The Annals of Statistics
1993, Vol. 21, No. 2, 753–766

ELFVING'S THEOREM FOR D-OPTIMALITY

BY HOLGER DETTE

Universität Göttingen

We consider a model robust version of the c-optimality criterion minimizing a weighted product with factors corresponding to the variances of the least squares estimates for linear combinations of the parameters in different models. A generalization of Elfving's theorem is proved for the optimal designs with respect to this criterion by an application of an equivalence theorem for mixtures of optimality criteria. As a special case an Elfving theorem for the D-optimal design problem is obtained. In the case of identical models the connection between the A-optimality criterion and the model robust criterion is investigated. The geometric characterizations of the optimal designs are illustrated by a couple of examples.

1. Introduction. Consider the usual linear regression model

$$g_1(x) = f_1'(x)\theta_1, \tag{1.1}$$

where $f_1'(x) = (f_{11}(x), \dots, f_{1k_1}(x))$ and x is the control variable which takes values in a compact space $\mathscr{X}$ with sigma field $\mathscr{B}$ including all one point sets and containing (at least) k_1 points $x_1, \dots, x_{k_1}$ such that $f_1(x_1), \dots, f_1(x_{k_1})$ are linearly independent (here the double index is used to be consistent with later notation). $\theta_1' = (\theta_{11}, \dots, \theta_{1k_1})$ is the vector of unknown parameters and the functions $f_{11}(x), \dots, f_{1k_1}(x)$ are assumed to be real valued and continuous on the design space $\mathscr{X}$. For every $x \in \mathscr{X}$ a random variable $Y_1(x)$ with mean $g_1(x) = f_1'(x)\theta_1$ and variance $\sigma^2 > 0$ can be observed where different observations are assumed to be uncorrelated. In this paper we consider approximate design theory where a design ξ is defined as probability measure on the sigma field $\mathscr{B}$ and the matrix

$$M_1(\xi) = \int_{\mathscr{X}} f_1(x) f_1'(x)\, d\xi(x)$$

is called the information matrix (or moment matrix) of the design ξ in the model g_1. An optimal design minimizes (or maximizes) an appropriate optimality criterion depending on $M_1^-(\xi)$ [or $M_1(\xi)$], where $M_1^-(\xi)$ denotes a generalized inverse of $M_1(\xi)$.

There are numerous optimality criteria in the literature to discriminate between competing designs [see, e.g., Kiefer (1974), Silvey (1980) or Pukelsheim (1980)] and we will only state the optimality criteria which are investigated from a geometric point of view in this paper. A design ξ is called D-optimal if ξ maximizes the determinant of the information matrix $M_1(\xi)$. For a given

Received November 1991; revised April 1992.
AMS 1991 *subject classifications*. Primary 62K05; secondary 62J05.
Key words and phrases. Elfving's theorem, c-optimal designs, A-optimal designs, model robust designs, geometric characterizations of optimal designs.

vector $c_1 \in \mathbb{R}^{k_1}$ we call a design c_1-optimal if it minimizes $c_1' M_1^-(\xi) c_1$. In order to satisfy the estimability of $c_1'\theta_1$ or testability of $c_1'\theta_1 = 0$ for a given design ξ we have to assume that $c_1 \in \text{range}(M_1(\xi))$ [see Pukelsheim (1980)]. Similarly, for a given matrix $A_1 \in \mathbb{R}^{k_1 \times s_1}$ a design ξ is called optimal for $A_1'\theta_1$ if $\text{range}(A_1) \subseteq \text{range}(M_1(\xi))$ and ξ minimizes $\text{trace}(A_1' M_1^-(\xi) A_1)$. For the statistical interpretation of the proposed optimality criteria we refer the reader to the books of Fedorov (1972) or Silvey (1980).

The theory described so far is based on the fact that the underlying model (1.1) is known by the experimenter before the experiments are carried out. This is not very realistic because in practice it is often the goal of the experiment to identify a suitable model for the description of the dependency of Y_1 from x. Sometimes an optimal design for a given model performs extremely badly in other models [see e.g., Huber (1981), page 243 or Box and Draper (1959)]. There are numerous publications in the literature to avoid these drawbacks of "classical" design theory [e.g., Box and Draper (1959), Stigler (1971), Atkinson (1972) and Atkinson and Fedorov (1975)]. In this paper we will investigate a model robust version of optimality for the estimation of parameter subsystems. To this end assume that the experimenter knows that the "true" model belongs to a given class of models (e.g., polynomials up to degree $m - 1$)

$$\mathscr{F}_m = \{g_1(x), \dots, g_m(x)\},$$

where $g_l(x) = f_l'(x)\theta_l$, $f_l'(x) = (f_{l1}(x), \dots, f_{lk_l}(x))$ and $\theta_l' = (\theta_{l1}, \dots, \theta_{lk_l})$. For every $l \in \{1, \dots, m\}$ the functions $f_{l1}, \dots, f_{lk_l}$ are assumed to be real valued and continuous on $\mathscr{X}$, where the design space $\mathscr{X}$ should contain (at least) $C = \max_{l=1}^m k_l$ different points $x_1, \dots, x_C$ such that for every $l = 1, \dots, m$ the vectors $f_l(x_1), \dots, f_l(x_{k_l}) \in \mathbb{R}^{k_l}$ are linearly independent. The information matrix of a design ξ in the model g_l is now given by $M_l(\xi) = \int_{\mathscr{X}} f_l(x) f_l'(x)\, d\xi(x)$, $l = 1, \dots, m$. Assume that the experimenter wants to estimate in every model $g_l \in \mathscr{F}_m$ the linear combination $c_l'\theta_l$, $l = 1, \dots, m$, where $c_1, \dots, c_m$ are given vectors with $c_l \in \mathbb{R}^{k_l}$ or different linear combination in the same model (in this case all vectors f_l would be the same) and let $\beta_1, \dots, \beta_m$ denote positive numbers with sum 1. The vector $\beta = (\beta_1, \dots, \beta_m)$ is called a prior for the class $\mathscr{F}_m$ and β_l reflects the experimenter's belief about the adequacy of the model g_l or the importance of the linear combination $c_l'\theta_l$ $l = 1, \dots, m$. The following definition is a natural generalization of the c_1-optimality criterion ($m = 1$) and was originally introduced by Läuter (1974) for a model robust version of the D-optimality criterion.

Definition 1.1. Let $\beta = (\beta_1, \dots, \beta_m)$ denote a vector of positive numbers satisfying $\sum_{l=1}^m \beta_l = 1$, $c_l \in \mathbb{R}^{k_l}$, $l = 1, \dots, m$, and $c = (c_1', \dots, c_m')' \in \mathbb{R}^k$ where $k = \sum_{l=1}^m k_l$. A design ξ for which all linear combinations $c_l'\theta_l$ are estimable is called *c-optimal* for the class $\mathscr{F}_m$ with respect to the prior $\beta = (\beta_1, \dots, \beta_m)$ if ξ minimizes the function

$$\Phi_\beta(\xi) = \sum_{l=1}^m \beta_l \log[c_l' M_l^-(\xi) c_l].$$

If all models in the set $\mathscr{F}_m$ are identical (i.e., $f_l = f_1,\ l = 1, \dots, m$) the minimizing design ξ is called *c-optimal with respect to the prior* $\beta = (\beta_1, \dots, \beta_m)$.

Note that in the case of identical models $g_l = g_1,\ l = 1, \dots, m$, this is an alternative approach compared to the determination of an optimal design for $A_1'\theta_1$ where $A_1 = (c_1, \dots, c_m) \in \mathbb{R}^{k_1 \times m}$ and we will discuss these two optimality criteria in more detail in Section 4.

A very useful method for the determination of the optimal design are geometric characterizations. For the D-optimal design problem Sibson (1972) and Silvey and Titterington (1973) used strong Lagrangian theory to show that the D-optimal design problem is the dual of the so called "minimal ellipsoid problem" [see also Silvey (1972)]. This means to find the ellipsoid centered at the origin containing the regression space $\{f_1(x) | x \in \mathscr{X}\}$ and having minimal content. The most famous geometric result in optimal design theory dates back to Elfving (1952, 1959) and yields to a characterization of the c_1-optimal design (i.e., $m = 1$, $\beta_1 = 1$ in Definition 1.1). Following the work of Elfving (1952, 1953, 1959) we define the Elfving set by

$$\mathscr{R}_1 = \operatorname{co}(\{f_1(x) | x \in \mathscr{X}\} \cup \{-f_1(x) | x \in \mathscr{X}\}) \subseteq \mathbb{R}^{k_1}, \tag{1.2}$$

where $\operatorname{co}(A)$ denotes the convex hull of a set $A \subseteq \mathbb{R}^{k_1}$. $\mathscr{R}_1$ is a compact, symmetric and convex set spanning $\mathbb{R}^{k_1}$ and containing the point 0. A c_1-optimal design can be characterized as a design which allows a representation of the intersection of the half line $\{\lambda c_1 | \lambda > 0\}$ with the boundary of the set $\mathscr{R}_1$ (see Section 2). This result was generalized by Studden (1971) to optimal designs for parameter systems $A_1'\theta_1$. Fellman [(1974), Theorem 2.1.2] used the set $\mathscr{R}_1$ describing the location of the support points of the optimal design where the optimality criterion is an arbitrary decreasing function of the information matrix $M_1(\xi)$ with respect to the Loewner ordering.

All of the above characterizations do not have too much in common, especially since there does not exist a geometric result of Elfving type for the D-optimal design problem. It is the purpose of this paper to present a geometric structure which can be used for the characterization of optimal designs with respect to all three optimality criteria. In Section 2 a generalized Elfving theorem for the model robust criteria is derived by the application of an equivalence theorem for mixtures of optimality criteria recently proved by Gutmair (1991). The D-optimal design problem now appears as a special case in this setup by the consideration of "nested" models and an Elfving type characterization for the most popular optimality criterion is stated in Section 3. The considered Elfving sets allow geometric interpretation of the optimality criterion for $A_1'\theta_1$ by the investigation of identical models ($f_l = f_1$, $l = 1, \dots, m$) which are given in Section 4. Thus all three optimality criteria can be treated considering the geometric properties of one type of Elfving set.

2. Elfving's theorem for model robust designs.

Elfving (1952) proved the following geometric characterization for the c_1-optimal design problem

($m = 1$, $\beta_1 = 1$ in Definition 1.1) using the set $\mathscr{R}_1$ defined in (1.2) [see also Elfving (1959) and Pukelsheim (1981)].

THEOREM 2.1 [Elfving (1952)]. *A design* $\xi = \begin{Bmatrix} x_\nu \\ p_\nu \end{Bmatrix}_{\nu=1}^{s}$ (*for which* $c_1'\theta_1$ *is estimable*) *is* c_1*-optimal* (*in the model* g_1) *if and only if there exist a positive number* $\gamma_1 > 0$ *and real numbers* $\varepsilon_{11}, \dots, \varepsilon_{1s}$ *with* $\varepsilon_{1\nu}^2 = 1$, $\nu = 1, \dots, s$, *such that the point* $\gamma_1 c_1 = \sum_{\nu=1}^{s} p_\nu \varepsilon_{1\nu} f_1(x_\nu)$ *is a boundary point of the set* $\mathscr{R}_1$ *defined by* (1.2).

The proof of a generalization of this theorem to the model robust setup requires the following equivalent condition for c-optimal designs for the class $\mathscr{F}_m$ with respect to a prior β. Its proof involves general arguments for mixtures of information functions and can be found in Gutmair (1991) or Dette (1991).

THEOREM 2.2. *A design* ξ (*for which all linear combinations* $c_l'\theta_l$ *are estimable*, $l = 1, \dots, m$) *is optimal for the class* $\mathscr{F}_m$ *with respect to the prior* β *if and only if there exist generalized inverses* $G_1, \dots, G_m$ *of* $M_1(\xi), \dots, M_m(\xi)$ *such that*

$$\sum_{l=1}^{m} \beta_l \frac{\left(c_l' G_l f_l(x)\right)^2}{c_l' M_l^-(\xi) c_l} \le 1$$

for all $x \in \mathscr{X}$. *The equal sign in this inequality appears for all support points of every c-optimal design for the class* $\mathscr{F}_m$ *with respect to the prior* β.

We will now define an Elfving set $\mathscr{R}_m^\beta$ generalizing the set $\mathscr{R}_1$ of the original Elfving theorem by

$$\mathscr{R}_m^\beta := \operatorname{co}\left(\left\{(\varepsilon_1 f_1'(x), \dots, \varepsilon_m f_m'(x))' \mid x \in \mathscr{X}, \varepsilon_l \in \mathbb{R}, \; l = 1, \dots, m, \; \sum_{l=1}^{m} \beta_l \varepsilon_l^2 = 1\right\}\right), \tag{2.1}$$

which is also a convex, symmetric and compact subset of $\mathbb{R}^k$, $k = \sum_{l=1}^{m} k_l$, containing the point 0. Note that in the case $m = 1$ ($\beta_1 = 1$) (2.1) gives exactly the set considered in Theorem 2.1. In general the structure of the k-dimensional set $\mathscr{R}_m^\beta$ is very complicated and will be illustrated in some examples of the following section. We are now, able to prove an analogous geometric characterization of the c-optimal design problem for the class $\mathscr{F}_m$ with respect to a prior β as given in Theorem 2.1.

THEOREM 2.3. *A design* $\xi = \begin{Bmatrix} x_\nu \\ p_\nu \end{Bmatrix}_{\nu=1}^{s}$ (*for which* $c_l'\theta_l$ *is estimable* $l = 1, \dots, m$) *is c-optimal for the class* $\mathscr{F}_m$ *with respect to the prior* β *if and only if there exist positive numbers* $\gamma_1, \dots, \gamma_m$ *and numbers* $\varepsilon_{11}, \dots, \varepsilon_{1s}, \varepsilon_{21}, \dots,$

$\varepsilon_{2s},\dots,\varepsilon_{m1},\dots,\varepsilon_{ms}$ *such that the following properties* (a), (b), (c) *and* (d) *hold*:

(a) $$\gamma_l c_l = \sum_{\nu=1}^{s} p_\nu \varepsilon_{l\nu} f_l(x_\nu) \qquad l = 1,\dots,m.$$

(b) *The point* $(\gamma_1 c_1',\dots,\gamma_m c_m')'$ *is a boundary point of the set* $\mathscr{R}_m^\beta$ *with a supporting hyperplane* $(a_1',\dots,a_m')'$.

(c) $$\gamma_l c_l' a_l = \beta_l \qquad l = 1,\dots,m.$$

(d) $$\sum_{l=1}^{m} \beta_l \varepsilon_{l\nu}^2 = 1 \qquad \nu = 1,\dots,s.$$

PROOF. Let $\xi = \begin{Bmatrix} x_\nu \\ p_\nu \end{Bmatrix}_{\nu=1}^{s}$ denote an optimal design for the class $\mathscr{F}_m$ with respect to the prior β. By Theorem 2.2 there exist generalized inverses $G_1,\dots,G_m$ of $M_1(\xi),\dots,M_m(\xi)$ such that

$$\sum_{l=1}^{m} \beta_l \frac{(c_l' G_l f_l(x))^2}{c_l' M_l^-(\xi) c_l} \le 1 \quad \text{for all } x \in \mathscr{X}, \tag{2.2}$$

$$\sum_{l=1}^{m} \beta_l \frac{(c_l' G_l f_l(x_\nu))^2}{c_l' M_l^-(\xi) c_l} = 1 \quad \text{for all } \nu = 1,\dots,s. \tag{2.3}$$

Let $\gamma_l^{-2} = c_l' M_l^-(\xi) c_l$ and $d_l = \gamma_l G_l c_l$, $l = 1,\dots,m$, then it follows from the estimability of $c_l'\theta_l$ by the design ξ that $\gamma_l c_l = M_l(\xi) d_l = \sum_{\nu=1}^{s} p_\nu \varepsilon_{l\nu} f_l(x_\nu)$, $l = 1,\dots,m$, where $\varepsilon_{l\nu} = f_l'(x_\nu) d_l$, $l = 1,\dots,m$, $\nu = 1,\dots,s$. This proves the representation given in (a). Equation (2.3) and the representation of $\gamma_l c_l$ yield

$$\sum_{l=1}^{m} \beta_l \gamma_l c_l' d_l = \sum_{\nu=1}^{s} p_\nu \sum_{l=1}^{m} \beta_l (d_l' f_l(x_\nu))^2 = \sum_{l=1}^{m} \beta_l (d_l' f_l(x_\nu))^2 = 1, \tag{2.4}$$

which implies $(\varepsilon_{l\nu} = d_l' f_l(x_\nu))$ $\sum_{l=1}^{m} \beta_l \varepsilon_{l\nu}^2 = 1$, $\nu = 1,\dots,s$, and shows condition (d).

From the inequality (2.2) and the Cauchy–Schwarz inequality we get

$$\left(\sum_{l=1}^{m} \beta_l d_l' (\varepsilon_l f_l(x)) \right)^2 \le \sum_{l=1}^{m} \beta_l \varepsilon_l^2 \cdot \sum_{l=1}^{m} \beta_l (d_l' f_l(x))^2 \le 1$$

for all $x \in \mathscr{X}$, whenever the numbers $\varepsilon_1,\dots,\varepsilon_m$ satisfy the equation $\sum_{l=1}^{m} \beta_l \varepsilon_l^2 = 1$. Observing (2.4) it is now easy to see that the point $(\gamma_1 c_1',\dots,\gamma_m c_m')'$ is a boundary point of $\mathscr{R}_m^\beta$ with supporting hyperplane $a = (\beta_1 d_1',\dots,\beta_m d_m')'$ which proves (b). Finally the condition (c) follows readily from the definition of γ_l and d_l.

To prove sufficiency let $(a_1',\dots,a_m')'$, $a_l \in \mathbb{R}^{k_l}$, denote a supporting hyperplane to the set $\mathscr{R}_m^\beta$ at the boundary point $(\gamma_1 c_1',\dots,\gamma_m c_m')'$ and let $a_l = \beta_l d_l$,

$l = 1, \dots, m$. Thus we have for all $x \in \mathscr{X}$, $\varepsilon_1, \dots, \varepsilon_m$ satisfying $\sum_{l=1}^m \beta_l \varepsilon_l^2 = 1$

$$\left| \sum_{l=1}^m (\beta_l d_l)'(\varepsilon_l f_l(x)) \right| \le 1. \tag{2.5}$$

Defining $\varepsilon_l(x) = d_l' f_l(x) / \sqrt{\sum_{l=1}^m \beta_l (d_l' f_l(x))^2}$, $l = 1, \dots, m$, we see that (2.5) implies

$$\sum_{l=1}^m \beta_l (d_l' f_l(x))^2 \le 1 \quad \text{for all } x \in \mathscr{X}. \tag{2.6}$$

Because $(\beta_1 d_1', \dots, \beta_m d_m')'$ is a supporting hyperplane to $\mathscr{R}_m^\beta$ at the boundary point $(\gamma_1 c_1', \dots, \gamma_m c_m')'$ we obtain from (2.5) (used at $x = x_\nu$) and the representation (a)

$$1 = \sum_{l=1}^m \beta_l \gamma_l c_l' d_l = \sum_{\nu=1}^s p_\nu \sum_{l=1}^m \varepsilon_{l\nu} \beta_l f_l'(x_\nu)\, d_l \le 1$$

and this implies $\sum_{l=1}^m \beta_l \varepsilon_{l\nu} f_l'(x_\nu)\, d_l = 1$, $\nu = 1, \dots, s$. By an application of the Cauchy–Schwarz inequality we now get for $\nu = 1, \dots, s$

$$1 = \left(\sum_{l=1}^m \beta_l \varepsilon_{l\nu} f_l'(x_\nu)\, d_l \right)^2 \le \sum_{l=1}^m \beta_l \varepsilon_{l\nu}^2 \sum_{l=1}^m \beta_l (d_l' f_l(x_\nu))^2 \le 1, \tag{2.7}$$

where the last inequality results from (2.6) and condition (d). Therefore we have $\varepsilon_{l\nu} = \lambda_\nu d_l' f_l(x_\nu)$ for some $\lambda_\nu \in \mathbb{R}$, $l = 1, \dots, m$, $\nu = 1, \dots, s$. From the normalizing conditions on the $\varepsilon_{l\nu}$ in (d) we thus obtain observing (2.7)

$$1 = \sum_{l=1}^m \beta_l \varepsilon_{l\nu}^2 = \lambda_\nu^2 \sum_{l=1}^m \beta_l (d_l' f_l(x_\nu))^2 = \lambda_\nu^2 \qquad \nu = 1, \dots, s. \tag{2.8}$$

On the other hand, we have from the property that $(\gamma_1 c_1', \dots, \gamma_m c_m')'$ is a boundary point of $\mathscr{R}_m^\beta$ with supporting hyperplane $(\beta_1 d_1', \dots, \beta_m d_m')'$

$$\begin{aligned} 1 &= \sum_{l=1}^m \beta_l \gamma_l d_l' c_l = \sum_{\nu=1}^s p_\nu \sum_{l=1}^m \beta_l \varepsilon_{l\nu} (d_l' f_l(x_\nu)) \\ &= \sum_{\nu=1}^s p_\nu \lambda_\nu \sum_{l=1}^m \beta_l (d_l' f_l(x_\nu))^2 = \sum_{\nu=1}^s p_\nu \lambda_\nu. \end{aligned}$$

Equation (2.8), $p_\nu \ge 0$ and $\sum_{\nu=1}^s p_\nu = 1$ now show that $\lambda_\nu = 1$ whenever $p_\nu > 0$ and this implies $\varepsilon_{l\nu} = d_l' f_l(x_\nu)$, $l = 1, \dots, m$, $\nu = 1, \dots, s$, where we have assumed (without loss of generality) that in the representation (a) all p_ν are positive. From this representation we finally obtain for $l = 1, \dots, m$,

$$\gamma_l c_l = \sum_{\nu=1}^s p_\nu \varepsilon_{l\nu} f_l(x_\nu) = \sum_{\nu=1}^s p_\nu f_l(x_\nu) f_l'(x_\nu)\, d_l = M_l(\xi)\, d_l.$$

By the definition of a generalized inverse [see Searle (1982), page 238] it follows that there exist generalized inverses $G_1, \dots, G_m$ of the matrices $M_1(\xi), \dots, M_m(\xi)$ such that

$$d_l = \gamma_l G_l c_l, \qquad l = 1, \dots, m.$$

Observing the condition (c) it follows (note that $a_l = \beta_l d_l$ and that $c_l' G_l c_l$ is invariant with respect to the choice of the generalized inverse because $c_l'\theta_l$ is estimable)

$$1 = \gamma_l c_l' d_l = \gamma_l^2 c_l' G_l c_l = \gamma_l^2 c_l' M_l^-(\xi) c_l, \qquad l = 1, \dots, m$$

and the inequality (2.6) yields that there exist generalized inverses $G_1, \dots, G_m$ such that (2.2) holds for all $x \in \mathscr{X}$. By an application of Theorem 2.2 it now follows that the design ξ is optimal for the class $\mathscr{F}_m$ with respect to the prior β, which completes the proof of Theorem 2.3. □

Comparing the original Theorem 2.1 of Elfving with the model robust version given in Theorem 2.3 we see that there appears the additional condition (c) in the generalization of the theorem. This requires the hyperplane $a = (a_1', \dots, a_m')'$ to $\mathscr{R}_m^\beta$ at the point $(\gamma_1 c_1', \dots, \gamma_m c_m')'$ to satisfy some "normalizing" conditions in the lower dimensional subsets of the components. The inner products of the vectors $\gamma_l c_l$ and a_l corresponding to different models have to be equal to the a priori probabilities β_l for the models.

The following theorem gives a dual characterization of the c-optimal design problem for the class $\mathscr{F}_m$ with respect to the prior β. Its proof can be performed by similar arguments as given in Pukelsheim (1980) [see Dette (1991) for details] and is therefore omitted.

THEOREM 2.4 (Duality). *Let Ξ denote the set of all probability measures on $\mathscr{X}$ such that $c_l \in \text{range}(M_l(\xi))$, $l = 1, \dots, m$, and*

$$\mathscr{D} := \left\{ d = (d_1', \dots, d_m')' \,\middle|\, \left| \sum_{l=1}^m \beta_l d_l' \varepsilon_l f_l(x) \right| \le 1 \right.$$
$$\left. \forall\, x \in \mathscr{X}\ \forall\, \varepsilon_1, \dots, \varepsilon_m \text{ with } \sum_{l=1}^m \beta_l \varepsilon_l^2 = 1 \right\}$$

denote the set of all covering halfspaces to $\mathscr{R}_m^\beta$. Then the problems

$$\min_{\xi \in \Xi} \left\{ \sum_{l=1}^m \beta_l \log[c_l' M_l^-(\xi) c_l] \right\} \quad \text{and} \quad \max_{d \in \mathscr{D}} \left\{ \sum_{l=1}^m \beta_l \log(c_l' d_l)^2 \right\}$$

are dual problems and share a common extreme value.

3. A characterization of D-optimality and some examples. In this section we will investigate a special case of Theorem 2.3, which is of particular interest because it will yield a geometric characterization of Elfving type for the D-optimal design problem. To this end consider the "nested" models

$$f_1'(x) = f_{11}(x), \qquad f_2'(x) = (f_{11}(x), f_{12}(x)), \dots,$$
$$f_m'(x) = (f_{11}(x), f_{12}(x), \dots, f_{1m}(x))$$

and the vectors "for the highest coefficient" $\tilde{c}_l = (0, \dots, 0, 1)' \in \mathbb{R}^l$, $l = 1, \dots, m$. For this special choice [which is of particular interest to decide how many regression functions $f_{1l}(x)$ have to be included in the model] the optimality criterion $\Phi_\beta(\xi)$ reduces to

$$\Phi_\beta(\xi) = -\sum_{l=1}^{m} \beta_l \log \frac{\det M_l(\xi)}{\det M_{l-1}(\xi)}$$

and for the uniform prior $\beta^* = (1/m, \dots, 1/m)$ we obtain the D-optimality criterion. Thus we have (by an application of Theorem 2.3) the following geometric characterization for the D-optimal design problem.

THEOREM 3.1 (Elfving's theorem for D-optimality). *A design* $\xi = \left\{ \begin{matrix} x_\nu \\ p_\nu \end{matrix} \right\}_{\nu=1}^{s}$ *is D-optimal for the model* $g_m(x) = \theta_1 f_{11}(x) + \cdots + \theta_m f_{1m}(x)$ *if and only if there exist positive numbers* $\gamma_l > 0$, $l = 1, \dots, m$, *and numbers* $\varepsilon_{11}, \dots, \varepsilon_{1s}, \dots, \varepsilon_{m1}, \dots, \varepsilon_{ms}$ *such that*

(a) $$\gamma_l \tilde{c}_l = (0, \dots, 0, \gamma_l)' = \sum_{\nu=1}^{s} p_\nu \varepsilon_{l\nu} f_l(x_\nu) \qquad l = 1, \dots, m.$$

(b) *The point* $(\gamma_1 \tilde{c}_1', \dots, \gamma_m \tilde{c}_m')' = (\gamma_1, 0, \gamma_2, 0, 0, \gamma_3, \dots, 0, \dots, 0, \gamma_m)' \in \mathbb{R}^{m(m+1)/2}$ *is a boundary point of the set* $\mathscr{R}_m^{\beta^*} \subseteq \mathbb{R}^{m(m+1)/2}$, $\beta^* = (1/m, \dots, 1/m)$, *with supporting hyperplane* $(1/m)(d_1', d_2', \dots, d_m')'$, $d_l' = (d_{l1}, \dots, d_{ll})$.

(c) $$\gamma_l d_{ll} = 1 \quad \textit{for } l = 1, \dots, m.$$

(d) $$\sum_{l=1}^{m} \varepsilon_{l\nu}^2 = m \quad \textit{for } \nu = 1, \dots, s.$$

We will finish this section giving two examples to get more insight into the geometric structure of the characterizations of the c-optimal design problem investigated in this section and in Section 2. Some more examples can be found in a paper of Dette (1991).

EXAMPLE 3.2. In this example we want to determine the D-optimal design for the model $g_2(x) = \theta_1(1-x) + \theta_2 x^2$ where $x \in [0,1]$. To this end let $m = 2$, $k_1 = 1$, $k_2 = 2 f_1'(x) = 1 - x$ and $f_2'(x) = (1 - x, x^2)$. Although we could apply Theorem 3.1 directly, we will solve the more general problem determining a c-optimal design for the class $\mathscr{F}_2$ with respect to the prior $\beta^* = (1/2, 1/2)$ where $c = (c_1, c_2')' = (1, h_1, h_2)$. The set $\mathscr{R}_2^{\beta^*}$ defined in Section 2 is given by

$$\mathscr{R}_2^{\beta^*} = \left\{ (x_1, x_2, x_3)' | x_1^2 + x_2^2 \le 2, |x_3| \le \sqrt{2} - \sqrt{x_1^2 + x_2^2} \right\}$$

and depicted in Figure 1. We have to distinguish the cases $h_2 = 0$ and $h_2 \neq 0$.

(a) $h_2 = 0$: In this case the vector c is given by $c = (c_1', c_2')' = (1, h_1, 0)'$ which shows that the vector $(\gamma_1 c_1', \gamma_2 c_2')$ can only intersect the boundary of $\mathscr{R}_2^{\beta^*}$ at the curve $\mathscr{K} = \{(x_1, x_2, 0)' | x_1^2 + x_2^2 = 2\}$ obtained from the point $(f_1'(0), f_2'(0))'$. Therefore the c-optimal design [for the vector $c = (1, h_1, 0)'$]

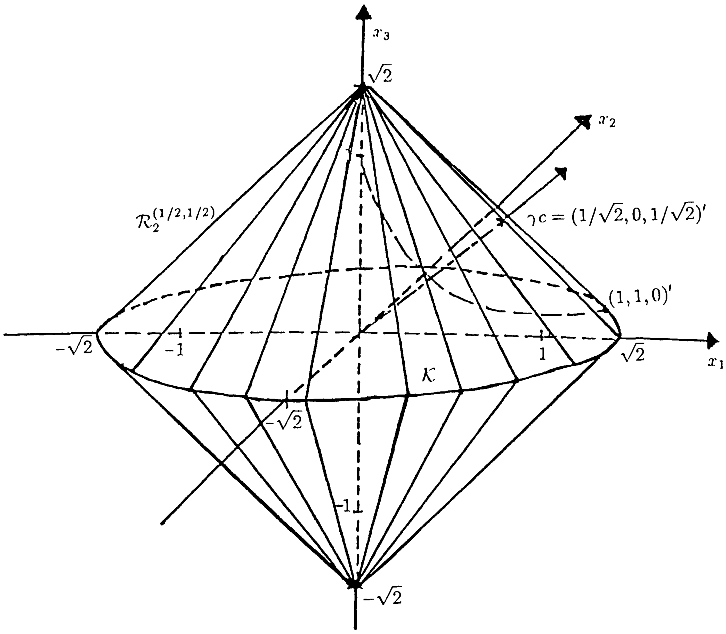

FIG. 1. *The set* $\mathscr{R}_2^{\beta^*}$ *for the models* $f_1'(x) = 1 - x$, $f_2'(x) = (1 - x, x^2)$.

puts mass 1 at the point 0 (note that this result holds also for an arbitrary prior β).

(b) $h_2 \neq 0$: In this case the vector $(\gamma_1 c_1', \gamma_2 c_2')'$ touches the boundary of $\partial \mathscr{R}_2^{\beta^*}$ at the some point $t \in \partial \mathscr{R}_2^{\beta^*}$ which is a convex combination of a point of $\mathscr{K}$ and one of the points $(0, 0, \sqrt{2})'$ or $(0, 0, -\sqrt{2})'$ depending on the sign of h_2. Therefore we see that the c-optimal design is supported at the points 0 and 1 (note that this result is independent of the prior β). From now on we will assume for definiteness that $h_2 > 0$. The calculation of the weights is more complicated because we have to determine the quantities $\gamma_1, \gamma_2, \varepsilon_{11}, \varepsilon_{12}, \varepsilon_{21}, \varepsilon_{22}$ of Theorem 2.3. First we remark that the condition $(\gamma_1 c_1', \gamma_2 c_2')' \in \partial \mathscr{R}_2^{\beta^*}$ implies

$$\gamma_2 h_2 = \sqrt{2} - \sqrt{\gamma_1^2 + \gamma_2^2 h_1^2}\,. \tag{3.1}$$

In the following let $d = (1/2)(\delta_1, \delta_2, \delta_3)'$ denote the supporting hyperplane to $\mathscr{R}_2^{\beta^*}$ at the point $(\gamma_1 c_1', \gamma_2 c_2')' = (\gamma_1, \gamma_2 h_1, \gamma_2 h_2)'$ then we have from condition (c) of Theorem 2.3

$$d = \left(\gamma_1^{-1}, \delta_2, (1 - \gamma_2 h_1 \delta_2)/(\gamma_2 h_2)\right)' \quad \text{for some } \delta_2 \in \mathbb{R} \tag{3.2}$$

and the supporting hyperplane property implies (note that we have assumed $h_2 > 0$)

$$\delta_2 = \frac{\sqrt{\gamma_1^2 + \gamma_2^2 h_1^2} - \gamma_2 h_2}{\gamma_2 h_1\left[\sqrt{\gamma_1^2 + \gamma_2^2 h_1^2} + \gamma_2 h_2\right]}.$$

Using (3.2) and (3.1) we thus obtain for the supporting hyperplane $(1/2)d$ at the point $(\gamma_1 c_1', \gamma_2 c_2')'$ by straightforward calculations $d = (\gamma_1^{-1}, (1-\sqrt{2}\gamma_2 h_2)/\gamma_2 h_1, \sqrt{2})' =: (d_1', d_2')'$ and from the definition $\varepsilon_{l\nu} = d_l' f_l(x_\nu)$ (compare with the proof of Theorem 2.3) we have $\varepsilon_{11} = \gamma_1^{-1}$, $\varepsilon_{21} = (1-\sqrt{2}\gamma_2 h_2)/(\gamma_2 h_1)$, $\varepsilon_{12} = 0$ and $\varepsilon_{22} = \sqrt{2}$. Condition (a) of Theorem 2.3 yields $p_1 = \gamma_1^2$, $p_2 = \gamma_2 h_2/\sqrt{2}$, $p_1 = (\gamma_2 h_1)^2/(1-\sqrt{2}\gamma_2 h_2)$ and $p_1 + p_2 = 1$. From these equations we get by straightforward algebra

$$\gamma_2 h_2 = \sqrt{2}\left(1-\gamma_1^2\right), \qquad p_1 = \gamma_1^2, \qquad p_2 = 1-\gamma_1^2, \tag{3.3}$$

$$2\gamma_1^4\left[1-\frac{h_1^2}{h_2^2}\right] - \gamma_1^2\left[1-4\frac{h_1^2}{h_2^2}\right] - 2\frac{h_1^2}{h_2^2} = 0. \tag{3.4}$$

From (3.1), (3.2) and (3.3) it can easily be shown that $(1/2)(\delta_1, \delta_2, \delta_3) \cdot (\gamma_1, \gamma_2 h_1, \gamma_2 h_2)' = 1$ and that the $\varepsilon_{l\nu}$ satisfy condition (d) of Theorem 2.3 [note that we have used all other conditions of this theorem to derive (3.2), (3.3) and (3.4)]. Thus we see that the c-optimal design for the class $\mathscr{F}_2$ with respect to the prior $\beta^* = (1/2, 1/2)$, $c = (1, h_1, h_2)'$, puts masses $p_1 = \gamma_1^2$ and $p_2 = 1 - \gamma_1^2$ at the points 0 and 1 where γ_1^2 is the positive solution of

$$2\gamma_1^4\left(h_2^2 - h_1^2\right) - \gamma_1^2\left(h_2^2 - 4h_1^2\right) - 2h_1^2 = 0. \tag{3.5}$$

Note that this result includes also the case $h_2 = 0$ for which (3.5) reduces to $(\gamma_1^2 - 1)^2 = 0$. For the vector $c = (1, 0, 1)$, $h_1 = 0$, $h_2 = 1$, we obtain the D-optimal design ($\gamma_1 = 1/\sqrt{2}$, $\gamma_2 = 1/\sqrt{2}$) for the model $\theta_1(1-x) + \theta_2 x^2$ (compare with Theorem 3.1) which puts equal masses at the points 0 and 1. If one is only interested in the D-optimal design a direct application of Theorem 3.1 would yield to an essential simplification of the calculations.

EXAMPLE 3.3. We will now show that the condition (c) is indeed necessary to obtain the equivalence of Theorem 2.3. To this end let $m = 2$, $\beta^* = (1/2, 1/2)$, $k_1 = 2$, $k_2 = 3$, $f_1'(x) = (1, x)$, $f_2'(x) = (1, x, x^2)$, $\mathscr{X} = [-1, 1]$ and $c' = (c_1', c_2')$, where $c_1' = (1, 2)$ and $c_2' = (1, 2, 4)$ (thus we want to extrapolate a linear or quadratic regression function at the point $x_0 = 2$). Consider the design ξ which puts masses $2/11, 3/11, 6/11$ at the points $-1, 0, 1$ and let $\varepsilon_{11} = -\sqrt{3/2}$, $\varepsilon_{12} = 0$, $\varepsilon_{13} = \sqrt{3/2}$, $\varepsilon_{21} = 1/\sqrt{2}$, $\varepsilon_{22} = -\sqrt{2}$, $\varepsilon_{23} = 1/\sqrt{2}$ and $\gamma_1 = (4/11)\sqrt{3/2}$, $\gamma_2 = \sqrt{2}/11$, then it is straightforward to show that the design ξ satisfies the conditions (a) and (d) of Theorem 2.3. To prove that the point $(\gamma_1 c_1', \gamma_2 c_2')'$ is a boundary point of $\mathscr{R}_2^{\beta^*} \subseteq \mathbb{R}^5$ define $d = (d_1', d_2')'$, where $d_1 = \sqrt{3/2}(0, 1)'$ and $d_2 = (3/\sqrt{2})(-2/3, 0, 1)'$, then we have $\gamma c' d = \gamma_1 c_1' d_1 + \gamma_2 c_2' d_2 = 2$ and by the Cauchy–Schwarz inequality

$$\begin{aligned}\left[\left(\varepsilon_1 f_1'(x)\, d_1\right) + \left(\varepsilon_2 f_2'(x)\, d_2\right)\right]^2 &\le \left(\varepsilon_1^2 + \varepsilon_2^2\right)\left[\left(f_1'(x)\, d_1\right)^2 + \left(f_2'(x)\, d_2\right)^2\right] \\ &= 9\left(x^2 - \tfrac{1}{2}\right)^2 + \tfrac{7}{4} \le 4\end{aligned}$$

whenever $\varepsilon_1^2 + \varepsilon_2^2 = 2$ and $x \in [-1, 1]$. Therefore it follows that the point $\gamma c' = (\gamma_1 c_1', \gamma_2 c_2')$ is a boundary point of the set $\mathscr{R}_2^{\beta^*}$ with supporting hyperplane $(1/2)d = ((1/2)d_1', (1/2)d_2')'$ and all conditions of Theorem 2.3 are satisfied except condition (c) [e.g., $(1/2)d_1'\gamma_1 c_1 = 6/11 \neq 1/2$]. If the design ξ would be c-optimal for the class $\mathscr{F}_2$ with respect to the prior $\beta = (1/2, 1/2)$ we obtain from Theorem 2.2 and straightforward algebra the inequality

$$\frac{1}{2}\sum_{l=1}^{2}\frac{\left(c_l' M_l^{-1}(\xi) f_l(x)\right)^2}{c_l' M_l^{-1}(\xi) c_l} = \frac{11}{16}x^2 + \frac{11}{10}\left(\frac{3}{2}x^2 - 1\right)^2 \leq 1$$

for all $x \in [-1, 1]$. But for the point $x = 0$ this inequality does not hold and thus ξ is not a c-optimal design for the class $\mathscr{F}_2$ with respect to the prior β. This shows that condition (c) of Theorem 2.3 cannot be omitted.

4. A-optimal designs. The results of Section 2 can easily be transferred to the case where the experimenter wants to estimate several linear combinations $A_l'\theta_l$ in different models $g_l \in \mathscr{F}_n$ where $A_l \in \mathbb{R}^{k_l \times s_l}$, $l = 1, \dots, n$, are given matrices. An A-optimal design for the class $\mathscr{F}_n$ with respect to the prior β, $A = (A_1', \dots, A_n')'$, allows the estimability of all linear combinations $A_l'\theta_l$ and minimizes the function $\sum_{l=1}^{n}\beta_l \log[\operatorname{trace}\{M_l^-(\xi)A_l A_l'\}]$.

THEOREM 4.1. *Let $\mathscr{R}_n^*$ denote the convex hull of the set*

$$\left\{(f_1(x)\varepsilon_1', \dots, f_n(x)\varepsilon_n')' | x \in \mathscr{X}, \varepsilon_l \in \mathbb{R}^{s_l}, \sum_{l=1}^{n}\beta_l\|\varepsilon_l\|_2^2 = 1\right\}$$
$$\subseteq \mathbb{R}^{k_1 \times s_1} \times \cdots \times \mathbb{R}^{k_n \times s_n}.$$

The design $\xi = \left\{\begin{matrix} x_\nu \\ p_\nu \end{matrix}\right\}_{\nu=1}^{s}$ (for which all linear combinations $A_l'\theta_l$ are estimable) is A-optimal for the class $\mathscr{F}_n$ with respect to the prior β if and only if there exist positive numbers $\gamma_1, \dots, \gamma_n$ and vectors $\varepsilon_{l\nu} \in \mathbb{R}^{s_l}$, $l = 1, \dots, n$, $\nu = 1, \dots, s$, such that the following properties (a*), (b*), (c*) *and* (d*) *hold*:

(a*) $$\gamma_l A_l = \sum_{\nu=1}^{s} p_\nu f_l(x_\nu)\varepsilon_{l\nu}' \qquad l = 1, \dots, n.$$

(b*) *$(\gamma_1 A_1', \dots, \gamma_n A_n')'$ is a boundary point of the set $\mathscr{R}_n^*$ with supporting hyperplane $D = (D_1', \dots, D_n')'$, $D_l \in \mathbb{R}^{k_l \times s_l}$.*

(c*) $$\gamma_l \operatorname{trace}(D_l' A_l) = \beta_l \qquad l = 1, \dots, n.$$

(d*) $$\sum_{l=1}^{n}\beta_l\|\varepsilon_{l\nu}\|_2^2 = 1 \qquad \nu = 1, \dots, s.$$

Note that in the case $n = 1$ and $\beta_1 = 1$ Theorem 4.1 gives exactly the Elfving theorem 1.1 proved by Studden (1971). In the remaining section we will tackle this special case and investigate the relationship between optimal designs for $A_1'\theta_1$ [i.e., optimal designs minimizing $\operatorname{trace}(A_1'M_1^-(\xi)A_1)$, $n = 1$ and $s_1 = m$ in Theorem 4.1] and the c-optimal designs for the class $\mathscr{F}_m$ with respect to a prior β. To this end we consider the case of m identical models $f_l = f_1$ (see Definition 1.1) and assume that the experimenter wants to estimate the linear combinations $c_1'\theta_1, \dots, c_m'\theta_1$ in one model $g_1(x) = f_1'(x)\theta_1$ for given vectors $c_1, \dots, c_m \in \mathbb{R}^{k_1}$. There are now two optimality criteria which can be used for the discrimination between competing designs, namely the minimization of $\operatorname{trace}(A_1' M_1^-(\xi) A_1)$ where $A_1 = (c_1, \dots, c_m) \in \mathbb{R}^{k_1 \times m}$, and the minimization of $\Phi_\beta(\xi) = \sum_{l=1}^m \beta_l \log[c_l' M_1^-(\xi) c_l]$ where $\beta = (\beta_1, \dots, \beta_m)$ is a given prior (see Definition 1.1). The first criterion is an equally weighted sum of the quantities $c_l' M_1^-(\xi) c_l$ and the second is essentially a weighted product of these terms. This has the consequence that the c-optimal design is not changing if a "component" c_l of the vector $c = (c_1' \dots, c_m')'$ is multiplied with some positive scalar while the optimal design for $A_1'\theta_1$ is changing with different scalings of the columns of the matrix A_1.

Another difference appears investigating the corresponding Elfving theorems for the two optimality criteria. To this end we recall the definition of the Elfving sets used in Theorem 1.1 of Studden (1971) (or Theorem 4.1 in the case $n = 1$ and $s_1 = m$)

$$\mathscr{S}_m := \operatorname{co}\big(\{ f_1(x)\varepsilon' | x \in \mathscr{X}, \varepsilon \in \mathbb{R}^m, \|\varepsilon\|_2 = 1\}\big) \subseteq \mathbb{R}^{k_1 \times m}.$$

By Theorem 1.1 of Studden (1971) (or Theorem 4.1 for $n = 1$ and $s_1 = m$) a number γ_1 has to be determined such that $\gamma_1 A_1$ is a boundary point of $\mathscr{S}_m$ and a representing design [satisfying (a*) for $n = 1$ in Theorem 4.1] is an optimal design for $A_1'\theta_1$. For the solution of the c-optimal design problem we have to find m positive numbers $\gamma_1, \dots, \gamma_m$ such that the point $(\gamma_1 c_1', \dots, \gamma_m c_m')'$ is a boundary point of the set $\mathscr{R}_m^\beta \subseteq \mathbb{R}^{k_1 m}$ and which also satisfies condition (c) of Theorem 2.3. This is, of course, a much harder task compared to Theorem 1.1 in Studden (1971) and can be seen as the price which has to be paid for the invariance property with respect to different scalings of the components in the optimality criterion. For a more detailed investigation we define the mapping ($a_l \in \mathbb{R}^{k_l}$, $l = 1, \dots, m$)

$$\Psi_\beta : \begin{cases} \mathbb{R}^{k_1 m} \to \mathbb{R}^{k_1 \times m}, \\ (a_1', \dots, a_m')' \to \big(\sqrt{\beta_1}\, a_1, \dots, \sqrt{\beta_m}\, a_m\big), \end{cases}$$

which is a one to one mapping from $\mathscr{R}_m^\beta$ onto $\mathscr{S}_m$, then it can easily be seen that $\partial \mathscr{S}_m = \Psi_\beta(\partial \mathscr{R}_m^\beta)$. The following results show the relation between the c-optimal design with respect to a prior $\beta(c = (c_1', \dots, c_m')')$ and the optimal design for $A_1'\theta_1$ ($A_1 = (c_1, \dots, c_m)$). The optimal design for $A_1'\theta_1$ is c-optimal with respect to a prior β depending on the supporting hyperplane to $\mathscr{S}_m$ at $\gamma_1 A_1$ while the c-optimal design with respect to a given prior is optimal for $\tilde{A}_1'\theta$ where the columns of $\tilde{A}_1$ are the scaled versions of the vectors c_l.

THEOREM 4.2. *Let* $A_1 = (c_1, \dots, c_m)$, $c = (c_1', \dots, c_m')'$, $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^s$ *denote an optimal design for* $A_1'\theta_1$ *and* $\gamma > 0$ *such that the point* γA_1 *is a boundary point of the set* $\mathscr{S}_m$ *with supporting hyperplane* $D = (d_1, \dots, d_m) \in \mathbb{R}^{k_1 \times m}$ [*see Theorem* 1.1 *of Studden* (1971)]. *The design* ξ *is also* c-*optimal with respect to the prior* $\beta = (\beta_1, \dots, \beta_m)$ *where the weights* β_l *are given by* $\beta_l = \gamma c_l' d_l$, $l = 1, \dots, m$.

PROOF. We will show that the design ξ satisfies the conditions (a)–(d) of Theorem 2.3. By Theorem 1.1 of Studden (1971) we have for γA_1 the representation $\gamma A_1 = \sum_{\nu=1}^{s} p_\nu f_1(x_\nu)\varepsilon_\nu' \in \partial\mathscr{S}_m$ where $\varepsilon_\nu = (\varepsilon_{1\nu}, \dots, \varepsilon_{m\nu})' \in \mathbb{R}^m$ with $\|\varepsilon_\nu\|_2 = 1$. Thus it follows that the point $\Psi_\beta^{-1}(\gamma A_1) = ((\gamma/\sqrt{\beta_1})c_1', \dots, (\gamma/\sqrt{\beta_m})c_m')'$ is a boundary point of the set $\mathscr{R}_m^\beta$ and we obtain the representation

$$\frac{\gamma}{\sqrt{\beta_l}}c_l = \sum_{\nu=1}^{s} p_\nu \tilde{\varepsilon}_{l\nu} f_1(x_\nu) \qquad l = 1, \dots, m,$$

where $\tilde{\varepsilon}_{l\nu} = (1/\sqrt{\beta_l})\varepsilon_{l\nu}$ and $\sum_{l=1}^{m}\beta_l\tilde{\varepsilon}_{l\nu}^2 = \|\varepsilon_\nu\|_2^2 = 1$. This shows that the conditions (a), (b) and (d) of Theorem 2.3 are satisfied. For the proof of the remaining condition (c) we remark that it is easy to see that the vector $(\sqrt{\beta_1}d_1', \dots, \sqrt{\beta_m}d_m')'$ defines a supporting hyperplane to $\mathscr{R}_m^\beta$ at the boundary point $\Psi_\beta^{-1}(\gamma A_1)$ and condition (c) is now obvious from the definition of $\beta_l = \gamma c_l' d_l$, $l = 1, \dots, m$. By an application of Theorem 2.3 we obtain that the design ξ is c-optimal with respect to the prior β where $\beta_l = \gamma c_l' d_l$. □

THEOREM 4.3. *Let $\gamma c = (\gamma_1 c_1', \dots, \gamma_m c_m')'$ denote an arbitrary boundary point of $\mathscr{R}_m^\beta$ and $\xi = \left\{ \begin{matrix} x_\nu \\ p_\nu \end{matrix} \right\}_{\nu=1}^{s}$ a design which allows the representation*

$$\gamma_l c_l = \sum_{\nu=1}^{s} p_\nu \varepsilon_{l\nu} f_1(x_\nu) \qquad l = 1, \dots, m \tag{4.1}$$

with real numbers $\varepsilon_{11}, \dots, \varepsilon_{1s}, \dots, \varepsilon_{m1}, \dots, \varepsilon_{ms}$ satisfying $\sum_{l=1}^{m}\beta_l\varepsilon_{l\nu}^2 = 1$, $l = 1, \dots, \nu$. The design ξ is optimal for $A_1'\theta_1$ where $A_1 = \Psi_\beta(\gamma c)$.

PROOF. The point $A_1 = \Psi_\beta(\gamma c) = (\gamma_1\sqrt{\beta_1}c_1, \dots, \gamma_m\sqrt{\beta_m}c_m)$ is a boundary point of the set $\mathscr{S}_m$ and it follows from (4.1) that A_1 has the representation

$$A_1 = \sum_{\nu=1}^{s} p_\nu f_1(x_\nu)\varepsilon_\nu'$$

with $\varepsilon_\nu = (\sqrt{\beta_1}\varepsilon_{1\nu}, \dots, \sqrt{\beta_m}\varepsilon_{m\nu})'$ and $\|\varepsilon_\nu\|_2 = 1$. By the Elfving theorem 1.1 of Studden (1971) we see that the design ξ is optimal for $A_1'\theta_1$. □

REMARK 4.4. Note that Theorem 4.3 does not require the design ξ to be c-optimal with respect to a given prior β. It is enough that the design allows a representation (4.1) of a boundary point of the set $\mathscr{R}_m^\beta$. In general the condition (c) of Theorem 2.3 is not fulfilled and thus ξ is not c-optimal for the given prior β.

Acknowledgments. This paper is based on a part of the author's Habilitationsschrift written at the University of Göttingen [Dette (1992)]. The author would like to thank Bill Studden for his helpful comments and the Department of Statistics, Purdue University, for its hospitality during the Spring semester 1991. The author is also grateful to the referees for their constructive comments which improved the presentation of the results.

REFERENCES

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

ATKINSON, A, C. and FEDOROV, V. V. (1975). Optimal design: Experiments for discriminating between several models. *Biometrika* **62** 289–303.

BOX, G. E. P. and DRAPER, N. R. (1959). A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.* **54** 622–654.

DETTE, H. (1991). Some generalizations of Elfving's theorem. Technical Report 16-91, Dept. Statistics, Purdue Univ.

DETTE, H. (1992). Geometric characterizations of model robust designs. Habilitationsschrift, Univ. Göttingen.

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

ELFVING, G. (1953). Convex sets in statistics. In *XII Congr. Math. Scand.* 34–39.

ELFVING, G. (1959). *Design of Linear Experiments*: *The Harald Cramér Volume* (U. Grenander, ed.) 58–74. Wiley, New York.

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

FELLMAN, J. (1974). On the allocation of linear observations. *Comment. Phys.-Math.* **44** 27–78.

GUTMAIR, S. (1991). Polars and subgradients of mixtures of information functions. Technical Report 286, Schwerpunktprogramm der Deutschen Forschungsgemeinschaft "Anwendungsbezogene Optimierung und Steuerung," Univ. Augsburg.

HUBER, P. (1981). *Robust Statistics*. Wiley, New York.

KIEFER, J. (1974). General equivalence theory for optimum designs. *Ann. Statist.* **2** 849–879.

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

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* **4** 339–364.

PUKELSHEIM, F. (1981). On c-optimal design measures. *Math. Operationsforsch. Statist. Ser. Statist.* **12** 13–20.

SEARLE, R. S. (1982). *Matrix Algebra Useful for Statistics*. Wiley, New York.

SIBSON, R. (1972). Discussion of a paper of H. P. Wynn. *J. Roy. Statist. Soc. Ser. B* **34** 181–183.

SILVEY, S. D. (1972). Discussion of a paper of H. P. Wynn. *J. Roy. Statist. Soc. Ser. B* **34** 174–175.

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

SILVEY, S. D. and TITTERINGTON, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60** 21–32.

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

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

INSTITUT FÜR MATHEMATISCHE
STOCHASTIK
UNIVERSITÄT GÖTTINGEN
LOTZESTR. 13
3400 GÖTTINGEN
GERMANY