left. x \in \mathscr{X}, \ \varepsilon \in L_{\mu}^2(k), \ \langle \varepsilon, \varepsilon \rangle_k = k \right\}\right),$$

where $\operatorname{conv}(\mathscr{A})$ denotes the convex hull of a set $\mathscr{A} \subseteq L_{\mu}^2(k(k+1)/2)$. In the following we will assume that the set

$$\Xi^* := \left\{ \xi \in \Xi \middle| \sup_{x \in \mathscr{X}} \int_{\Theta} d_{\xi}(x, \vartheta)\, d\mu(\vartheta) < \infty, \ \gamma_l^{\varepsilon} \in L_{\mu}^2, \ h_l^{\varepsilon} \in L_{\mu}^2(l), \ l = 1, \dots, k \right\}$$

is not empty. The following theorem gives a geometric characterization of the Bayesian D-optimal design problem. The proof can be obtained by combining the reasoning in Dette (1993) with the equivalence theorem in Läuter (1974a) [applied to the set of models $\mathscr{F}_{\Theta}$ defined in (2.4)] and is omitted for the sake of brevity.

THEOREM 2.1. *A design $\xi \in \Xi^*$ is Bayesian D-optimal if and only if there exist positive functions $\gamma_l \in L_{\mu}^2$ and for all $x \in \operatorname{supp}(\xi)$ real-valued functions $\varepsilon_l(x, \cdot) \in L_{\mu}^2$, $l = 1, \dots, k$, such that the following four conditions are satisfied*:

(a) $\gamma_l(\vartheta) c_l = \int_{\mathscr{X}} \varepsilon_l(x, \vartheta) f_l(x, \vartheta)\, d\xi(x) \quad \mu\text{-}a.e. \qquad l = 1, \dots, k.$

(b) *The function*

$$\gamma : \begin{cases} \Theta \to \mathbb{R}^{k(k+1)/2}, \\ \vartheta \to \gamma(\vartheta) := (\gamma_1(\vartheta), 0, \gamma_2(\vartheta), 0, \ldots, 0, \gamma_k(\vartheta))' \end{cases}$$

is a boundary point of the Bayesian Elfving set $\mathscr{R}_\Theta^D$ *with "supporting hyperplane"*

$$h = \frac{1}{k}(h_1, \ldots, h_k)' \in L_\mu^2\left(\frac{k(k+1)}{2}\right), \qquad h_l \in L_\mu^2(l), \, l = 1, \ldots, k.$$

(c) $\gamma_l(\vartheta) c_l' h_l(\vartheta) = 1 \quad \mu\text{-}a.e. \qquad l = 1, \ldots, k.$

(d) *The function* $\varepsilon(x, \cdot) = (\varepsilon_1(x, \cdot), \ldots, \varepsilon_k(x, \cdot))' \in L_\mu^2(k)$ *satisfies*

$$\langle \varepsilon(x, \cdot), \varepsilon(x, \cdot) \rangle_k = k \quad \textit{for all } x \in \operatorname{supp}(\xi).$$

Moreover the functions γ_l, h_l *and* ε_l *are* (μ-*a.e.*) *uniquely determined by* (2.6), (2.7) *and* (2.8), *respectively.*

REMARK 2.2. Theorem 2.1 gives some more insight into the complicated structure of the Bayesian D-optimal design problem. The maximization of the function in (2.3) is equivalent to the determination of a supporting hyperplane to a convex subset in $L_\mu^2(k(k+1)/2)$ at a specific boundary point of the Bayesian Elfving set $\mathscr{R}_\Theta^D$. This is usually a very hard problem and can only be done explicitly in special cases. In general, numerical methods have to be applied for the determination of a Bayesian D-optimal design [see, e.g., Chaloner and Larntz (1989, 1992)]. However, Theorem 2.1 turns out to be useful for proving or disproving if a given design is Bayesian D-optimal (see the examples in Section 4).

REMARK 2.3. If $f_k(x, \vartheta)$ does not depend on ϑ [e.g., if $\eta(x, \vartheta)$ is a linear model], it follows from (2.6), (2.7) and (2.8) that the functions γ_l^ξ, ε_l^ξ and h_l^ξ are constant in ϑ. It is then easy to see that conditions (a) through (d) in Theorem 2.1 are also independent of ϑ and Theorem 2.1 reduces to a similar statement as given in Dette (1993).

3. Bayesian *c*-optimality. Throughout this section let $c \in L_\mu^2(k)$ denote a function with quadratic integrable components. A design ξ with $c(\vartheta) \in \operatorname{range}(M(\xi, \vartheta))$ for all $\vartheta \in \Theta$ is called Bayesian c-optimal if ξ minimizes

$$\begin{aligned} C(\xi) &:= E_\mu[c(\vartheta)' M_k^-(\xi, \vartheta) c(\vartheta)] \\ &= \int_\Theta c(\vartheta)' M_k^-(\xi, \vartheta) c(\vartheta) \, d\mu(\vartheta) \end{aligned} \tag{3.1}$$

among all designs satisfying $C(\xi) < \infty$ [here $M_k^-(\xi, \vartheta)$ denotes an arbitrary generalized inverse of $M_k(\xi, \vartheta)$]. The criterion (3.1) could be used if the experimenter is interested in a specific real-valued function, say $b(\vartheta)$, of the unknown parameters $\vartheta \in \Theta$ in the model $y = \eta(x, \vartheta)$. For the choice $c(\vartheta) =$

$(\partial/\partial\vartheta)b(\vartheta)$ a Bayesian c-optimal design minimizes an average (with respect to μ) of the asymptotic variance of the maximum likelihood estimator of $b(\vartheta)$ [see, e.g., Silvey (1980), page 4]. This criterion was applied by Chaloner (1989) to determine optimal designs for the estimation of the turning point of a quadratic regression. In the case of a prior distribution with finite support, (3.1) reduces to the well-known A-optimality criterion for which an Elfving theorem was proved by Studden (1971).

In the following define for a design ξ with $c(\vartheta) \in \operatorname{range}(M_k(\xi, \vartheta))$ (μ-a.e.) and $C(\xi) < \infty$,

$$\gamma^{\xi} = \left\{E_{\mu}\left[c(\vartheta)' M_k^{-}(\xi, \vartheta) c(\vartheta)\right]\right\}^{-1/2} \tag{3.2}$$

for $x \in \operatorname{supp}(\xi)$ functions $h^{\xi}\colon \Theta \to \mathbb{R}^k$, $\varepsilon^{\xi}(x)\colon \Theta \to \mathbb{R}$ by

$$h^{\xi}(\vartheta) = \gamma^{\xi} M_k^{-}(\xi, \vartheta) c(\vartheta), \qquad \vartheta \in \Theta, \tag{3.3}$$

$$\varepsilon^{\xi}(x, \vartheta) = h(\vartheta)' f_k(x, \vartheta), \qquad \vartheta \in \Theta, \tag{3.4}$$

and a set of designs by

$$\begin{aligned} \Xi^{**} = \Big\{\xi \Big| c(\vartheta) \in \operatorname{range}(M_k(\xi, \vartheta))\ \forall\, \vartheta \in \Theta,\ h^{\xi} \in L^2_{\mu}(k), \\ \varepsilon^{\xi}(x) \in L^2_{\mu}\ \forall\ x \in \operatorname{supp}(\xi)\Big\}. \end{aligned}$$

For Bayesian c-optimality the Elfving set

$$\begin{aligned} \mathscr{R}^{c}_{\Theta} := \operatorname{conv}\Big(\Big\{g \in L^2_{\mu}(k) \Big| g(\vartheta) = \varepsilon(\vartheta) f_k(x, \vartheta),\ x \in \mathscr{X}, \\ \varepsilon \in L^2_{\mu},\ \langle \varepsilon, \varepsilon\rangle_1 = 1\Big\}\Big) \end{aligned} \tag{3.5}$$

turns out to be useful for a geometric characterization of Bayesian c-optimal designs. The proof of the following theorem can be obtained either by a similar reasoning as given in Studden (1971) or by an application of an equivalence theorem for Bayesian c-optimality and is therefore omitted.

THEOREM 3.1. *A design $\xi^* \in \Xi^{**}$ is Bayesian c-optimal if and only if there exists a constant $\gamma > 0$ and for all $x \in \operatorname{supp}(\xi^*)$ real-valued functions $\varepsilon(x, \cdot) \in L^2_{\mu}$ with $\int \varepsilon^2(x, \vartheta)\, d\mu(\vartheta) = 1$ such that the function $\gamma c(\cdot)$ has the representation*

$$\gamma c(\vartheta) = \int_{\mathscr{X}} f_k(x, \vartheta) \varepsilon(x, \vartheta)\, d\xi^*(x) \quad \mu\text{-a.e.} \tag{3.6}$$

and is a boundary point of the Elfving set $\mathscr{R}^{c}_{\Theta}$ defined in (3.5). *Moreover, the constant γ is uniquely determined by* (3.2) *and the function $\varepsilon(x, \cdot)$ and the supporting hyperplane h at $\gamma c \in \partial \mathscr{R}^{c}_{\Theta}$ are uniquely determined by* (3.3) *and* (3.4) *at all points $\vartheta \in \operatorname{supp}(\Theta)$ for which* $\det(M_k(\xi, \vartheta)) > 0$.

REMARK 3.2. If $\#\operatorname{supp}(\mu) = 1$, then Theorem 3.1 gives the classical Elfving theorem for locally optimal designs [see Elfving (1952)], and a couple of examples for the geometric construction of optimal designs can be found in a recent paper by Ford, Torsney and Wu (1992). In the case $\#\mathrm{supp}(\mu) = k < \infty$, the criterion in (3.1) gives the so-called A-optimality criterion and Theorem 3.1 reduces to the geometric characterization for quadratic loss in Studden (1971). An interesting case appears when the function $f(x, \vartheta)$ is independent of ϑ [e.g., $\eta(x, \vartheta) = \vartheta' x$]. Here the optimality criterion (3.1) also reduces to the A-optimality criterion where $A \in \mathbb{R}^{k \times s}$ is any matrix satisfying $AA' = E_\mu(c(\vartheta)c(\vartheta)')$. In contrast to the first case, there are now two geometric characterizations available. On the one hand, we can apply Theorem 3.1 using μ-a.e. the representation (3.6) and the set $\mathscr{R}_\Theta^c \subseteq L_\mu^2(k)$. On the other hand, we can use Theorem 1.1 in Studden (1971) for any square root $A \in \mathbb{R}^{k \times s}$ of the matrix $E_\mu(c(\vartheta)c(\vartheta)')$ and an Elfving set in $\mathbb{R}^{k \times s}$. Which of these results is easier to apply will usually depend on the specific situation. It is also worthwhile mentioning that for $\eta(x, \vartheta) = \vartheta' x$, $f(x, \vartheta) = x$, $\Theta = \mathscr{X}$ and $c(\vartheta) = \vartheta$, Theorem 3.1 gives a new geometric characterization for the integrated variance criterion as considered in Studden (1977) and Cook and Nachtsheim (1982).

REMARK 3.3. In this paper we have concentrated on the Bayesian c- and D-optimality criteria because these criteria have an interpretation from a Bayesian point of view in terms of a utility function [see, e.g., Chaloner and Larntz (1989)]. But it is also worth mentioning that there exist geometric characterizations for many other optimality criteria with a similar form as the criteria defined in (2.3) and (3.1). As a further example (which also has a Bayesian motivation), we consider a generalization of the Bayesian c-optimality criterion, namely the minimization of

$$E_\mu\left[tr\left(A(\vartheta)' M_k^-(\xi, \vartheta) A(\vartheta)\right)\right], \tag{3.7}$$

where $A(\vartheta) \in L_\mu^2(k \times s)$ is a matrix-valued function with quadratic integrable elements. It can then be shown that a design ξ with $\mathrm{range}(A(\vartheta)) \subseteq \mathrm{range}(M_k(\xi, \vartheta))$ (μ-a.e.) minimizes (3.7) if and only if there exists a constant $\gamma > 0$ and for all $x \in \mathrm{supp}(\xi)$ functions $\varepsilon(x, \cdot) \in L_\mu^2(s)$ with $\langle \varepsilon(x, \cdot), \varepsilon(x, \cdot)\rangle_s = 1$ such that the function $\gamma A(\cdot)$ has the representation

$$\gamma A(\vartheta) = \int f_k(x, \vartheta) \varepsilon(x, \vartheta)'\, d\xi(x) \quad \mu\text{-a.e.}$$

and is a boundary point of the Elfving set

$$\mathscr{R}_\Theta^A = \mathrm{conv}\Big(\Big\{g \in L_\mu^2(k \times s) \,\Big|\, g(\vartheta) = f_k(x, \vartheta)\varepsilon(\vartheta)',\ x \in \mathscr{X},\ \varepsilon \in L_\mu^2(s),\ \langle \varepsilon, \varepsilon\rangle_s = 1\Big\}\Big).$$

4. Examples.

EXAMPLE 4.1. Consider the exponential growth model $\eta(x, \beta) = e^{-\beta x}$, $\beta > 0$, $f_1(x, \beta) = -xe^{-\beta x}$, with gamma prior

$$\frac{d\mu(\beta)}{d\beta} = \frac{\alpha^{m+1}}{\Gamma(m+1)} \beta^m e^{-\alpha\beta} I\{\beta > 0\}, \tag{4.1}$$

$\alpha > 0$, $m > 1$, and design space $= [0, x_0]$, where $x_0 > \alpha/(m+1)$. Let ξ^* denote the design which puts all mass at $x^* = \alpha/(m+1)$. In order to show the Bayesian D-optimality of this design, we calculate the functions appearing in Theorem 2.1 from (2.6), (2.7) and (2.8) as follows:

$$\gamma_1(\beta) = \frac{\alpha}{m+1}\exp\left(-\frac{\alpha\beta}{m+1}\right), \qquad h_1(\beta) = \frac{m+1}{\alpha}\exp\left(\frac{\alpha\beta}{m+1}\right),$$
$$\varepsilon_1(\beta) \equiv -1.$$

Condition (a), (c) and (d) of Theorem 2.1 ($k = 1$) are obviously satisfied by this choice (note that $h_1 \in L^2_\mu$ because $m > 1$). Finally, we have that

$$\left|\int_0^\infty f_1(x,\beta)h_1(\beta)\varepsilon(\beta)\,d\mu(\beta)\right| \le x^2\left(\frac{m+1}{\alpha}\right)^2\left(\frac{2}{\alpha}x - \frac{2}{m+1} + 1\right)^{-m-1} \le 1$$

for all $x \in [0, x_0]$ and all $\varepsilon \in L^2_\mu$ with $\langle \varepsilon, \varepsilon \rangle_1 = 1$. This shows that $h_1(\beta)$ defines a supporting hyperplane to the Elfving set

$$\mathscr{R}^D_\Theta = \operatorname{conv}\left(\left\{g \in L^2_\mu \middle| g(\beta) = \varepsilon(\beta) f_1(x,\beta),\ x \in \mathscr{X}, \langle \varepsilon, \varepsilon \rangle_1 = 1\right\}\right)$$

at the boundary point γ_1. By Theorem 2.1 the one-point design ξ^* is Bayesian D-optimal.

EXAMPLE 4.2. Consider the exponential growth model with two parameters $\eta(x, \alpha, \beta) = \alpha e^{-\beta x}$, $\alpha > 0$, $\beta > 0$, $\mathscr{X} = [0, 3]$. In the notation of Section 2 we have $k = 2$, $f_1(x, \beta) = e^{-\beta x}$, $f_2(x, \beta) = e^{-\beta x}(1, -\alpha x)'$. As a prior we use the gamma density in (4.1) with the special choice $m = \alpha = 2$ and consider the design ξ^* which puts equal masses at the points $x_0^* = 0$ and $x_1^* = \frac{2}{3}$. By straightforward but somewhat tedious calculations, we find the quantities appearing in Theorem 2.1 and it is easy to see that conditions (a), (c) and (d) of this theorem are satisfied. For the supporting hyperplane in condition (b), we have from (2.7)

$$h_1(\beta) = \left\{\tfrac{1}{2}(1 + e^{-4/3\beta})\right\}^{-1/2},$$
$$h_2(\beta) = \tfrac{3}{4}e^{2/3\beta}h_1(\beta)\cdot\begin{pmatrix} 4/3e^{-4/3\beta} \\ 1 + e^{-4/3\beta}\end{pmatrix}$$

and obtain for all $\varepsilon \in L^2_\mu(2)$ satisfying $\langle \varepsilon, \varepsilon \rangle_2 = 2$ that

$$\begin{aligned}
&\left|\frac{1}{2}\int_0^\infty \sum_{j=1}^2 \varepsilon_j(\beta) f_j(x,\beta)' h_j(\beta)\,d\mu(\beta)\right|^2 \\
&\quad \le \frac{1}{2}\int_0^\infty \sum_{j=1}^2 \left(f_j(x,\beta)'h_j(\beta)\right)^2 d\mu(\beta) \\
&\quad = 9\int_0^\infty \left\{\left(x - \frac{2}{3}\right)^2 e^{2\beta(x+1)} + x^2 e^{-2\beta(x+1/3)}\right\}\beta^2\,d\beta \\
&\quad = \frac{1}{4}\left\{\frac{(3x-2)^2}{(x+1)^3} + \frac{9x^2}{(x+(1/3))^3}\right\} \le 1
\end{aligned}$$

for all $x \in [0,3]$ (here the last inequality follows by straightforward algebra showing that the derivative of the left-hand side has only the nonnegative zeros $x = 0$ and $x = \frac{2}{3}$). Therefore, $\frac{1}{2}(h_1(\beta), h_2(\beta)')'$ defines a supporting hyperplane to the Bayesian Elfving set $\mathscr{R}_\Theta^D$ [note that the components of h_1 and h_2 are quadratic integrable with respect to the measure $d\mu(\beta)$] and by Theorem 2.1 the design ξ^* with equal masses at $x_0^* = 0$ and $x_1^* = \frac{2}{3}$ is Bayesian D-optimal.

EXAMPLE 4.3. Let $\eta(x, \beta) = xe^{-\beta x}$ $[f_1(x,\beta) = -x^2 e^{-\beta x}]$, $x \geq 0$, $\beta > 0$. Then the maximum of η is attained for $x_{\max} = b(\beta) = 1/\beta$. In order to find a design which minimizes the asymptotic variance of the maximum likelihood estimator for $x_{\max}$, we use the criterion (3.1) with $c(\beta) = 1/\beta^2$ and gamma prior in (4.1) where $m > 7$, $\alpha > 0$. Consider the one-point design ξ^* at $x^* = 2\alpha/(m+1)$. Then we calculate from (3.1), (3.2) and (3.3) the quantities appearing in Theorem 3.1 as

$$\gamma = \left\{ \frac{\Gamma(m-3)}{\Gamma(m+1)} \left(\frac{x^*}{\alpha} \right)^{-4} \left(1 - 2\frac{x^*}{\alpha} \right)^{-m+3} \right\}^{-1/2},$$

$$h(\beta) = \gamma \beta^{-2} (x^*)^{-4} e^{2\beta x^*},$$

$$\varepsilon(x^*, \beta) = -\gamma(\beta x^*)^{-2} e^{\beta x^*}.$$

Obviously, we have

$$\int_0^\infty \varepsilon(x^*, \beta)^2 \, d\mu(\beta) = 1,$$

$$\gamma \int_0^\infty c(\beta) h(\beta) \, d\mu(\beta) = 1$$

and obtain from this

$$\begin{aligned}\left(\int_0^\infty \varepsilon(\beta) f_1(x,\beta) \, d\mu(\beta) \right)^2 &\leq \int_0^\infty (h(\beta) f_1(x,\beta))^2 \, d\mu(\beta) \\ &= \gamma^2 (x^*)^{-8} x^4 \frac{\alpha^{m+1}}{\Gamma(m+1)} \int_0^\infty \beta^{m-4} e^{-\beta(\alpha + 2x - 4x^*)} \, d\beta \\ &= \left(\frac{x}{x^*} \right)^4 \left(\frac{\alpha - 2x^*}{\alpha + 2x - 4x^*} \right)^{m-3} \leq 1\end{aligned}$$

for all $x \in [0, \infty)$ and $\varepsilon \in L_\mu^2$ with $\int_0^\infty \varepsilon^2(\beta) \, d\mu(\beta) = 1$. This shows that the function $\gamma c(\beta) = \gamma/\beta^2$ is a boundary point of the set $\mathscr{R}_\Theta^c$ with a representation (3.6) using ξ^*. By Theorem 3.1 the design ξ^* with mass 1 at the point $x^* = 2\alpha/(m+1)$ is Bayesian c-optimal for the estimation of the maximum of the nonlinear regression function $\eta(x, \beta) = xe^{-\beta x}$.

Acknowledgments. Parts of this paper were written during a visit to Purdue University in the summer of 1992. The author would like to thank the Department of Statistics for its hospitality and the Deutsche Forschungsgemeinschaft for the financial support that made this visit possible. The author is also grateful to L. Haines, H. M. Neugebauer, to an unknown referee and to Associate Editor for constructive comments on an earlier version of this paper.

REFERENCES

CHALONER, K. (1987). An approach to experimental design for generalized linear models. In *Model Orientated Data Analysis. Lecture Notes in Econom. and Math. Systems* **297** 3–12. Springer, Berlin.

CHALONER, K. (1989). Bayesian design for estimating the turning point of a quadratic regression. *Comm. Statist. Theory Methods* **18** 1385–1400.

CHALONER, K. (1993). A note on Bayesian optimal designs for non-linear problems. *J. Statist. Plann. Inference* **37** 229–236.

CHALONER, K. and LARNTZ, K. (1989). Optimal Bayesian design applied to logistic experiments. *J. Statist. Plann. Inference* **18** 191–208.

CHALONER, K. and LARNTZ, K. (1992). Bayesian design for accelerated life testing. *J. Statist. Plann. Inference* **33** 245–260.

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

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

DETTE, H. (1993). Elfving's theorem for D-optimality. *Ann. Statist.* **21** 753–766.

DETTE, H. and NEUGEBAUER, H. M. (1996). Bayesian optimal one point designs for one parameter nonlinear models. *J. Statist. Plann. Inference* **52** 17–31.

ELFVING, G. (1952). Optimum allocation in linear regression. *Ann. Math. Statist.* **23** 255–262.

FORD, I., TORSNEY, B. and WU, C. F. J. (1992). The use of canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Statist. Soc. Ser. B* **54** 569–583.

KIEFER, J. (1971). On the role of symmetry and approximation in exact design optimality. In *Statistical Decision Theory and Related Topics* (S. S. Gupta and J. Yackel, eds.) 109–118. Academic Press, New York.

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

LÄUTER, E. (1974b). Optimal design for nonlinear models. *Math. Operationsforsch Statist. Ser. Statist.* **5** 625–636 (in Russian).

PRONZATO, L. and WALTER, W. (1985). Robust experimental design via stochastic approximation. *Math. Biosci.* **75** 103–120.

SILVEY, S. D. (1980). *Optimal Design*. Chapman and Hall, New York.

STUDDEN, W. J. (1971). Elfving's theorem and optimal designs for quadratic loss. *Ann. Math. Statist.* **42** 1613–1621.

STUDDEN, W. J. (1977). Optimal designs for integrated variance in polynomial regression. In *Statistical Decision Theory and Related Topics* **2** 411–420. Academic Press, New York.

ZACKS, S. (1977). Problems and approaches in design of experiments for estimation and testing in nonlinear problems. In *Multivariate Analysis IV* (P. R. Krishnaiah, ed.) 209–223. North-Holland, Amsterdam.

INSTITUT UND FAKULTÄT FÜR MATHEMATIK
RUHR-UNIVERSITÄT BOCHUM
44780 BOCHUM
GERMANY
E-MAIL: helger.dette@rz.ruhr-uni-bochum.de