SOME GENERALIZATIONS OF ELFING'S THEOREM

by

Holger Dette
Purdue University
and
University of Göttingen

Technical Report #91-16C

Department of Statistics
Purdue University

April 1991

◇

Some Generalizations of Elfing's Theorem

by

Holger Dette

Abstract

We consider a model robust version of the c-optimality criterion and give geometric characterizations analogous to Elfing's (1952) characterization for the classical c-optimal design problem. As a special case we obtain an Elfing's theorem for the D-optimal design problem. Some examples are given in order to get insight into the geometric structure of the underlying problem. The results are derived from general equivalence theorems for model robust designs which are proved in the appendix.

1. Introduction

Consider the linear regression models

$$g_\ell(x) = f_\ell^T(x) \cdot \theta_\ell \qquad \ell = 1, \dots, n$$

where $f_\ell^T(x) = (f_{\ell,1}(x), \dots, f_{\ell,k_\ell}(x))$, x is the control variable, $\theta_\ell^T = (\theta_{\ell 1}, \dots, \theta_{\ell k_\ell})$ are the vectors of unknown parameters ($\ell = 1, \dots, n$). The design space $\mathcal{X}$ is assumed to be compact containing at least $k = k_1 + k_2 + \dots + k_n$ points and the regression functions $f_{\ell 1}(x), \dots, f_{\ell k_\ell}(x)$ are assumed to be independent for every $\ell = 1, \dots, n$.

A design ξ is a probability measure on $\mathcal{X}$ and the matrix

$$M_\ell(\xi) = \int_{\mathcal{X}} f_\ell(x) f_\ell^T(x) d\xi(x)$$

is called the information matrix of the design ξ in the model g_ℓ. If ξ is supported at m points $x_1, \dots, x_m$ with masses $\xi(x_i) = \frac{n_i}{N}$ the experimenter takes n_i uncorrelated observations at each point x_i ($i = 1, \dots, m$, $\sum_{i=1}^{m} n_i = N$). The covariance matrix of the least squares estimator for θ_ℓ (in the model $g_\ell(x)$) is proportional to the inverse of the information matrix and an optimal design minimizes (or maximizes) an appropriate optimality criterion depending on $M_\ell^-(\xi)$ (or $M_\ell(\xi)$), where $M_\ell^-(\xi)$ denotes a generalized inverse of $M_\ell(\xi)$.

In this paper we are interested in designs which allow good estimates for a given linear combination of the unknown parameter vectors θ_ℓ, i.e. $c_\ell^T\theta_\ell$, where $c_\ell \in \mathbb{R}^{k_\ell}$ is a given vector ($\ell = 1,\ldots,n$). We will start our investigations with one model (i.e. $n = 1$). A design is called c_1-optimal if it minimizes $c_1^T M_1^-(\xi)c_1$ (see Silvey (1980) p. 49). Usual choices for c_1 are $c_1 = (0,\ldots,0,1,0,\ldots,0)$ (precise estimation of one parameter of $\theta_1 = (\theta_{1,1},\ldots,\theta_{1,k_1})^T$) or $c_1 = (f_{11}(x_0),\ldots,f_{1k_1}(x_0))^T$ (precise estimation of the regression at the point x_0). The c_1 optimal design also maximizes the power of the F-test for the hypotheses $H_0 : c_1^T\theta_1 = 0$. In order to satisfy the estimability of $c_1^T\theta_1$ or testability of $c_1^T\theta_1 = 0$ for a given design ξ we have to assume that $c_1 \in$ range $(M(\xi))$. Define

$$\mathcal{R}_1 = \text{ convex hull of } f_1(\mathcal{X}) \cup -f_1(\mathcal{X})$$

which is compact, symmetric and convex set, spanning $\mathbb{R}^{k_1}$ and containing the point 0. The following geometric characterization of the c_1-optimal designs is due to Elfing (1952) (see also Pukelsheim (1979) for its geometric interpretation and some examples).

Theorem 1.1. (Elfing (1952)) A design $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{m}$ (for which $c_1^T\theta_1$ is estimable) is c_1-optimal (in the model g_1) if and only if there exist a positive number $\gamma_1 > 0$ and numbers $\varepsilon_{11},\ldots,\varepsilon_{1m}$ with $\varepsilon_{1j}^2 = 1$ ($j = 1,\ldots,m$) such that the point

$$\gamma_1 c_1 = \sum_{\nu=1}^{m} p_\nu \varepsilon_{1\nu} f_1(x_\nu)$$

is a boundary point of the set $\mathcal{R}_1$.

The theory described so far is based on the fact that the underlying model (namely g_1) is known by the experimenter before the experiments are carried out, which is seldom the case in practical applications. We will now consider optimal designs which are robust with respect to a given set of models.

To this end assume that the "true" model belongs to the given set of models

$$\mathcal{F}_n = \{g_1(x),\ldots,g_n(x)\}.$$

For every model we want to estimate a linear combination of the parameters $c_\ell^T\theta_\ell$ ($\ell = 1,\ldots,n$). For example, if we want to estimate the (unknown) regression function $g(x)$ at a given point x_0 we choose $c_\ell = f_\ell(x_0)$ ($\ell = 1,\ldots,n$). Let $\beta_1,\ldots,\beta_n$ denote positive numbers with sum 1, the vector $\beta = (\beta_1,\ldots,\beta_n)$ is called a prior for the class $\mathcal{F}_n$ and β_ℓ reflects the experimenters belief about the adequacy of the model g_ℓ ($\ell = 1,\ldots,n$). Let $c = (c_1^T,\ldots,c_n^T)^T$, a design ξ (for which all linear combinations $c_\ell^T\theta_\ell$ are estimable) is called c-optimal for the class $\mathcal{F}_n$ with respect to the prior β if it minimizes

$$\Phi_\beta(\xi) = \sum_{\ell=1}^{n} \beta_\ell \log\,[c_\ell^T M_\ell^-(\xi)c_\ell].$$

Note that the condition of the estimability of the linear combinations $c_\ell^T\theta_\ell$ implies

$$c_\ell \in \text{ range } M_\ell(\xi) \qquad \text{for } \ell = 1,\ldots,n$$

and that an c-optimal design will allow good estimates of the linear combination $c_\ell^T\theta_\ell$ in every model of $\mathcal{F}_n$.

The outline of the paper is as follows. In section 2 we will give a geometric characterization of the c-optimal designs similar to the characterization of Elfing (1952), which is obtained as a special case of the general theory. In section 3 we will use these results to prove an Elfing type theorem for the D-optimality criterion. Section 4 gives examples to get some insight into the geometric structure of the underlying problems. Finally in the appendix we give proofs of some theorems used in section 2 by an application of the results given in Pukelsheim (1980).

2. The Generalized Elfing Theorem

With the notation of section 1 we have the following equivalent condition for c-optimal designs for the class $\mathcal{F}_n$ with respect to the prior β. The proof involves general arguments of information functionals as given in Pukelsheim (1980) and is deferred to the appendix.

Theorem 2.1. A design ξ (for which $c_\ell^T\theta_\ell$ is estimable $\ell = 1,\ldots,n$) is optimal for the class $\mathcal{F}_n$ with respect to the prior β if and only if there exist generalized inverses $G_1,\ldots,G_n$ of $M_1(\xi),\ldots,M_n(\xi)$ such that

$$\sum_{\ell=1}^{n} \beta_\ell \frac{(c_\ell^T G_\ell f_\ell(x))^2}{c_\ell^T M_\ell^-(\xi)c_\ell} \leq 1$$

for all $x \in \mathcal{X}$. The equal sign in this inequality appears only for the support points of the optimal design.

We will now define a set $\mathcal{R}_n$ corresponding to the set $\mathcal{R}_1$ of the original Elfing theorem. Let $\mathcal{R}_n$ denote the convex hull of the set

$$\{(\varepsilon_1 f_1^T(x), \dots, \varepsilon_n f_n^T(x))^T | \qquad x \in \mathcal{X}, \qquad \sum_{\ell=1}^{n} \beta_\ell \varepsilon_\ell^2 = 1\}$$

which is also a convex, symmetric and compact subset of $\mathbb{R}^{k_1+\dots+k_n}$ containing the point 0. Note that in the case $n = 1$ ($\beta_1 = 1$) this set is exactly the set given in Theorem 1.1. In general the structure of $\mathcal{R}_n$ is very complicated and will be illustrated in some examples of section 4. We are now able to prove an analogous geometric characterization of c-optimal designs for the class $\mathcal{F}_n$ with respect to a prior β as given in Theorem 1.1.

Theorem 2.2. A design $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{m}$ (for which $c_\ell^T \theta_\ell$ is estimable $\ell = 1, \dots, n$) is c-optimal for the class $\mathcal{F}_n$ with respect to the prior β if and only if there exist positive numbers $\gamma_1, \dots, \gamma_n$ and numbers $\varepsilon_{11}, \dots, \varepsilon_{1m}$, $\varepsilon_{21}, \dots, \varepsilon_{2m}, \dots, \varepsilon_{n1}, \dots, \varepsilon_{nm}$ such that

(A)
$$\gamma_\ell c_\ell = \sum_{\nu=1}^{m} p_\nu \varepsilon_{\ell\nu} f_\ell(x_\nu) \qquad \ell = 1, \dots, n$$

(B) The point $(\gamma_1 c_1^T, \dots, \gamma_n c_n^T)^T$ is a boundary point of the set $\mathcal{R}_n$ with a supporting hyperplane $(a_1^T, \dots, a_n^T)^T$.

(C)
$$\gamma_\ell c_\ell^T a_\ell = \beta_\ell \qquad \ell = 1, \dots, n$$

(D)
$$\sum_{\ell=1}^{n} \beta_\ell \varepsilon_{\ell\nu}^2 = 1 \qquad \nu = 1, \dots, m$$

Proof: To prove the necessity of the theorem we will use the equivalence theorem 2.1. Let $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{m}$ denote an optimal design for the class $\mathcal{F}_n$ with respect to the prior β. By Theorem 2.1. there exist generalized inverses of $M_1(\xi), \dots, M_n(\xi)$ such that

(2.1)
$$\sum_{\ell=1}^{n} \beta_\ell \frac{(c_\ell^T G_\ell f_\ell(x))^2}{c_\ell^T M_\ell^-(\xi) c_\ell} \leq 1 \qquad \text{for all } x \in \mathcal{X}$$

and

$$(2.2) \qquad \sum_{\ell=1}^{n} \beta_\ell \frac{(c_\ell^T G_\ell f_\ell(x_\nu))^2}{c_\ell^T M_\ell^-(\xi) c_\ell} = 1 \qquad \text{for all } \nu = 1, \ldots, m.$$

Let $\gamma_\ell^{-2} = c_\ell^T M_\ell^-(\xi) c_\ell$ and $d_\ell = \gamma_\ell G_\ell c_\ell$ ($\ell = 1, \ldots, n$) then we have from the estimability of $c_\ell^T \theta_\ell$ by the design ξ

$$\gamma_\ell c_\ell = M_\ell(\xi) d_\ell = \sum_{\nu=1}^{m} p_\nu \varepsilon_{\ell\nu} f_\ell(x_\nu) \qquad \ell = 1, \ldots, m$$

where $\varepsilon_{\ell\nu} = f_\ell^T(x_\nu) d_\ell$ ($\ell = 1, \ldots n$, $\nu = 1, \ldots, m$). This proves the representation given in (A). Equation (2.2) and the representation of $\gamma_\ell c_\ell$ yield

$$(2.3) \qquad \sum_{\ell=1}^{n} \beta_\ell \gamma_\ell c_\ell^T d_\ell = \sum_{\ell=1}^{n} \beta_\ell (d_\ell^T f_\ell(x_\nu))^2 = \sum_{\ell=1}^{n} \beta_\ell \varepsilon_{\ell\nu}^2 = 1$$

which shows condition (D). From the inequality (2.1) and the Cauchy-Schwarz inequality we get

$$\left(\sum_{\ell=1}^{n} \beta_\ell (d_\ell^T \varepsilon_\ell f_\ell(x)) \right)^2 \leq \sum_{\ell=1}^{n} \beta_\ell \varepsilon_\ell^2 \cdot \sum_{\ell=1}^{n} \beta_\ell (d_\ell^T f_\ell(x))^2 \leq 1$$

for all $x \in \mathcal{X}$, whenever the numbers $\varepsilon_1, \ldots, \varepsilon_n$ satisfy the equation $\sum_{\ell=1}^{n} \beta_\ell \varepsilon_\ell^2 = 1$. Observing (2.3) we thus see that the point $(\gamma_1 c_1^T, \ldots, \gamma_n c_n^T)^T$ is a boundary point with supporting hyperplane $(\beta_1 d_1^T, \ldots, \beta_n d_n^T)^T$ which proves (B). Finally the condition (C) follows readily from the definition of γ_ℓ and d_ℓ.

To prove sufficiency let $(a_1^T, \ldots, a_n^T)^T$ denote a supporting hyperplane of $\mathcal{R}_n$ at the boundary point $(\gamma_1 c_1^T, \ldots, \gamma_n c_n^T)^T$ and let $a_\ell = \beta_\ell d_\ell$ ($\ell = 1, \ldots, n$). Thus we have for all $x \in \mathcal{X}$, $(\varepsilon_1, \ldots, \varepsilon_n)^T$ with $\sum_{\ell=1}^{n} \beta_\ell \varepsilon_\ell^2 = 1$

$$(2.4) \qquad \left| \sum_{\ell=1}^{n} (\beta_\ell d_\ell)^T (\varepsilon_\ell f_\ell(x)) \right| \leq 1$$

We will now show that (2.4) implies the inequality

$$(2.5) \qquad \sum_{\ell=1}^{n} \beta_\ell (d_\ell^T f_\ell(x))^2 \leq 1 \qquad \text{for all } x \in \mathcal{X}$$

To do this we have to distinguish two cases. At first consider a $x \in \mathcal{X}$ for which $d_\ell^T f_\ell(x) = 0$ $(\ell = 1, \dots, n)$, in this case (2.5) is obvious. In the other case define

$$\varepsilon_\ell(x) = \frac{d_\ell^T f_\ell(x)}{\sqrt{\sum\limits_{\ell=1}^n \beta_\ell (d_\ell^T f_\ell(x))^2}} \qquad \ell = 1, \dots, n,$$

then it follows that $\sum\limits_{\ell=1}^n \beta_\ell \varepsilon_\ell^2(x) = 1$ and (2.4) must hold for this vector which yields

$$1 \geq \left[\sum_{\ell=1}^n (\beta_\ell d_\ell)^T (\varepsilon_\ell(x) f_\ell(x)) \right]^2 = \sum_{\ell=1}^n \beta_\ell (d_\ell^T f_\ell(x))^2$$

and proves (2.5). Because $(\beta_1 d_1^T, \dots, \beta_n d_n^T)^T$ is a supporting hyperplane at the point $(\gamma_1 c_1^T, \dots, \gamma_n c_n^T)^T$ we obtain from (2.4) (used at $x = x_\nu$) and the representation (A)

$$1 = \sum_{\ell=1}^n \beta_\ell \gamma_\ell c_\ell^T d_\ell = \sum_{\nu=1}^m p_\nu \sum_{\ell=1}^n \varepsilon_{\ell\nu} \beta_\ell f_\ell^T(x_\nu) d_\ell \leq 1$$

which implies (note that $| \sum\limits_{\ell=1}^n \beta_\ell \varepsilon_{\ell\nu} f_\ell^T(x_\nu) d_\ell | \leq 1$)

$$\sum_{\ell=1}^n \beta_\ell \varepsilon_{\ell\nu} f_\ell^T(x_\nu) d_\ell = 1 \qquad (\nu = 1, \dots, m)$$

By an application of the Cauchy Schwarz inequality we now get for $\nu = 1, \dots, m$

$$1 = (\sum_{\ell=1}^n \beta_\ell \varepsilon_{\ell\nu} f_\ell^T(x_\nu) d_\ell)^2 \leq \sum_{\ell=1}^n \beta_\ell \varepsilon_{\ell\nu}^2 \sum_{\ell=1}^n \beta_\ell (d_\ell^T f_\ell(x_\nu))^2 \leq 1 \tag{2.6}$$

where the last inequality results from (2.5) and condition (D). Therefore we have equality in the Cauchy-Schwarz inequality and it follows

$$\sqrt{\beta_\ell} \varepsilon_{\ell\nu} = \lambda_\nu \sqrt{\beta_\ell}\, d_\ell^T f_\ell(x_\nu) \qquad \ell = 1, \dots, n, \;\; \nu = 1, \dots, m$$

or equivalently

$$\varepsilon_{\ell\nu} = \lambda_\nu d_\ell^T f_\ell(x_\nu) \qquad \ell = 1, \dots, n, \;\; \nu = 1, \dots, m.$$

From the normalizing conditions on the $\varepsilon_{\ell\nu}$ in (D) we obtain observing (2.6) $(\nu = 1, \dots, m)$

$$1 = \sum_{\ell=1}^n \beta_\ell \varepsilon_{\ell\nu}^2 = \lambda_\nu^2 \sum_{\ell=1}^n \beta_\ell (d_\ell^T f_\ell(x_\nu))^2 = \lambda_\nu^2. \tag{2.7}$$

On the other hand we have from the property that $(\gamma_1 c_1^T, \ldots, \gamma_n c_n^T)^T$ is a boundary point of $\mathcal{R}_n$ with supporting hyperplane $(\beta_1 d_1^T, \ldots, \beta_n d_n^T)^T$

$$\begin{aligned} 1 = \sum_{\ell=1}^{n} \beta_\ell \gamma_\ell c_\ell^T d_\ell &= \sum_{\nu=1}^{m} p_\nu \sum_{\ell=1}^{n} \beta_\ell \varepsilon_{\ell\nu} (d_\ell^T f_\ell(x_\nu)) \\ &= \sum_{\nu=1}^{m} p_\nu \lambda_\nu \sum_{\ell=1}^{m} \beta_\ell (d_\ell^T f_\ell(x_\nu))^2 = \sum_{\nu=1}^{m} p_\nu \lambda_\nu. \end{aligned}$$

Equation (2.7), $p_\nu \geq 0$, $\sum_{\nu=1}^{m} p_\nu = 1$ now show that $\lambda_\nu = 1$ whenever $p_\nu > 0$ and this implies

$$\varepsilon_{\ell\nu} = d_\ell^T f_\ell(x_\nu) \qquad \ell = 1, \ldots, n, \quad \nu = 1, \ldots, m$$

where we have assumed that in the representation (A) all p_ν are positive (without loss of generality). From this representation we thus obtain for $\ell = 1, \ldots, n$

$$\gamma_\ell c_\ell = \sum_{\nu=1}^{m} p_\nu \varepsilon_{\ell\nu} f_\ell(x_\nu) = \sum_{\nu=1}^{m} p_\nu f_\ell(x_\nu) f_\ell^T(x_\nu) d_\ell = M_\ell(\xi) d_\ell$$

From the definition of a generalized inverse (see Searle (1982), p. 238) it follows that there exist generalized inverses $G_1, \ldots, G_n$ of the matrices $M_1(\xi), \ldots, M_n(\xi)$ such that

$$d_\ell = \gamma_\ell G_\ell c_\ell \qquad \ell = 1, \ldots, n.$$

By the condition (D) we thus have (note that $a_\ell = \beta_\ell d_\ell$ and that $c_\ell^T G_\ell c_\ell$ is invariant with respect to the choice of the generalized inverse because $c_\ell^T \theta_\ell$ is estimable)

$$1 = \gamma_\ell c_\ell^T d_\ell = \gamma_\ell^2 c_\ell^T G_\ell c_\ell = \gamma_\ell^2 c_\ell^T M_\ell^-(\xi) c_\ell \qquad (\ell = 1, \ldots, n)$$

and the inequality (2.5) yields that there exist generalized inverses $G_1, \ldots, G_n$ such that

$$\sum_{\ell=1}^{n} \beta_\ell \frac{(c_\ell^T G_\ell f_\ell(x))^2}{c_\ell^T M_\ell^-(\xi) c_\ell} \leq 1 \qquad \text{for all } x \in \mathcal{X}.$$

By an application of Theorem 2.1 we now see that the design ξ is optimal for the class $\mathcal{F}_n$ with respect to the prior β, which completes the proof of Theorem 2.2.

Note that Theorem 2.2 reduces to the original Elfing theorem if $n = 1$ ($\beta_1 = 1$) and that in this case the condition (C) is obvious while condition (D) yields $\varepsilon_{1\nu} = \mp 1$

($\nu = 1, \ldots, m$). The Theorem can easily be transferred to the case where the experimenter is not only interested to estimate one linear combination $c_\ell^T \theta_\ell$ in the model g_ℓ but several combinations, i.e. $A_\ell^T \theta_\ell$ for some matrix $A_\ell \in \mathbb{R}^{k_\ell \times s_\ell}$ ($\ell = 1, \ldots, n$). Let $A = (A_1^T, \ldots, A_n^T)^T$, a design ξ is called A optimal for the class $\mathcal{F}_n$ with respect to the prior β if it minimizes

$$\sum_{\ell=1}^{n} \beta_\ell \log \left[\text{trace } \{M_\ell^-(\xi) A_\ell A_\ell^T\}\right]$$

By an application of an analogous equivalence theorem as given in Theorem 2.1 (see Theorem 5.8 in the appendix) we can prove a similar geometric characterization for A optimal designs for the class $\mathcal{F}_n$ with respect to the prior β. The proceeding is the same as in the proof of Theorem 2.2 and the details therefore omitted. Note that this result contains the Elfing theorem of Studden (1971) for A-optimal designs as the special case $n = 1$, $\beta_1 = 1$.

Theorem 2.3. Let $\mathcal{R}_n$ denote the convex hull of the set

$$\{((f_1(x)\varepsilon_1)^T, \ldots, (f_n(x)\varepsilon_n)^T)^T | \ \ x \in \mathcal{X}, \ \ \varepsilon_\ell \in \mathbb{R}^{1 \times s_\ell}, \ \ \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (\varepsilon_\ell^T \varepsilon_\ell) = 1\}$$

The design $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{m}$ is A-optimal for the class $\mathcal{F}_n$ with respect to the prior β if and only if there exist positive numbers $\gamma_1, \ldots, \gamma_n$ and vectors $\varepsilon_{\ell\nu} \in \mathbb{R}^{1 \times s_\ell}$ ($\ell = 1, \ldots, n$, $\nu = 1, \ldots, m$) such that

(A) $$\gamma_\ell A_\ell = \sum_{\nu=1}^{m} p_\nu f_\ell(x_\nu) \varepsilon_{\ell\nu} \qquad \ell = 1, \ldots, n$$

(B) There exists a "supporting hyperplane" $D = (D_1^T, \ldots, D_n^T)^T$ ($D_\ell \in \mathbb{R}^{k_\ell \times s_\ell}$) of $\mathcal{R}_n$ at the "boundary point" $(\gamma_1 A_1^T, \ldots, \gamma_n A_n^T)^T$ such that

(1) $$\sum_{\ell=1}^{n} \text{trace } \gamma_\ell D_\ell^T A_\ell = 1 \qquad (\ell = 1, \ldots, n)$$

(2) $$\sum_{\ell=1}^{n} \text{trace } (D_\ell^T f_\ell(x) \varepsilon_\ell) \leq 1$$

for all $x \in \mathcal{X}$ and $\varepsilon_\ell \in \mathbb{R}^{1 \times s_\ell}$ with $\sum_{\ell=1}^{n} \beta_\ell \text{ trace } (\varepsilon_\ell^T \varepsilon_\ell) = 1$

(C) $$\gamma_\ell \text{ trace } D_\ell^T A_\ell = \beta_\ell \qquad (\ell = 1, \ldots, n)$$

(D) $$\sum_{\ell=1}^{n} \beta_\ell \text{ trace } (\varepsilon_{\ell\nu}^T \varepsilon_{\ell\nu}) = 1 \qquad \nu = 1, \dots, m$$

It is obvious that Theorem 2.3 gives Theorem 2.2 for the case $s_\ell = 1$ $\ell = 1, \dots, n$. However, the set $\mathcal{R}_n$ in this theorem is a subset of a very complicated space (namely $\mathbb{R}^{k_1 \times s_1} \times \mathbb{R}^{k_2 \times s_2} \times \dots \times \mathbb{R}^{k_n \times s_n}$) and the geometric structure of the derived results is more intuitive for the case considered in Theorem 2.2, which is the reason why we have given the proof for the last named Theorem. This Theorem will also allow to give an Elfing characterization for the D-optimality criterion, which is considered in the following section.

3. A Characterization of D-optimality

In this section we will investigate a special case of Theorem 2.2, which is of particular interest because it will give a geometric characterization of Elfing type for the D-optimal design problem. To this end consider the "nested" models

(3.1) $$\begin{cases} f_1^T(x) &= f_{11}(x) \\ f_2^T(x) &= (f_{11}(x), f_{12}(x)) \\ \vdots & \\ f_n^T(x) &= (f_{11}(x), f_{12}(x), \dots, f_{1n}(x)) \end{cases}$$

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

(3.2) $$\Phi_\beta(\xi) = -\sum_{\ell=1}^{n} \beta_\ell \log \frac{\det M_\ell(\xi)}{\det M_{\ell-1}(\xi)},$$

and for the prior $\beta_1 = \dots = \beta_n = \frac{1}{n}$ we obtain the D-optimality criterion. Thus we have (by an application of Theorem 2.2) the following geometric characterization of D-optimal designs.

Theorem 3.1. A design $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{m}$ is D-optimal for the model $g_n(x) = \alpha_1 f_{11}(x) + \ldots + \alpha_n f_{1n}(x)$ if and only if there exist positive numbers $\gamma_\ell > 0$ $(\ell = 1, \ldots, n)$ and numbers $\varepsilon_{11}, \ldots, \varepsilon_{1m}, \ldots, \varepsilon_{n1}, \ldots, \varepsilon_{nm}$ such that

(A)
$$\gamma_\ell \tilde{c}_\ell = (0, \ldots, 0, \gamma_\ell)^T = \sum_{\nu=1}^{m} p_\nu \varepsilon_{\ell\nu} f_\ell(x_\nu) \qquad \ell = 1, \ldots, n$$

(B) The point $(\gamma_1 \tilde{c}_1^T, \ldots, \gamma_n \tilde{c}_n^T)^T = (\gamma_1, 0, \gamma_2, 0, 0, \gamma_3, \ldots, 0, \ldots, 0, \gamma_n)^T$ is a boundary point of the convex hull of the set

$$\mathcal{R}_n = \{(\varepsilon_1 f_1^T(x), \ldots, \varepsilon_n f_n^T(x))^T | \qquad x \in \mathcal{X}, \ \sum_{\ell=1}^{n} \varepsilon_\ell^2 = n\} \ \subseteq \ \mathbb{R}^{n(n+1)/2}$$

with supporting hyperplane $\frac{1}{n}\ (d_1^T, d_2^T, \ldots, d_n^T)^T$, $d_\ell^T = (d_{\ell 1}, \ldots, d_{\ell n})$

(C)
$$\gamma_\ell d_{\ell\ell} = 1 \qquad \text{for } \ell = 1, \ldots, n$$

(D)
$$\sum_{\ell=1}^{n} \varepsilon_{\ell\nu}^2 = n \qquad \text{for } \nu = 1, \ldots, m.$$

Examples to illustrate the theorem are given in the next section. We will finish this section proving a certain orthogonality condition of linear combinations of the functions given in (3.1). To this end consider the optimality criterion in (3.2) and an optimal design with respect to this criterion (i.e. an c-optimal design for the class $\mathcal{F}_n$ with respect to the prior β where the functions $f_\ell^T(x)$ are given by (3.1) and $c = (\tilde{c}_1^T, \ldots, \tilde{c}_n^T)^T$ with $\tilde{c}_\ell^T = (0, \ldots, 0, 1)^T$). Define as in the proof of Theorem 2.2 (note that for the vector $c = (\tilde{c}_1^T, \ldots, \tilde{c}_n^T)^T$ we may assume that all information matrices $M_\ell(\xi_\beta)$ of the optimal design ξ_β are non singular)

$$\gamma_\ell^{-2} = \tilde{c}_\ell^T M_\ell^{-1}(\xi_\beta) \tilde{c}_\ell, \qquad d_\ell = \gamma_\ell M_\ell^{-1}(\xi_\beta) \tilde{c}_\ell \qquad (\ell = 1, \ldots, n)$$

and consider the "polynomials"

(3.3)
$$P_\ell(x) = d_\ell^T f_\ell(x) \qquad \ell = 1, \ldots, n.$$

Then we have

$$\int_{\mathcal{X}} P_\ell(x) f_\ell^T(x) d\xi_\beta(x) = d_\ell^T \int_{\mathcal{X}} f_\ell(x) f_\ell^T(x) d\xi_\beta(x) =$$
$$\gamma_\ell \tilde{c}_\ell^T M_\ell^{-1}(\xi) M_\ell(\xi) = \gamma_\ell \tilde{c}_\ell = (0, \ldots, 0, \gamma_\ell)$$

and

$$\int_{\mathcal{X}} P_\ell^2(x) d\xi_\beta(x) = d_\ell^T M_\ell(\xi_\beta) d_\ell = \gamma_\ell^2 \tilde{c}_\ell^T M_\ell \tilde{c}_\ell = 1.$$

Thus we have proved the following theorem.

Proposition 3.2. The "polynomials" $\{P_\ell(x)\}_{\ell \geq 0}$ defined by (3.3) are orthonormal with respect to the measure $d\xi_\beta(x)$, where ξ_β is the c-optimal design for the class $\mathcal{F}_n$ with respect to the prior β and $c = (\tilde{c}_1^T, \ldots, \tilde{c}_n^T)^T$ ($\tilde{c}_\ell = (0, \ldots, 0, 1)^T$).

4. Examples

In this section we will give some examples to get more insight into the geometric structure of the optimal design problems investigated in section 2 and 3.

Example 4.1. (Linear or quadratic regression through the origin). Let $n = 2$, $k_1 = 1$, $k_2 = 1$, $f_1(x) = x$, $f_2(x) = x^2$ and $\mathcal{X} = [0, 1]$, by the definition of $\mathcal{R}_2$ as the convex hull of the set

$$\{(\varepsilon_1 x, \varepsilon_2 x^2)^T | \ x \in [0, 1], \qquad \beta_1 \varepsilon_1^2 + \beta_2 \varepsilon_2^2 = 1\}$$

we have that $\mathcal{R}_2$ is given by the ellipsoid $\mathcal{R}_2 = \{(x, y)^T | \ \beta_1 x^2 + \beta_2 y^2 \leq 1\}$. Now let $c_1 = 2$, $c_2 = 4$ (suppose that we want to estimate $f_1(x)$ or $f_2(x)$ at the point $x_0 = 2$), $\gamma_1 = \frac{1}{2}$, $\gamma_2 = \frac{1}{4}$ and ξ_β the design which concentrates mass 1 at the point 1. Thus we have

$$\begin{pmatrix} \gamma_1 c_1 \\ \gamma_2 c_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = p_1 \begin{pmatrix} \varepsilon_1 f_1(1) \\ \varepsilon_2 f_2(1) \end{pmatrix}$$

where $p_1 = 1$, $\varepsilon_1 = \varepsilon_2 = 1$. The supporting hyperplane of $\mathcal{R}_2$ at the point $(1, 1)^T$ is given by (β_1, β_2) and the conditions of Theorem 2.2 readily checked which shows that the c-optimal design for the class $\mathcal{F}_2$ is supported at the point 1 (independent of the prior β).

Example 4.2. Let $n = 2$, $k_1 = k_2 = 1$, $f_1(x) = x$, $f_2(x) = 1 - x$ and $\mathcal{X} = [0,1]$. For a prior $\beta = (\beta_1, \beta_2)$ the set $\mathcal{R}_2$ is the convex hull of the set

$$\left\{ \begin{pmatrix} 0 \\ 1/\sqrt{\beta_2} \end{pmatrix}, \begin{pmatrix} 0 \\ -1\sqrt{\beta_2} \end{pmatrix}, \begin{pmatrix} 1/\sqrt{\beta_1} \\ 0 \end{pmatrix}, \begin{pmatrix} -1/\sqrt{\beta_1} \\ 0 \end{pmatrix} \right\}$$

which is given in Figure 4.1.

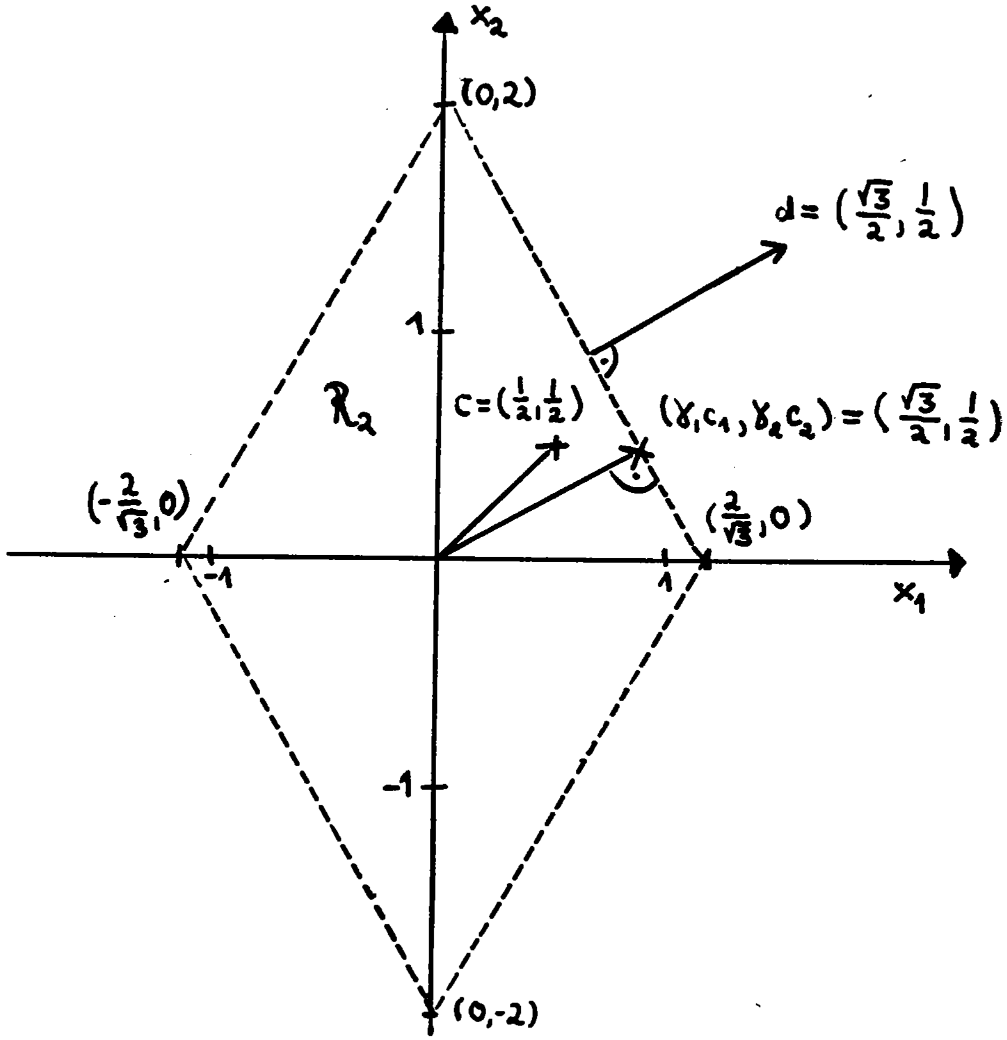

Figure 4.1. The set $\mathcal{R}_2$ for $f_1(x) = x$, $f_2(x) = 1 - x$, $\beta_1 = 3/4$, $\beta_2 = 1/4$

Suppose we want to estimate one of the regression functions $f_1(x)$ or $f_2(x)$ at the point $x_0 = \frac{1}{2}$. Thus we choose $c_1 = c_2 = \frac{1}{2}$. From Figure 4.1 it is evident that the point

$$\begin{pmatrix} 2\sqrt{\beta_1} \cdot \frac{1}{2} \\ 2\sqrt{\beta_2} \cdot \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta_1} \\ \sqrt{\beta_2} \end{pmatrix} = \beta_2 \cdot \begin{pmatrix} 0 \\ \frac{1}{\sqrt{\beta_2}} 1 \end{pmatrix} + \beta_1 \begin{pmatrix} \frac{1}{\sqrt{\beta_1}} 1 \\ 0 \end{pmatrix}$$

is a boundary point of $\mathcal{R}_2$ with supporting hyperplane $d = (\sqrt{\beta_1}, \sqrt{\beta_2})^T$ (note that the boundary of $\mathcal{R}_2$ in this part consists of the set $\{(x, -\sqrt{\beta_1/\beta_2}(x - 1/\sqrt{\beta_1}))^T | x \in [0, 1/\sqrt{\beta_1}]\})$. The quantities in the representation (A) of Theorem 2.2 are given by $p_1 = \beta_2$ $p_2 = \beta_1$, $x_1 = 0$, $x_2 = 1$, $\varepsilon_{11} = 0$, $\varepsilon_{12} = 1/\sqrt{\beta_1}$ $\varepsilon_{21} = 1\sqrt{\beta_2}$, $\varepsilon_{22} = 0$. The conditions (C) and (D) are easily checked and we conclude that the c-optimal design for the class $\mathcal{F}_2 = \{x, 1-x\}$ is supported at the points 0 and 1 with masses β_2 and β_1.

Example 4.3. In this example we want to determine the D-optimal design for the model $g_2(x) = \alpha_1(1-x) + \alpha_2 x^2$ where $x \in [0,1]$. To this end let $n = 2$, $k_1 = 1$, $k_2 = 2$ $f_1^T(x) = 1 - x$ and $f_2^T(x) = (1-x, x^2)$. We will solve the more general problem of determining a c-optimal design for the class $\mathcal{F}_2$ with respect to the prior $\beta = (1/2, 1/2)$ where $c = (c_1, c_2^T)^T = (1, h_1, h_2)$. The set $\mathcal{R}_2$ defined in section 2 is given by

$$\mathcal{R}_2 = \{(x_1, x_2, x_3)^T | \ x_1^2 + x_2^2 \le 2, \ |x_3| \le \sqrt{2} - \sqrt{x_1^2 + x_2^2}\} \tag{4.1}$$

and is depicted in Figure 4.2.

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 $c = (c_1^T, c_2^T)^T = (1, h_1, 0)^T$ which shows that the vector $(\gamma_1 c_1^T, \gamma_2 c_2^T)$ can only intersect the boundary of $\mathcal{R}_2$ at the curve $\mathcal{K} = \{(x_1, x_2, 0)^T | \ x_1^2 + x_2^2 = 2\}$ which is obtained from the point $(f_1^T(0), f_2^T(0))^T$, i.e.

$$\mathcal{K} = \left\{ \begin{pmatrix} \varepsilon_1 f_1(0) \\ \varepsilon_2 f_2(0) \end{pmatrix} | \ \varepsilon_1^2 + \varepsilon_2^2 = 2 \right\}.$$

Therefore the c-optimal design (for the vector $c = (1, h_1, 0)^T$) puts mass 1 at the point 0.

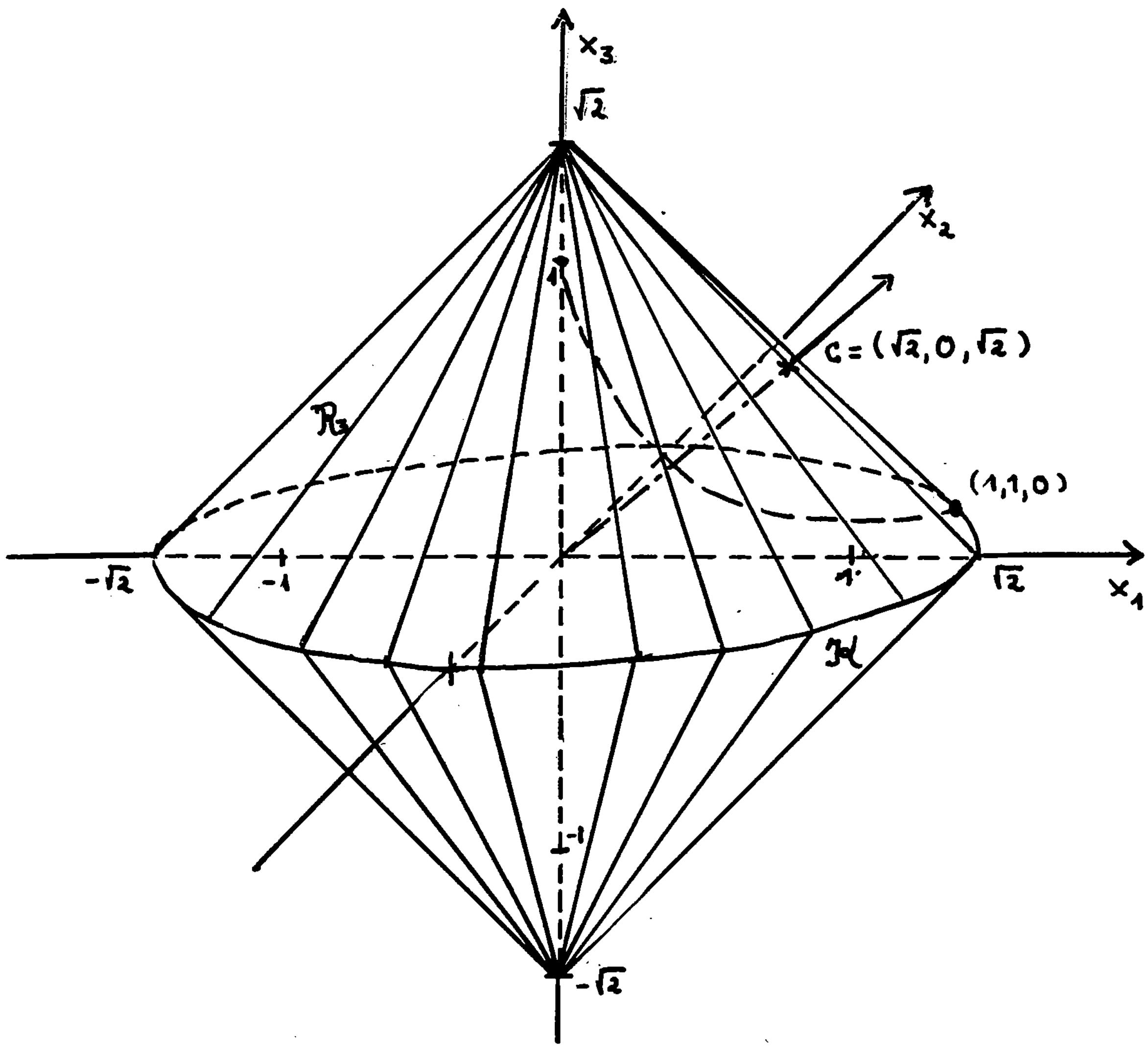

Figure 4.2. The set $\mathcal{R}_2$ for the models $f_1^T(x) = 1 - x$, $f_2^T(x) = (1 - x, x^2)$

(B) $h_2 \neq 0$: In this case the vector $(\gamma_1 c_1^T, \gamma_2 c_2^T)^T$ touches the boundary of $\partial\mathcal{R}_2$ at the some point $t \in \partial\mathcal{R}_2$ which is a convex combination of a point of $\mathcal{K}$ and one of the points $(0, 0, \sqrt{2})^T$ or $(0, 0, -\sqrt{2})^T$ which depends on the sign of h_2. Therefore we see that the c-optimal design is supported at the points 0 and 1. 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 unknown quantities γ_1, γ_2, ε_{11}, ε_{12}, ε_{21}, ε_{22} used in Theorem 2.2. First we remark that the condition $(\gamma_1 c_1^T, \gamma_2 c_2^T)^T \in \partial\mathcal{R}_2$ implies

$$\gamma_2 h_2 = \sqrt{2} - \sqrt{\gamma_1^2 + \gamma_2^2 h_1^2}. \tag{4.2}$$

Let $d = \frac{1}{2}(d_1, d_2, d_3)^T$ denote the supporting hyperplane of $\mathcal{R}_2$ at the point $(\gamma_1 c_1^T, \gamma_2 c_2^T)^T = (\gamma_1, \gamma_2 h_1, \gamma_2 h_2)^T$ then we have from the condition (C) of Theorem 2.2

$$d = \begin{pmatrix} \gamma_1^{-1} \\ d_2 \\ (1 - \gamma_2 h_1 d_2)/(\gamma_2 h_2) \end{pmatrix} \qquad \text{for some } d_2 \in \mathbb{R} \tag{4.3}$$

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

$$d^T \left[\frac{\sqrt{2}}{\sqrt{\gamma_1^2 + \gamma_2^2 h_1^2}} \begin{pmatrix} \gamma_1 \\ \gamma_2 h_1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ \sqrt{2} \end{pmatrix} \right] = 0$$

which implies

$$d_2 = \frac{\sqrt{\gamma_1^2 + \gamma_2^2 h_1^2} - \gamma_2 h_2}{\gamma_2 h_1 [\sqrt{\gamma_1^2 + \gamma_2^2 h_1^2} + \gamma_2 h_2]}.$$

From (4.3) and (4.2) we thus obtain for the supporting hyperplane $\frac{1}{2}d$ at the point $(\gamma_1 c_1^T, \gamma_2 c_2^T)^T$ by straight forward calculations

$$d = (\gamma_1^{-1}, \frac{1 - \sqrt{2}\gamma_2 h_2}{\gamma_2 h_1}, \sqrt{2})^T = (d_1^T, d_2^T)^T$$

and by the definition $\varepsilon_{\ell\nu} = d_\ell^T f_\ell(x_\nu)$ (compare with the proof of Theorem 2.2) we have

$$\begin{array}{lllll} \varepsilon_{11} = d_1 f_1(0) = \gamma_1^{-1} & & \varepsilon_{21} = d_2^T f_2(0) = \frac{1-\sqrt{2}\gamma_2 h_2}{\gamma_2 h_1} \\ \varepsilon_{12} = d_1 f_1(1) = 0 & & \varepsilon_{22} = d_2^T f_2(1) = \sqrt{2} \end{array}$$

From condition (A) of Theorem 2.2 we have the equation

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 h_1 \\ \gamma_2 h_2 \end{pmatrix} = p_1 \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ 0 \end{pmatrix} + p_2 \begin{pmatrix} 0 \\ 0 \\ \varepsilon_{22} \end{pmatrix}$$

which yields $p_1 = \gamma_1^2$, $p_2 = \frac{\gamma_2 h_2}{\sqrt{2}}$, $p_1 = \frac{(\gamma_2 h_1)^2}{1-\sqrt{2}\gamma_2 h_2}$ and $p_1 + p_2 = 1$. From these equations we get

$$\gamma_2 h_2 = \sqrt{2}(1 - \gamma_1^2) \tag{4.4}$$

$$p_1 = \gamma_1^2, \quad p_2 = 1 - \gamma_1^2 \tag{4.5}$$

$$2\gamma_1^4\left[1 - \frac{h_1^2}{h_2^2}\right] - \gamma_1^2\left[1 - 4\frac{h_1^2}{h_2^2}\right] - 2\frac{h_1^2}{h_2^2} = 0 \tag{4.6}$$

From (4.3), (4.4) and (4.5) it can easily be shown that $\frac{1}{2}(d_1, d_2, d_3) \cdot (1, h_1, h_2)^T = 1$ and that the $\varepsilon_{\ell\nu}$ satisfy condition (D) of Theorem 2.2 (note that we have used all other conditions of this theorem to derive (4.3), (4.4) and (4.5)). Thus we see that the c-optimal design for the class $\mathcal{F}_2$ with respect to the prior $\beta = (1/2, 1/2)$ ($c = (1, h_1, h_2)^T$) 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 the equation

$$2\gamma_1^4(h_2^2 - h_1^2) - \gamma_1^2(h_2^2 - 4h_1^2) - 2h_1^2 = 0. \tag{4.7}$$

Note that this result also includes the case $h_2 = 0$ for which (4.7) reduces to $(\gamma_1^2 - 1)^2 = 0$. For the vector $c = (1, 0, 1)$ ($h_1 = 0_1, h_2 = 1$) we obtain the D-optimal design for the model $\alpha_1(1-x) + \alpha_2 x^2$ (compare with Theorem 3.1) which puts equal masses at the points 0 and 1.

Example 4.4. We will now show that the condition (C) is necessary to obtain the equivalence of Theorem 2.2. To this end let $n = 2$, $\beta_1 = \beta_2 = \frac{1}{2}$, $k_1 = 2, k_2 = 3$, $f_1^T(x) = (1, x)$, $f_2^T(x) = (1, x, x^2)$, $c^T = (c_1^T, c_2^T)$ where $c_1^T = (1, 2)$ and $c_2^T = (1, 2, 4)$ (thus we want to estimate a linear or quadratic regression 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 $\varepsilon_{11} = -\sqrt{3/2}$, $\varepsilon_{12} = 0$, $\varepsilon_{13} = \sqrt{3/2}$, $\varepsilon_{21} = 1/\sqrt{2}$, $\varepsilon_{22} = -\sqrt{2}$ and $\varepsilon_{23} = 1/\sqrt{2}$, then we have the following representations

$$\varepsilon_{11}\frac{2}{11}\begin{pmatrix}1\\-1\end{pmatrix} + \varepsilon_{12}\cdot\frac{3}{11}\begin{pmatrix}1\\0\end{pmatrix} + \varepsilon_{13}\frac{6}{11}\begin{pmatrix}1\\1\end{pmatrix} = \frac{4}{11}\sqrt{\frac{3}{2}}\begin{pmatrix}1\\2\end{pmatrix} = \gamma_1 c_1$$

$$\varepsilon_{21}\frac{2}{11}\begin{pmatrix}1\\-1\\1\end{pmatrix} + \varepsilon_{22}\frac{3}{11}\begin{pmatrix}1\\0\\0\end{pmatrix} + \varepsilon_{23}\frac{6}{11}\begin{pmatrix}1\\1\\1\end{pmatrix} = \frac{1}{11}\sqrt{2}\begin{pmatrix}1\\2\\4\end{pmatrix} = \gamma_2 c_2$$

where $\gamma_1 = \frac{4}{11}\sqrt{\frac{3}{2}}$, $\gamma_2 = \frac{\sqrt{2}}{11}$. Thus the design ξ satisfies the conditions (A) and (D) of Theorem 2.2. To prove that the point $(\gamma_1 c_1^T, \gamma_2 c_2^T)^T$ is a boundary point of $\mathcal{R}_2 \subseteq \mathbb{R}^5$ define $d = (d_1^T, d_2^T)^T$ where

$$d_1 = \sqrt{\frac{3}{2}}\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad d_2 = \frac{3}{\sqrt{2}}\begin{pmatrix} -2/3 \\ 0 \\ 1 \end{pmatrix}$$

then we have $\gamma c^T d = \gamma_1 c_1^T d_1 + \gamma_2 c_2^T d_2 = 2$ and by the Cauchy Schwarz inequality

$$(\varepsilon_1 f_1^T(x) d_1) + (\varepsilon_2^T f_2(x) d_2) \leq (\varepsilon_1^2 + \varepsilon_2^2)[(f_1^T(x)d_1)^2 + (f_2^T(x)d_2)^2]$$

$$= \frac{3}{2}x^2 + \frac{9}{2}(x^2 - 2/3)^2 = \frac{9}{2}(x^2 - \frac{1}{2})^2 + \frac{7}{8} \leq 2$$

whenever $\varepsilon_1^2 + \varepsilon_2^2 = 2$ and $x \in [-1,1]$. Therefore the point $\gamma c_1^T = (\gamma_1 c^T, \gamma_2 c_2^T)$ is a boundary point of the set $\mathcal{R}_2$ with supporting hyperplane $\frac{1}{2}d = (\frac{1}{2}d_1, \frac{1}{2}d_2)^T$. Thus all conditions of Theorem 2.2 are satisfied except for condition (C) (for example $\frac{1}{2}d_1^T \gamma_1 c_1 = \frac{12}{11}$). If the design ξ would be c-optimal with respect to the prior $\beta = (\frac{1}{2}, \frac{1}{2})$ we obtain from Theorem 2.1 and straight forward algebra the inequality

$$\frac{1}{2}\sum_{\ell=1}^{2} \frac{(c_\ell^T M_\ell^{-1}(\xi) f_\ell(x))^2}{c_\ell^T M_\ell^{-1}(\xi) c_\ell} = \frac{11}{16}x^2 + \frac{11}{10}(\frac{3}{2}x^2 - 1) \leq 1$$

for all $x \in [-1,1]$. But for $x = 0$ this inequality does not hold and thus ξ is not a c-optimal design for $\mathcal{F}_n$ which shows that condition (C) of Theorem 2.2 can not be omitted.

5. Appendix (General equivalence theorems for model robust designs).

In this section we will give a proof of the equivalence theorem 2.2. To this end we will derive similar results as given in Pukelsheim (1980) and prove a general equivalence theorem in the model robust setup, which contains Theorem 2.2 as a special case. We will also use the same notation as given by Pukelsheim (1980).

Let $NND(k)$ denote the set of all nonnegative definite matrices $A \in \mathbb{R}^{k \times k}$ and similarly $PD(k)$ the set of all positive definite matrices. For a fixed matrix $K \in \mathbb{R}^{k \times s}$ of rank s define A_K as the set of all $A \in NND(k)$ whose range contains the range of K. Let J denote the function from $NND(k)$ into $NND(s)$ which maps every $A \in NND(k)$ into

$(K^TA^-K)^{-1} \in NND(s)$ if $A \in A_K$ and into 0 otherwise. An information functional j on $NND(s)$ is a real valued function which satisfies the conditions

$$
(5.1) \qquad \begin{cases} \text{(i)} \quad j(B) \geq 0 & \forall\ B \in NND(s) \\ \text{(ii)} \quad j(B) > 0 & \forall\ B \in PD(s) \\ \text{(iii)} \quad j(\lambda B) = \lambda j(B) & \forall\ B \in NND(s) \qquad \forall\, \lambda > 0 \\ \text{(iv)} \quad j(B+C) \geq j(B) + j(C) & \forall\ B, C \in NND(s). \end{cases}
$$

The conditions (iii) and (iv) are called positive homogenity and superadditivity. Typical examples of information functionals are the Φ_p-optimality criteria (Kiefer (1974), Pukelsheim (1980))

$$
(5.2) \qquad j_p(C) = \begin{cases} (\text{trace } C^p/s)^{1/p} & \text{if } -\infty < p \leq 1, p \neq 0 \\ (\det\, C)^{1/s} & \text{if } p = 0 \\ \lambda_{\min}(C) & \text{if } p = -\infty \end{cases}
$$

The polar function j° (defined on NND(s)) is given by

$$
(5.3) \qquad j^\circ(D) = \inf\{\frac{\text{trace } (C^TD)}{j(C)} \mid\ C \in PD(s)\}
$$

Let $\mathcal{M}$ denote a compact convex subset of $NND(k)$ for which $\mathcal{M} \cap A_K \neq \emptyset$, then we have the following theorem (see Pukelsheim (1980), Theorem 5).

Theorem 5.1. Let $M \in \mathcal{M} \cap A_K$, $C = J(M) = (K^TM^-K)^{-1}$, then M maximizes the function $j \circ J$ with respect to $\mathcal{M}$ if and only if there exists a generalized inverse G of M and a matrix $D \in NND(s)$ with the properties

$$
(1) \qquad j(C) \cdot j^\circ(D) = \text{ trace } (C^TD) = 1
$$

$$
(2) \qquad \text{trace } (K^TGAG^TKCDC) \leq 1 \qquad \text{for all } A \in \mathcal{M}
$$

If M maximizes $j \circ J$ then for every contracting generalized inverse of M and a matrix D with these properties we have equality in (2) for every matrix A which maximizes $j \circ J$.

Note that the above theorem can be applied to derive an equivalence theorem for the optimal design problem described in section 1 when $n = 1$ ($K = c_1, \mathcal{M} = \{M(\xi)|\xi$ is a probability measure on $\mathcal{X}\}$, $j(c_1) = c_1$, $j^\circ(d_1) = d_1$). However, you can not apply this theorem in the model robust situation described in section 1 (i.e. $n \geq 2$). We will now derive an equivalence theorem for this situation which contains Theorem 5.1 as a special case. The proceeding is similar to that given in Pukelsheim (1980) and we will only state the main theorems and sketch the differences in the proofs.

For $\ell = 1, \ldots, n$ let $K_\ell \in \mathbb{R}^{k_\ell \times s_\ell}$ denote a fixed $k_\ell \times s_\ell$ matrix of rank s_ℓ and consider n information functionals $j_\ell : \quad NND(s_\ell) \to \mathbb{R}_{\geq 0}$. For an adequate description of the model robust set up consider the set

$$NND(k_1, \ldots, k_n) := \{C = (C_1, \ldots, C_n)^T | \;\; C_\ell \in NND(k_\ell) \;\; \ell = 1, \ldots, n\}$$

of all matrix valued vectors C, where the ℓ-th coordinate consists of a nonnegative $k_\ell \times k_\ell$ matrix ($\ell = 1, \ldots, n$). We define the function j from $NND(s_1, \ldots, s_n)$ into $\mathbb{R}_{\geq 0}$ by

$$(5.4) \qquad j(C) = \prod_{\ell=1}^{n} \{j_\ell(C_\ell)\}^{\beta_\ell} \qquad \forall \;\; C \in NND(s_1, \ldots, s_n),$$

where the β_ℓ are positive numbers with sum 1.

Proposition 5.2. Let j_ℓ denote information functionals defined on $NND(s_\ell)$, $\beta_\ell > 0$ ($\ell = 1, \ldots, n$) and $\sum\limits_{\ell=1}^{n} \beta_\ell = 1$. The function j defined by (5.2) is an information functional on $NND(s_1, \ldots, s_n)$.

Proof: Condition (i) and (ii) of (5.1) are obvious, while (iii) follows from $\sum\limits_{\ell=1}^{n} \beta_\ell = 1$ and the fact that the j_ℓ are information functionals. To prove (iv) we use the well known inequality (see Roberts and Varberg (1973), p. 190)

$$(5.5) \qquad \prod_{\ell=1}^{n} x_\ell^{\alpha_\ell} \leq \sum_{\ell=1}^{n} \alpha_\ell x_\ell \qquad \forall \;\; x_\ell \geq 0, \;\; \alpha_\ell > 0, \;\; \sum_{\ell=1}^{n} \alpha_\ell = 1$$

which implies for "matrices" $B=(B_1,\dots,B_n)^T$, $C=(C_1,\dots,C_n)^T\in NND(s_1,\dots,s_n)$

$$j(B+C)=\prod_{\ell=1}^n (j_\ell(B_\ell+C_\ell))^{\beta_\ell}\geq \frac{\prod_{\ell=1}^n (j_\ell(B_\ell)+j_\ell(C_\ell))^{\beta_\ell}}{\prod_{\ell=1}^n j_\ell(B_\ell)^{\beta_\ell}+\prod_{\ell=1}^n j_\ell(C_\ell)^{\beta_\ell}}\left[\prod_{\ell=1}^n j_\ell(B_\ell)^{\beta_\ell}+\prod_{\ell=1}^n j_\ell(C_\ell)^{\beta_\ell}\right]$$

$$=\left[\prod_{\ell=1}^n j_\ell(B_\ell)^{\beta_\ell}+\prod_{\ell=1}^n j_\ell(C_\ell)^{\beta_\ell}\right]\left\{\prod_{\ell=1}^n\left(\frac{j_\ell(B_\ell)}{j_\ell(B_\ell)+j_\ell(C_\ell)}\right)^{\beta_\ell}+\prod_{\ell=1}^n\left(\frac{j_\ell(C_\ell)}{j_\ell(B_\ell)+j_\ell(C_\ell)}\right)^{\beta_\ell}\right\}^{-1}$$

$$\geq\left[\prod_{\ell=1}^n j_\ell(B_\ell)^{\beta_\ell}+\prod_{\ell=1}^n j_\ell(C_\ell)^{\beta_\ell}\right]\left\{\sum_{\ell=1}^n\beta_\ell\frac{j_\ell(B_\ell)}{j_\ell(B_\ell)+j_\ell(C_\ell)}+\sum_{\ell=1}^n\beta_\ell\frac{j_\ell(C_\ell)}{j_\ell(B_\ell)+j_\ell(C_\ell)}\right\}^{-1}$$

$$=\prod_{\ell=1}^n j_\ell(B_\ell)^{\beta_\ell}+\prod_{\ell=1}^n j_\ell(C_\ell)^{\beta_\ell}=j(B)+j(C).$$

This shows property (iv) of (5.1) and completes the proof of the proposition.

For the vectors $B,C\in NND(s_1,\dots,s_n)$ define a scalar product by

$$<B,C>=\sum_{\ell=1}^n\beta_\ell \text{ trace } (B_\ell^T C_\ell).$$

The polar function j° (from $NND(s_1,\dots,s_n)$ into $\mathbb{R}_{\geq 0}$) is defined in a similar way as in (5.3)

$$j^\circ(D)=\inf\{\frac{<C,D>}{j(C)}|\ \ C\in PD(s_1,\dots,s_n)\}\tag{5.6}$$

where $PD(s_1,\dots,s_n)$ denotes the set of all vectors $C=(C_1,\dots,C_n)^T\in NND(s_1,\dots,s_n)$ whose components C_ℓ are positive definite matrices.

Theorem 5.3. Let $j_1,\dots,j_n$ denote an information functionals and $j(C)=\prod_{\ell=1}^n j_\ell(C_\ell)^{\beta_\ell}$ then the polar function j° defined by (5.6) has the representation

$$j^\circ(D)=\prod_{\ell=1}^n j_\ell^\circ(D_\ell)^{\beta_\ell}\qquad\forall\ \ D\in NND(s_1,\dots,s_n)\tag{5.7}$$

where j_ℓ° denotes the polar function of j_ℓ $(\ell = 1, \dots, n)$.

Proof: We will first assume that some of the $j_\ell^\circ(D_\ell)$ are vanishing. Without loss of generalization assume $j_\ell^\circ(D_\ell) > 0$ $(\ell = 1, \dots, n-1)$, $j_n^\circ(D_n) = 0$ and let for every $\varepsilon > 0$ C_n denote a positive definite matrix such that

$$0 = j_n^\circ(D_n) \le \frac{\text{trace } (C_n^T D_n)}{j_n(C_n)} < j_n^\circ(D_n) + \varepsilon = \varepsilon.$$

Defining $\tilde{C}_n = C_n / j_n(C_n)$ we have for every $\varepsilon > 0$ a matrix $\tilde{C}_n \in PD(s_n)$ with

$$0 \le \text{ trace } (\tilde{C}_n^T D_n) < \varepsilon, \quad j_n(\tilde{C}_n) = 1 \tag{5.8}$$

Now let $\tilde{D}_\ell = D_\ell + I_\ell$ $(\ell = 1, \dots, n-1)$ (where $I_\ell \in PD(s_\ell)$ denotes the identity matrix) and $\tilde{C} = (\varepsilon \tilde{D}_1, \dots, \varepsilon \tilde{D}_{n-1}, \tilde{C}_n)^T$, then we have for every $\varepsilon > 0$

$$0 \le j^\circ(D) \le \frac{< \tilde{C}, D >}{j(\tilde{C})} \le \frac{(\sum_{\ell=1}^{n-1} \beta_\ell \text{ trace } (\tilde{D}_\ell^T D_\ell) + \beta_n)\varepsilon}{\prod_{\ell=1}^{n-1} \varepsilon^{\beta_\ell} j_\ell(\tilde{D}_\ell)^{\beta_\ell}} = \varepsilon^{\beta_n} \frac{\sum_{\ell=1}^{n-1} \beta_\ell \text{ trace } (\tilde{D}_\ell^T D_\ell) + \beta_n}{\prod_{\ell=1}^{n-1} j_\ell(\tilde{D}_\ell)^{\beta_\ell}}$$

which shows that $j^\circ(D) = 0$.

Now assume that $j_\ell^\circ(D_\ell) > 0$ for all $\ell = 1, \dots, n$. From the definition of j_ℓ° we have in combination with (5.5)

$$\begin{aligned} \frac{< C, D >}{j(C)} &= \frac{\sum_{\ell=1}^{n} \beta_\ell \text{ trace } (C_\ell^T D_\ell)}{j(C)} \ge \frac{\sum_{\ell=1}^{n} \beta_\ell j_\ell^\circ(D_\ell) j_\ell(C_\ell)}{j(C)} \\ &\ge \frac{\prod_{\ell=1}^{n} (j_\ell^\circ(D_\ell) j_\ell(C_\ell))^{\beta_\ell}}{j(C)} = \prod_{\ell=1}^{n} (j_\ell^\circ(D_\ell))^{\beta_\ell} \end{aligned}$$

for all matrices $C \in PD(s_1, \dots, s_n)$ which implies

$$j^\circ(D) \ge \prod_{\ell=1}^{n} (j_\ell^\circ(D_\ell))^{\beta_\ell}.$$

To prove the converse inequality, we remark that (by the same reasoning as given in (5.8)) for every $\varepsilon > 0$ there exist matrices $\tilde{C}_\ell \in PD(s_\ell)$ $(\ell = 1, \dots, n)$ such that

$$0 < j_\ell^\circ(D_\ell) \le \frac{\text{trace } (\tilde{C}_\ell^T D_\ell)}{j_\ell(\tilde{C}_\ell)} < j_\ell^\circ(D_\ell) + \varepsilon, \quad j_\ell(\tilde{C}_\ell) j_\ell^\circ(D_\ell) = 1$$

Let $\tilde{C} = (\tilde{C}_1, \dots, \tilde{C}_n)^T$ then we obtain from these inequalities for every $\varepsilon > 0$

$$j^\circ(D) \leq \frac{\sum_{\ell=1}^{n} \beta_\ell \text{ trace } (\tilde{C}_\ell^T D_\ell)}{\prod_{\ell=1}^{n} j_\ell(\tilde{C}_\ell)^{\beta_\ell}} \leq \frac{\sum_{\ell=1}^{n} \beta_\ell j_\ell(\tilde{C}_\ell)(j_\ell^\circ(D_\ell) + \varepsilon)}{\prod_{\ell=1}^{n} j_\ell(\tilde{C}_\ell)^{\beta_\ell}}$$
$$= \frac{1 + \varepsilon \sum_{\ell=1}^{n} \beta_\ell j_\ell^\circ(D_\ell)^{-1}}{\prod_{\ell=1}^{n} j_\ell^\circ(D_\ell)^{-\beta_\ell}} = \prod_{\ell=1}^{n} j_\ell(D_\ell)^{\beta_\ell} \left[1 + \varepsilon \sum_{\ell=1}^{n} \frac{\beta_\ell}{j_\ell^\circ(D_\ell)}\right],$$

which proves the theorem.

For a given set $\mathcal{M} \subseteq NND(k_1, \dots, k_n)$ define its polar set

$$\mathcal{N} = \{B \in NND(k_1, \dots, k_n) | < M, B > \ \leq 1 \qquad \forall \ M \in \mathcal{M}\}$$

and consider for fixed matrices $K_\ell \in \mathbb{R}^{k_\ell \times s_\ell}$ $(\ell = 1, \dots, n)$ the mappings $(\ell = 1, \dots, n)$

$$J_\ell : \begin{cases} NND(k_\ell) & \rightarrow NND(s_\ell) \\ M_\ell & \rightarrow J_\ell(M_\ell) = \begin{cases} (K_\ell^T M_\ell^- K_\ell)^{-1} & \text{if } M_\ell \in A_{K_\ell} \\ 0 & \text{else} \end{cases} \end{cases}$$

We now define a function J from $NND(k_1, \dots, k_n)$ into $NND(s_1, \dots, s_n)$ by

$$J(M) = (J_1(M_1), \dots, J_n(M_n))^T \qquad \text{for all } M \in NND(k_1, \dots k_n)$$

then the dual problem of the maximization problem

$$\text{(P)} \qquad \begin{array}{ll} \text{Maximize} & j \circ J(M) \\ \text{subject to} & M \in \mathcal{M} \end{array}$$

is given by $(K^T N K = (K_1^T N_1 K_1, \dots, K_n^T N_n K_n)^T)$

$$\text{(D)} \qquad \begin{array}{ll} \text{Minimize} & 1/j^\circ(K^T N K) \\ \text{subject to} & N \in \mathcal{N} \end{array}$$

(see Pukelsheim (1980)). We thus have the following theorems

Theorem 5.4. For every vector $M = (M_1, \ldots, M_n)^T \in \mathcal{M}$ and for every vector $N = (N_1, \ldots, N_n)^T \in \mathcal{N}$ we have

$$j^\circ J(M) \leq 1/j^\circ(K^T N K)$$

with equality if and only if $M_\ell \in A_{K_\ell} (\ell = 1, \ldots, n)$ and the matrices $C = J(M)$ and $D_\ell = K_\ell^T N_\ell K_\ell$ satisfy the conditions

(1) $\quad 1 = \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (M_\ell^T N_\ell)$

(2) $\quad M_\ell N_\ell = K_\ell C_\ell K_\ell^T N_\ell \quad (\ell = 1, \ldots, n)$

(3) $\quad j_\ell(C_\ell) j_\ell^\circ(D_\ell) = \text{ trace } (C_\ell^T D_\ell) = 1$

Proof: From Pukelsheim (1980) we have $(\ell = 1, \ldots, n)$ trace $(M_\ell^T N_\ell) \geq$ trace $(C_\ell^T D_\ell)$ with equality if and only if $M_\ell N_\ell = K_\ell C_\ell K_\ell^T N_\ell$. Observing the definition of the set $\mathcal{N}$ we obtain

$$\begin{aligned} 1 \geq \; < M, N > &= \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (M_\ell^T N_\ell) \geq \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (C_\ell^T D_\ell) \\ &\geq \sum_{\ell=1}^{n} \beta_\ell j_\ell(C_\ell) j_\ell^\circ(D_\ell) \geq j(C) \cdot j^\circ(D) \end{aligned}$$

where the last inequality follows from (5.5). The conditions (1), (2), (3) correspond to the equal signs in this inequality (note that the products $j_\ell(C_\ell) j_\ell^\circ(D_\ell)$ must have a common value, which corresponds to equality in (5.5) and that this value is 1 by (1) and (2))

Theorem 5.5.

$$\sup_{M \in \mathcal{M}} j \circ J(M) = \min_{N \in \mathcal{N}} \frac{1}{j^\circ(K^T N K)}$$

Proof: Consider the logarithms of the above functions log $j \circ J(M)$ and $-$log $j^\circ(K^T N K)$. By a similar argument as in the proof of Proposition 5.2 it can be shown that the function log $j \circ J(M)$ is concave on $\mathcal{M} \subseteq NND(k_1, \ldots, k_n)$. Define the functions

$$f(A) = \begin{cases} 0 & \text{if } A \in \mathcal{M} \subseteq NND(k_1, \ldots, k_n) \\ \infty & \text{else } (A \in \mathbb{R}^{k_1 \times k_1} \times \ldots \times \mathbb{R}^{k_n \times k_n}) \end{cases}$$

$$g(A) = \begin{cases} \log \ j \circ J(A) & \text{if } A = (A_1, \ldots, A_n)^T, \ A_\ell \in A_{K_\ell} \\ -\infty & \text{else } (A \in \mathrm{R}^{k_1 \times k_1} \times \ldots \times \mathrm{R}^{k_n \times k_n}) \end{cases}$$

and assume at first that $\mathcal{M} \cap PD(k_1, \ldots, k_n) \neq \emptyset$. By a slight modification of Fenchel's Theorem (Rockafellar (1970) (note that we use a "weighted" scalar product $< x, y >= \sum_{i=1}^{n} \beta_i x_i y_i$ instead of the common product) we obtain

$$\sup_A \{g(A) - f(A)\} = \min_B \{f^*(B) - g^*(B)\} \tag{5.9}$$

where f^* and g^* are the conjugate functions of f and g. By definition we have for every vector $B = (B_1, \ldots, B_n)$ with symmetric elements B_ℓ

$$\begin{aligned} g^*(B) &= \inf\{< A, B > -g(A) \ \mid \ A_\ell \in A_{k_\ell}, \ \ell = 1, \ldots, n\} \\ &= \inf\{\sum_{\ell=1}^{n} \beta_\ell [\text{trace } (A_\ell^T B_\ell) - \log \ j_\ell \circ J_\ell(A_\ell)] \ \mid \ A_\ell \in A_{K_\ell}, \ \ell = 1, \ldots, n\} \\ &= \sum_{\ell=1}^{n} \beta_\ell \inf\{\text{trace } (A_\ell^T B_\ell) - \log \ j_\ell \circ J_\ell(A_\ell) \ \mid \ A_\ell \in A_{K_\ell}\} \\ &= 1 + \sum_{\ell=1}^{n} \beta_\ell \log \ j_\ell^\circ(K_\ell^T B_\ell K_\ell) = 1 + \log \ j^\circ(K^T B K) \end{aligned}$$

where we have used the identity ($\ell = 1, \ldots, n$)

$$\inf\{\text{trace } (A_\ell^T B_\ell) - \log \ j_\ell \circ J_\ell(A_\ell) \ \mid \ A_\ell \in A_{K_\ell}\} = 1 + \log j_\ell^\circ(K_\ell^T B_\ell K_\ell)$$

(which was proved by Pukelsheim (1980)) and Theorem 5.3. The functions f^* and g^* have the same value at B and $\frac{1}{2}[B + B^T]$ and thus the minimization problem can be carried out over the set of vectors whose components are symmetric matrices. Take an arbitrary $B \neq 0$ with $g^*(B) > -\infty$ and this property, then we have from the representation

$$f^*(B) = \sup\{< M, B > \mid \ M \in \mathcal{M}\}$$

that $f^*(B)$ is positive and that the function

$$h(\alpha) = f^*(\alpha B) - g^*(\alpha B) = \alpha f^*(B) - 1 - \log \ \alpha - \log \ j^\circ(K^T B K)$$

attains it unique minimum at $\alpha = 1/f^*(B)$. This minimum is given by

$$\log\ f^*(B) - \log\ j^\circ(K^T BK) = \log\ j^\circ(K^T NK)$$

where $N = B/f^*(B) \in \mathcal{N}$. This proves the theorem in the case $\mathcal{M} \cap PD(k_1, \ldots, k_n) \neq 0$. The second part ($\mathcal{M} \cap PD(k_1, \ldots, k_n) = \emptyset$) is treated in the same way as in Pukelsheim (1980) (apply all arguments of this paper to the components of the vectors) and therefore omitted.

Theorem 5.6. Let $M = (M_1, \ldots, M_n)^T \in \mathcal{M}$ where $M_\ell \in A_{K_\ell}$ $(\ell = 1, \ldots, n)$ and let $C = J(M) = ((K_1^T M_1^- K_1)^{-1}, \ldots, (K_n^T M_n^- K_n)^{-1})^T$. Then M maximizes the function $j \circ J$ if and only if there exist generalized inverses $G_1, \ldots, G_n$ of $M_1, \ldots, M_n$ and a matrix $D \in NND(s_1, \ldots, s_n)$ such that

$$\text{(A)} \qquad j(C)j^\circ(D) = \prod_{\ell=1}^{n} (j_\ell(C_\ell) j_\ell^\circ(D_\ell))^{\beta_\ell} = \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (C_\ell^T D_\ell) = 1$$

$$\text{(B)} \qquad \sum_{\ell=1}^{n} \beta_\ell \text{ trace } K_\ell^T G_\ell A_\ell G_\ell^T K_\ell C_\ell D_\ell C_\ell \leq 1 \qquad \text{for all } A = (A_1, \ldots, A_m)^T \in \mathcal{M}$$

$$\text{(C)} \qquad j_\ell(C_\ell) j_\ell^\circ(D_\ell) = \text{ trace } (C_\ell^T D_\ell) = 1 \qquad (\ell = 1, \ldots, n)$$

Whenever M maximizes $j \circ J$ there exist matrices $G_1, \ldots, G_n$ fulfilling (5.10) and $(D_1, \ldots, D_n)$ with these properties and we have for every matrix A also maximizing $j \circ J$ equality in (B).

Proof: For the direct part let $M = (M_1, \ldots, M_n)^T$ denote an optimal matrix. By a similar reasoning as given in Pukelsheim (1980) it can be shown that there exist generalized inverses $G_1, \ldots, G_n$ of $M_1, \ldots, M_n$ and an optimal solution of the dual problem $N = (N_1, \ldots, N_n)^T$ such that

$$(5.10) \qquad \sum_{\ell=1}^{n} \beta_\ell u_\ell^T N_\ell u_\ell = \sum_{\ell=1}^{n} \beta_\ell u_\ell^T G_\ell^T M_\ell N_\ell M_\ell G_\ell u_\ell \quad \text{for all} \quad u_\ell \in \mathbb{R}^{k_\ell} \quad (\ell = 1, \ldots, n)$$

Let $D = (K_1^T N_1 K_1, \ldots, K_n^T N_n K_n) = K^T N K$. Since N is a optimal solution of the dual problem (D) and M is an optimal solution of (P) we have equality in Theorem 5.4 which shows (A) and (C)

$$1 = j(C) j^\circ(D) = \prod_{\ell=1}^{n} (j_\ell(C_\ell) j_\ell^\circ(D_\ell))^{\beta_\ell} = \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (C_\ell^T D_\ell)$$

From condition (2) of the same theorem we obtain (by multiplying with $N^{1/2^+}$)

$$M_\ell N_\ell M_\ell = K_\ell C_\ell D_\ell C_\ell K_\ell^T \qquad (\ell = 1, \ldots, n)$$

and from (5.10) it follows that

$$(5.11) \qquad \begin{cases} \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (K_\ell^T G_\ell A_\ell G_\ell^T K_\ell C_\ell D_\ell C_\ell) = \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (A_\ell G_\ell^T M_\ell N_\ell M_\ell G_\ell) \\ \qquad = \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (A_\ell N_\ell) = < A, N > \leq 1 \end{cases}$$

To prove the converse direction let $G_1, \ldots, G_n$, $D = (D_1, \ldots, D_n)^T$ as stated in the theorem and define $N = (N_1, \ldots, N_n)^T$ by $N_\ell = G_\ell^T K_\ell C_\ell D_\ell C_\ell K_\ell^T G_\ell$. From $M_\ell \in A_{K_\ell}$ we have $K_\ell^T G_\ell K_\ell C_\ell = I_\ell \in PD(s_\ell)$ $(\ell = 1, \ldots, n)$ which shows (observing (C) and Theorem 5.3)

$$j^\circ(K^T N K) = \prod_{\ell=1}^{n} j_\ell^\circ(K_\ell^T N_\ell K_\ell)^{\beta_\ell} = \prod_{\ell=1}^{n} j_\ell^\circ(D_\ell)^{\beta_\ell} = \prod_{\ell=1}^{n} j_\ell(C_\ell)^{-\beta_\ell} = 1/j(C)$$

From condition (B) it follows that $N \in \mathcal{N}$ which shows that M is the optimal solution of the problem (P) and N is the optimal solution of the problem (D).

Whenever M and $A = (A_1, \ldots, A_n)^T$ are optimal we have from (2) and (3) of Theorem 5.4 and (5.11)

$$\sum_{\ell=1}^{n} \beta_\ell \text{ trace } (K_\ell^T G_\ell A_\ell G_\ell^T K_\ell C_\ell D_\ell C_\ell) = \sum_{\ell=1}^{n} \beta_\ell \text{ trace } (A_\ell^T N_\ell) = 1$$

which completes the proof.

We are now going back to the optimal design theory and the situation considered in section 1 to section 4. Recall that the regression functions are given by

$$f_\ell(x) = (f_{\ell 1}(x), \ldots, f_{\ell k_\ell}(x))^T \qquad \ell = 1, \ldots, n,$$

the set $\mathcal{M}$ is now given by

$$\mathcal{M} = \{M(\xi) = (M_1(\xi), \ldots, M_n(\xi))^T \mid \xi \text{ is a probability measure on } \mathcal{X}\}$$

and a design is called j-optimal for the class $\mathcal{F}_n$ with respect to the prior β (for the estimation of the parameter combinations $K_1^T\theta_1, \ldots, K_n^T\theta_n$) if $M_\ell(\xi) \in A_{K_\ell}$ $(\ell = 1, \ldots, n)$ and ξ maximizes $j \circ J(M(\xi))$, where j is defined by (5.4). Then we have the following Theorem.

Theorem 5.7. A design ξ is j-optimal for the class of $\mathcal{F}_n$ with respect to the prior β if and only if there exist generalized inverses $G_1, \ldots, G_n$ of $M_1(\xi), \ldots, M_n(\xi)$ and a vector $D = (D_1, \ldots, D_n)^T$ such that

$$j_\ell(C_\ell)j_\ell^\circ(D_\ell) = \text{ trace } (C_\ell^T D_\ell) = 1$$

$$\sum_{\ell=1}^{n} \beta_\ell \text{ trace } (f_\ell^T(x)G_\ell^T K_\ell C_\ell D_\ell C_\ell K_\ell^T G_\ell f_\ell(x)) \leq 1 \qquad \text{for all } x \in \mathcal{X}$$

where we have equality in this equation for all support points of every design which is j-optimal for the class $\mathcal{F}_n$ with respect to the prior β.

Remark. Note that the support points of the optimal design have to satisfy the equation

$$\sum_{\ell=1}^{n} \beta_\ell f_\ell^T(x) N_\ell f_\ell(x) = 1$$

whenever $N = (N_1, \ldots, N_n)^T$ is an optimal covering cylinder of $f(\mathcal{X})$
(i.e. $< f(x)f^T(x), N > \; = \Sigma\beta_\ell f_\ell^T(x)N_\ell f_\ell(x) \leq 1$ for all $x \in \mathcal{X}$).

For the j_p criteria given in (5.2) Pukelsheim (1980) showed that the polar function of j_p is given by sj_q where $p+q = pq$ and that $D \in NND(s)$ solves the equation $j_p(C)(j_p)^\circ(D) =$ trace $(C^TD) = 1$ if and only if $D = C^{p-1}/\text{trace }(C^p)$ $(p > -\infty)$. Let us consider the case $p = -1$ and let $K_\ell = A_\ell \in \mathrm{R}^{k_\ell \times s_\ell}$ then we have the following corollary which contains

Theorem 2.1 as the special case $s_1 = \ldots = s_n = 1$ and is used in the proof of Theorem 2.3 $(j_\ell(C) = j_{-1}(C),\ \ell = 1, \ldots, n)$.

Theorem 5.8. A design ξ (for which $A_\ell^T\theta_\ell$ is estimable $\ell = 1, \ldots, n$) is A-optimal for the class $\mathcal{F}_n$ with respect to the prior β, if and only if there exist generalized inverses $G_1, \ldots, G_n$ of $M_1(\xi), \ldots, M_n(\xi)$ such that

$$\sum_{\ell=1}^{n} \beta_\ell \frac{\text{trace } (A_\ell^T G_\ell f_\ell(x) f_\ell^T(x) G_\ell^T A_\ell)}{\text{trace } (A_\ell^T M_\ell^- A_\ell)} \leq 1 \quad \text{for all } x \in \mathcal{X}$$

where we have equality if and only if x is a support point of an A-optimal design for the class $\mathcal{F}_n$ with respect to the prior β.

Acknowledgements

I am deeply indebted to Bill Studden for many valuable discussions and suggestions which were essential for a better understanding of the geometry of the Elfing theorem.

References

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

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

Pukelsheim, F. (1980). On linear regression designs which maximize information. *Journal of Statistical Planning and Inference*, **4**, 339–364.

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

Roberts, A. W., Varberg, E. V. (1973). Convex Functions, Academic Press, New York.

Rockafellar, R. T. (1970). Convex Analysis. Princeton, NJ, Princeton University Press.

Searle, R. S. (1982). Matrix Algebra useful for Statistics, Wiley, New York.

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

Studden, W. J. (1971). Elfing's Theorem and optimal designs for quadratic loss. *Ann. Math. Statist.* **42**, 1613–1621.