A GEOMETRIC SOLUTION OF THE BAYESIAN E-OPTIMAL DESIGN PROBLEM

by

H. Dette and W.J. Studden
Universität Göttingen — Purdue University

Technical Report # 92-27

Department of Statistics
Purdue University

June, 1992

A GEOMETRIC SOLUTION OF THE BAYESIAN E-OPTIMAL DESIGN PROBLEM

by

H. Dette*
Institut für Mathematische Stochastik
Universität Göttingen
Lotzestr. 13
3400 Göttingen
GERMANY

and

W.J. Studden**
Department of Statistics
Purdue University
1399 Mathematical Sciences Building
West Lafayette, IN 47907-1399
USA

Abstract

We consider the Bayesian E-optimal design problem in the usual linear model. Using a Bayesian version of Elfving's Theorem for quadratic loss, sufficient conditions are given such that the Bayesian E-optimal design and the classical E-optimal design (without the assumption of a prior distribution for the parameter vector) are supported at the same set of points or are identical. If the minimum eigenvalue of the classical E-optimal information matrix has multiplicity 1 (as in the case of polynomial regression) it is shown that for a sufficiently large number of observations the classical and the Bayesian E-optimal design are supported at the same set of points.

*Research supported in part by the Deutsche Forschungsgemeinschaft
**Research supported in part by NSF Grant DMS-9101730

1. Introduction. Consider the linear regression model

$$y = f'(x)\theta + \varepsilon \tag{1.1}$$

where $f'(x) = (f_1(x), \ldots, f_k(x))$ is the vector of regression functions, x the control variable, $\theta' = (\theta_1, \ldots, \theta_k)$ the vector of unknown parameters and ε is a normally distributed random variable with mean 0 and variance σ^2. We assume that $\mathcal{X}$ is a compact set containing at least k points with sigma field including all one point sets. The regression functions are k linearly independent real valued continuous functions on the design space $\mathcal{X}$. Let $\pi(\theta, \sigma^2)$ denote a prior distribution on (θ, σ^2) such that the conditional prior distribution of θ given σ^2 is a normal distribution with mean μ and covariance matrix $\sigma^2 R^{-1}$ where R is a given positive definite $k \times k$ "precision" matrix. A design ξ is a probability measure on $\mathcal{X}$ (or on its sigma field) and the $k \times k$ matrix

$$M_B(\xi) := M(\xi) + \frac{1}{n}R := \int_{\mathcal{X}} f(x)f'(x)d\xi(x) + \frac{1}{n}R \tag{1.2}$$

is called the Bayesian information matrix of the given design ξ where n denotes the number of observations taken by the experimenter. If ξ concentrates masses $\frac{n_i}{n}$ at s different points x_i, the experimenter takes n uncorrelated observations, n_i at each x_i. In this case (under quadratic loss) the covariance matrix of the posterior conditional distribution of θ, given the observations at these points and σ^2, is proportional to the inverse of the Bayesian information matrix $M_B^{-1}(\xi)$.

A Bayesian optimal design maxi- or minimizes an appropriate optimality criterion depending on $M_B(\xi)$ or its inverse (see Pilz (1991), Chaloner (1984) and El-Krunz and Studden (1991)) and there are numerous criteria which can be chosen to compare competing designs. In this paper we will investigate the A- and E-optimality criterion from a geometric point of view. For a given matrix $A \in \mathbb{R}^{k \times m}$ a design ξ is called Bayesian optimal for $A'\theta$ if ξ minimizes $tr(A'M_B^{-1}(\xi)A)$ and we call a design Bayesian E-optimal if ξ maximizes the minimum eigenvalue $\lambda_{\min}(M_B(\xi))$ of the Bayesian information matrix $M_B(\xi)$. The designs minimizing, and maximizing the corresponding functionals for the "classical" information matrix $M(\xi) = \int_{\mathcal{X}} f(x)f'(x)d\xi(x)$ (that is the information matrix of ξ in the model (1.1) without the assumption of a prior distribution on (θ, σ^2)) are called "classical" optimal for $A'\theta$ and "classical" E-optimal.

*

It is the purpose of this paper to investigate conditions guaranteeing that the Bayesian and classical E-optimal design are supported at the same set of points or are identical. In Section 2 some general results are established investigating the relation between the Bayesian E-optimal and the Bayesian optimal design for $A'\theta$. In Section 3 we use these results and show that for sufficiently large n the Bayesian and the classical E-optimal design are supported at the same set of points if the minimum eigenvalue of the information matrix of the classical E-optimal design has multiplicity 1. If this multiplicity is greater than 1 this property will generally depend on the precision matrix R and we are able to state sufficient conditions for it. In the same section we give some conditions such that the Bayesian and classical E-optimal design coincide. The analysis is based on the consideration of the geometric properties of certain convex subsets in $\mathbb{R}^{k\times m}$ introduced by Elfving (1952) (for $m = 1$) and generalized by Studden (1971) (see also Dette and Studden (1992) and El-Krunz and Studden (1991)). Finally in Section 4, some applications and examples are given in the case of polynomial regression where the classical E-optimal design was recently determined by Pukelsheim and Studden (1992).

2. E-optimal and optimal designs for $A'\theta$.

In this section we present some general results concerning Bayesian E-optimality and Bayesian optimality for $A'\theta$. Some familiarity with the work of Elfving (1952) and Studden (1971) will be helpful. Proofs are only given if they involve new arguments not obtainable from the literature. In all other cases the proofs can either be found in the papers of El-Krunz and Studden (1991) or Dette (1992) or are obvious modifications of the proofs of the corresponding statements for the classical problem given by Dette and Studden (1992). We state all results for the Bayesian optimal design problem, the classical case can be formally obtained by replacing the precision matrix R by a matrix containing only zeros as elements. Our first results are immediate consequences of Theorem 3 of Pukelsheim (1980) and are the basis for all further investigations (see also Pilz (1991) for a proof of Theorem 2.2).

Theorem 2.1. A design ξ is Bayesian E-optimal if and only if there exists a matrix $E \in co(\mathcal{S})$ such that

$$f'(x)Ef(x) \leq \lambda_{\min}(M_B(\xi)) - \frac{1}{n}tr(RE)$$

for all $x \in \mathcal{X}$. Here $\mathcal{S}$ denotes the set of all $k \times k$ matrices of the form zz', with

$\|z\|_2 = 1$, such that z is an eigenvector of $M_B(\xi)$ corresponding to the minimum eigenvalue $\lambda_{\min}(M_B(\xi))$ of the Bayesian information matrix $M_B(\xi)$ and $co(A)$ denotes the convex hull of A.

Theorem 2.2. A design ξ is Bayesian optimal for $A'\theta$ if and only if

$$f'(x)M_B^{-1}(\xi)AA'M_B^{-1}(\xi)f(x) \le tr(A'M_B^{-1}(\xi)M(\xi)M_B^{-1}(\xi)A)$$

for all $x \in \mathcal{X}$.

In what follows $\lambda_{\min}^{(n)}$ and $\lambda_{\min}$ denote the minimum eigenvalue of the Bayesian and classical E-optimal information matrix while $\lambda_{\min}(B)$ is the minimum eigenvalue of a symmetric matrix B. For the classical E-optimal design Dette and Studden (1992) used the so called "Elfving" set (see Elfving (1952) and Studden (1971))

$$\mathcal{S}_m = co(\{f(x)\varepsilon' \mid x \in \mathcal{X},\ \varepsilon \in \mathbb{R}^m,\ \|\varepsilon\|_2 = 1\})$$

for the characterization of the E-optimal design. In the Bayesian context we will need the following set (introduced by El-Krunz and Studden (1991) investigating Bayesian optimal designs for $A'\theta$)

$$\mathcal{K}_m^{(n)} = \left\{ \left(1 + \frac{1}{n} tr(D'RD)\right)^{-\frac{1}{2}} \left(U + \frac{1}{n}RD\right) \mid U \in \partial\mathcal{S}_m,\ D \in \mathcal{C}(U) \right\}$$

where $\mathcal{C}(U) = \{D \mid D \text{ is supporting hyperplane to } \mathcal{S}_m \text{ at } U \in \partial\mathcal{S}_m,\ trD'U = 1\}$.

Theorem 2.3. A design $\xi = \left\{ \begin{smallmatrix} x_\nu \\ \hat{p}_\nu \end{smallmatrix} \right\}_{\nu=1}^{s}$ is Bayesian optimal for $A'\theta$ if and only if there exists a (unique) number $\delta > 0$ and (unique) vectors $\varepsilon_1, \ldots, \varepsilon_s \in \mathbb{R}^m$ with norm $\|\varepsilon_\nu\|_2 = 1$ $(\nu = 1, \ldots, s)$ such that

$$\delta A = \left(1 + \frac{1}{n} D'RD\right)^{-\frac{1}{2}} \left(\hat{U} + \frac{1}{n}RD\right) \in \mathcal{K}_m^{(n)}$$

where $\hat{U} = \sum_{\nu=1}^{s} \hat{p}_\nu f(x_\nu)\varepsilon'_\nu$ and $D \in \mathcal{C}(\hat{U})$.

Remark 2.4. For $m = 1$, $c = A \in \mathbb{R}^{k\times 1}$ the optimality criterion for $A'\theta$ reduces to the well known c-optimality criterion (see e.g. Pukelsheim 1981)). Thus putting formally

$R = 0$ Theorem 2.3 reduces to the well known Elfving theorem (see Elfving (1952)). Thus a design $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{s}$ is classical c-optimal if and only if there exists a positive number $\gamma > 0$, $\varepsilon_\nu = \mp 1$ $(\nu = 1, \ldots, s)$ such that

$$\gamma c = \sum_{\nu=1}^{s} p_\nu f(x_\nu) \varepsilon_\nu \tag{2.1}$$

is a boundary point of the set $\mathcal{S}_1 = co(\{f(x)\varepsilon |\ x \in \mathcal{X},\ \varepsilon = \mp 1\})$.

Theorem 2.5. Let $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ denote a classical optimal design for $A'\theta$ supported at k different points such that $F = [f(x_1), \ldots, f(x_k)]$ is nonsingular with inverse $T = F^{-1}$. Let R denote a precision matrix such that $R^* = TRT' = diag(r_1^*, \ldots, r_k^*)$ is diagonal,

$$\hat{p}_\nu = p_\nu \left(1 + \frac{1}{n} \sum_{i=1}^{k} r_i^* \right) - \frac{1}{n} r_\nu \qquad (\nu = 1, \ldots, k) \tag{2.2}$$

and $n_0 := \min\{n | \hat{p}_\nu > 0,\quad \nu = 1, \ldots, k\}$. Whenever $n \geq n_0$ the design $\xi_B = \left\{ \begin{smallmatrix} x_\nu \\ \tilde{p}_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ (which puts masses $\tilde{p}_\nu$ at the support points x_ν of ξ) is Bayesian optimal for $A'\theta$ if and only if $\tilde{p}_\nu = \hat{p}_\nu$ $(\nu = 1, \ldots, k)$.

Corollary 2.6. For $c \in \mathbb{R}^k$ let $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{s}$ denote the classical c-optimal design supported at k different points with an Elfving representation (2.1) such that $f(x_1), \ldots, f(x_k)$ are linearly independent, $R^* = TRT' = (r_{ij}^*)_{i,j=1}^{k}$,

$$\hat{p}_\nu = p_\nu \left(1 + \frac{1}{n} \sum_{i,j=1}^{k} \varepsilon_i \varepsilon_j r_{ij}^* \right) - \frac{1}{n} \sum_{j=1}^{k} \varepsilon_\nu \varepsilon_j r_{\nu j}^* \qquad (\nu = 1, \ldots, k) \tag{2.3}$$

and $n_0 = \min\{n | \hat{p}_\nu > 0,\quad \nu = 1, \ldots, k\}$. For all $n \geq n_0$ the design $\xi_B = \left\{ \begin{smallmatrix} x_\nu \\ \tilde{p}_\nu \end{smallmatrix} \right\}_{\nu=1}^{s}$ (which puts masses $\tilde{p}_\nu$ at the support points x_ν of the classical c-optimal design) is Bayesian c-optimal if and only if $\tilde{p}_\nu = \hat{p}_\nu$ $(\nu = 1, \ldots, k)$.

Note that Corollary 2.6 does not require a diagonal matrix R^* in contrast with Theorem 2.5 and that there always exists a c-optimal design supported at at most k points (see Fellman (1974), Theorem 3.1.4) which is not necessarily the case for the optimal design for $A'\theta$ if $m > 1$. In what follows define

$$\bar{\mathcal{K}}_m^{(n)} := \{tK |\ t \in [0, 1],\ K \in \mathcal{K}_m^{(n)}\}$$

and let

$$\mathcal{D}^* = \left\{ \left(1 + \frac{1}{n} trace(D'RD)\right)^{-\frac{1}{2}} D \mid D \in \mathcal{C}(U) \text{ for some } U \in \partial\mathcal{S}_m \right\}. \tag{2.4}$$

We will need the following two auxiliary results.

Lemma 2.7. For every matrix $K_0 \in \mathcal{K}_m^{(n)}$ there exists a matrix $D^* \in \mathcal{D}^*$ such that

(i) $trace(D^{*\prime} K_0) = 1$

(ii) $trace(D^{*\prime} K) \leq 1$ for all $K \in \mathcal{K}_m^{(n)}$.

Proof. For

$$K_0 = \left(1 + \frac{1}{n} tr(D_0' R D_0)\right)^{-\frac{1}{2}} \left(U_0 + \frac{1}{n} R D_0\right) \in \mathcal{K}_m^{(n)}$$

and

$$D^* = \left(1 + \frac{1}{n} tr(D_0' R D_0)\right)^{-\frac{1}{2}} D_0 \in \mathcal{D}^*$$

we readily obtain $tr(D^{*\prime} K_0) = 1$. Let $K = (1 + \frac{1}{n} tr(D'RD))^{-\frac{1}{2}}(U + \frac{1}{n} RD)$ denote an arbitrary element of $\mathcal{K}_m^{(n)} (U \in \partial\mathcal{S}_m,\ D \in \mathcal{C}(U))$. From the positive definiteness of the precision matrix R we have

$$2tr(D_0' R D) \leq tr(D_0' R D_0) + tr(D' R D). \tag{2.5}$$

Using Schwarz's inequality it follows that

$$tr^2(D_0' R D) \leq tr(D_0' R D_0) \cdot tr(D' R D)$$

and this implies (observing (2.5))

$$\left(1 + \frac{1}{n} tr(D_0' R D)\right) \leq \sqrt{1 + \frac{1}{n} tr(D_0' R D_0)} \sqrt{1 + \frac{1}{n} tr(D' R D)} \ . \tag{2.6}$$

Because D_0 is a supporting hyperplane to the set $\mathcal{S}_m$ at the boundary point U_0 we have $tr(D_0' U) \leq 1$ for all $U \in \partial\mathcal{S}_m$ and obtain for all $K \in \mathcal{K}_m^{(n)}$ from (2.6)

$$tr(D^{*\prime}K) \le \frac{1+\frac{1}{n}tr(D_0'RD)}{\sqrt{1+\frac{1}{n}tr(D_0'RD_0)}\sqrt{1+\frac{1}{n}tr(D'RD)}} \le 1.$$

■

Lemma 2.8. The set $\bar{\mathcal{K}}_m^{(n)}$ is convex with boundary $\mathcal{K}_m^{(n)}$.

Proof. Let $K_1,\ K_2 \in \bar{\mathcal{K}}_m^{(n)}$, $\alpha \in (0,1)$ and $C = \alpha K_1 + (1-\alpha)K_2$. It was shown by El-Krunz and Studden (1991) that the line $\{\lambda C | \lambda > 0\}$ intersects the set $\mathcal{K}_m = \{U + \frac{1}{n}RD | U \in \partial\mathcal{S}_m,\ D \in \mathcal{C}(U)\}$ at a (unique) point γC, that is

$$\gamma C = U + \frac{1}{n}RD \in \mathcal{K}_m$$

for some $\gamma > 0$, $U \in \partial\mathcal{S}_m$ and $D \in \mathcal{C}(U)$. This implies for $\delta = \gamma(1+\frac{1}{n}tr(D'RD))^{-\frac{1}{2}}$ that

$$K_C := \delta C = \left(1 + \frac{1}{n}tr(D'RD)\right)^{-\frac{1}{2}} \gamma C \in \mathcal{K}_m^{(n)} \tag{2.7}$$

and by Lemma 2.7 there exists a $D^* \in \mathcal{D}^*$ such that $tr(D^{*\prime}K) \le 1 = tr(D^{*\prime}K_C)$ for all $K \in \mathcal{K}_m^{(n)}$. Thus we obtain by a multiplication of (2.7) with $D^{*\prime}$

$$1 = tr(D^{*\prime}K_C) = \delta tr(D^{*\prime}(\alpha K_1 + (1-\alpha)K_2)) \le \delta$$

which shows that $C = \frac{1}{\delta}K_C \in \bar{\mathcal{K}}_m^{(n)}$ and proves the convexity of $\bar{\mathcal{K}}_m^{(n)}$. From Lemma 2.7 we have that $\mathcal{K}_m^{(n)} \subseteq \partial\bar{\mathcal{K}}_m^{(n)}$ and for the converse inclusion we consider a $K_0 \in \partial\bar{\mathcal{K}}_m^{(n)}$. Because $\bar{\mathcal{K}}_m^{(n)}$ is closed it follows that there exists a $t \in (0,1]$ and a $K_1 \in \mathcal{K}_m^{(n)}$ such that $K_0 = tK_1$. Let D_0 denote the supporting hyperplane to the (convex) set $\bar{\mathcal{K}}_m^{(n)}$ at the boundary point K_0, then we obtain

$$tr(D_0'K) \le 1 = tr(D_0'K_0)$$

for all $K \in \bar{\mathcal{K}}_m^{(n)}$. Inserting in this inequality $K_1 \in \mathcal{K}_m^{(n)} \subseteq \bar{\mathcal{K}}_m^{(n)}$ we have

$$1 \ge tr(D_0'K_1) = \frac{1}{t}tr(D_0'K_0) = \frac{1}{t}$$

which shows that $t = 1$ or equivalently $K_0 = K_1 \in \mathcal{K}_m^{(n)}$. ■

Theorem 2.9. Let ξ denote a Bayesian E-optimal design and E a matrix satisfying the conditions of Theorem 2.1 with a convex representation

$$E = \sum_{i=1}^{k_0} \alpha_i z_i z_i' \tag{2.8}$$

by normalized eigenvectors z_i of $M_B(\xi)$ corresponding to the minimum eigenvalue $\lambda_{\min}^{(n)}$. Then the design ξ is also Bayesian optimal for $A'\theta$, where the matrix A is given by $A = (\sqrt{\alpha_1} z_1, \ldots, \sqrt{\alpha_{k_0}} z_{k_0}) \in \mathbb{R}^{k \times k_0}$.

By the convexity of the Elfving set $\bar{\mathcal{K}}_m^{(n)}$ (see Lemma 2.8) it now makes sense to investigate the "inball" of $\bar{\mathcal{K}}_m^{(n)}$ which is the largest symmetric ball centered at the origin and included in $\bar{\mathcal{K}}_m^{(n)}$. The radius of this ball

$$r_m = \inf\{\sqrt{tr(K'K)} \mid K \in \mathcal{K}_m^{(n)}\}$$

is called the "inball radius" of the convex set $\bar{\mathcal{K}}_m^{(n)}$ (note that $\mathcal{K}_m^{(n)} = \partial\bar{\mathcal{K}}_m^{(n)}$ by Lemma 2.8). A vector $K \in \mathcal{K}_m^{(n)}$ with $tr(K'K) = r_m^2$ is called an "inball vector" of $\bar{\mathcal{K}}_m^{(n)}$. The following Lemma gives an alternative representation of the inball radius r_m using the set $\mathcal{D}^*$ defined by (2.4).

Lemma 2.10.

$$r_m^2 = \inf\left\{\frac{1}{tr(D^{*\prime}D^*)} \mid D^* \in \mathcal{D}^*\right\}$$

Proof. By Lemma 2.7 it is obvious that $\mathcal{D}^*$ is a subset of the set of all supporting hyperplanes to $\bar{\mathcal{K}}_m^{(n)}$ and this implies

$$\begin{aligned} r_m^2 &= \inf\left\{\frac{1}{tr(D'D)} \mid D \text{ is a supporting hyperplane to } \bar{\mathcal{K}}_m^{(n)}\right\} \\ &\leq \inf\left\{\frac{1}{tr(D^{*\prime}D^*)} \mid D^* \in \mathcal{D}^*\right\}. \end{aligned}$$

Let $K \in \bar{\mathcal{K}}_m^{(n)}$ denote an inball vector of $\bar{\mathcal{K}}_m^{(n)}$ (note that $\bar{\mathcal{K}}_m^{(n)}$ is compact and the infimum is in fact a minimum), then the supporting hyperplane D^* to $\bar{\mathcal{K}}_m^{(n)}$ at K is unique and by Lemma 2.7 an element of $\mathcal{D}^*$, which proves the assertion. ■

Theorem 2.11. Let $\xi = \left\{\begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix}\right\}_{\nu=1}^{s}$ denote a Bayesian E-optimal design and E the matrix of the equivalence Theorem 2.1 with a representation (2.8). For every $m \geq k_0$ the matrix

*

$\sqrt{\lambda_{\min}^{(n)}}A = \sqrt{\lambda_{\min}^{(n)}}\ (\sqrt{\alpha_1}z_1,\ldots,\sqrt{\alpha_{k_0}}z_{k_0},\ 0,\ldots,0) \in \mathbb{R}^{k\times m}$ is an inball vector of the set $\bar{\mathcal{K}}_m^{(n)}$ and the squared inball radius of $\bar{\mathcal{K}}_m^{(n)}$ is given by $\lambda_{\min}^{(n)}$.

Proof. Consider the Bayesian optimal design problem for $A'\theta$. Theorem 2.9 shows that the Bayesian E-optimal design is also Bayesian optimal for $A'\theta$ and thus Theorem 2.3 implies that there exist $\delta > 0,\ \varepsilon_1,\ldots,\varepsilon_s \in \mathbb{R}^m\ \ (\|\varepsilon_\nu\|_2 = 1)$ satisfying

$$\delta A = \left(1 + \frac{1}{n}tr(D'RD)\right)^{-\frac{1}{2}} \left(\sum_{\nu=1}^{s} \hat{p}_\nu f(x_\nu)\varepsilon_\nu' + \frac{1}{n}RD\right) \in \mathcal{K}_m^{(n)}.$$

By the proof of this theorem (see El-Krunz and Studden (1991) or Dette (1992)) we obtain $D = \delta^* M_B^{-1}(\xi)A = \frac{\delta^*}{\lambda_{\min}^{(n)}}A$ where

$$(\delta^*)^{-2} = tr\left(A'M_B^{-1}(\xi)M(\xi)M_B^{-1}(\xi)A\right) = tr\left(A'M_B^{-1}(\xi)A - \frac{1}{n}A'M_B^{-1}(\xi)RM_B^{-1}(\xi)A\right)$$
$$= \left(\frac{1}{\lambda_{\min}^{(n)}}\right)^2 tr\left(\lambda_{\min}^{(n)}A'A - \frac{1}{n}A'RA\right) = \left(\frac{1}{\lambda_{\min}^{(n)}}\right)^2 \left[\lambda_{\min}^{(n)} - \frac{1}{n}tr(A'RA)\right]$$

and $\delta^* A = \sum_{\nu=1}^{s} p_\nu f(x_\nu)\varepsilon_\nu' + \frac{1}{n}RD \in \mathcal{K}_m$. This implies

$$\delta = \left(1 + \frac{1}{n}tr(D'RD)\right)^{-\frac{1}{2}} \delta^*$$
$$= \lambda_{\min}^{(n)} \left[1 + \frac{1}{n}\frac{tr(A'RA)}{\lambda_{\min}^{(n)} - \frac{1}{n}tr(A'RA)}\right]^{-\frac{1}{2}} \left[\lambda_{\min}^{(n)} - \frac{1}{n}tr(A'RA)\right]^{-\frac{1}{2}} = \sqrt{\lambda_{\min}^{(n)}}$$

and the definition of the inball radius yields

$$r_m^2 \le tr(\delta A'\delta A) = \lambda_{\min}^{(n)}.$$

The converse inequality follows by an application of Lemma 2.10 and similar arguments as given in Dette and Studden (1992) is therefore omitted. ■

Theorem 2.12. Let ξ denote a Bayesian E-optimal design and E the matrix of the equivalence Theorem 2.1 with a representation (2.8). For all $m \ge k_0$ the Bayesian E-optimal design ξ is also Bayesian optimal for $A_n'\theta$ where A_n is any inball vector of the Bayesian Elfving set $\bar{\mathcal{K}}_m^{(n)}$. Moreover, if $D \in \mathcal{D}^*$ is a supporting hyperplane to $\bar{\mathcal{K}}_m^{(n)}$ at the inball vector A_n we have

$$\|D'f(x_\nu)\|_2^2 = 1 + \frac{1}{n}tr(D'RD)$$

for all support points of the Bayesian E-optimal design.

3. Bayesian and classical E-optimal designs. The results of the previous section indicate that the inball vectors of the classical Elfving set $\mathcal{S}_m$ and the Bayesian Elfving set $\bar{\mathcal{K}}_m^{(n)}$ play a particular role for the E-optimal design problem. The following theorem shows that the results of Theorem 2.5 can be transferred to the E-optimality criterion provided that there exist inball vectors of these sets with the same direction. In what follows k_0 will always denote the larger of the two numbers in the representation (2.8) corresponding to the matrix E in the equivalence Theorem 2.1 for a Bayesian and a classical E-optimal design (in the last case we formally put $R = 0$ in Theorem 2.1, see also Pukelsheim (1980)).

Theorem 3.1. Let $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ denote a classical E-optimal design supported at k points, $F = [f(x_1), \ldots, f(x_k)]$ be non singular and $T = F^{-1}$. Let R denote a precision matrix such that $R^* = TRT' = diag(r_1^*, \ldots, r_k^*)$ is diagonal,

$$\hat{p}_\nu = p_\nu \left(1 + \frac{1}{n} \sum_{i=1}^{k} r_i^* \right) - \frac{1}{n} r_\nu^* \qquad (\nu = 1, \ldots, n) \tag{3.1}$$

and $n_0 := \min\{n | \hat{p}_\nu > 0 \quad \nu = 1, \ldots, k\}$. If $n \geq n_0$, $m \geq k_0$ and there exist inball vectors A and $A^{(n)}$ of $\mathcal{S}_m$ and $\bar{\mathcal{K}}_m^{(n)}$ with the same direction (i.e. $A^{(n)} = \rho_n A$ for some $\rho_n > 1$), then the design $\xi_B = \left\{ \begin{smallmatrix} x_\nu \\ \hat{p}_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ (which puts masses $\hat{p}_\nu$ at the support points x_ν of the classical E-optimal design) is a Bayesian E-optimal design.

Proof. Assume that $n \geq n_0$ and that A and $A^{(n)}$ are inball vectors of the Elfving sets $\mathcal{S}_m$ and $\bar{\mathcal{K}}_m^{(n)}$ with the same direction, that is $A^{(n)} = \rho_n A$. Denoting by $\lambda_{\min}$ and $\lambda_{\min}^{(n)}$ the minimum eigenvalues of the classical and Bayesian E-optimal information matrix, it follows from Theorem 2.11 and Theorem 3.3 of Dette and Studden (1992) that

$$\rho_n^2 \lambda_{\min} = \lambda_{\min}^{(n)}.$$

Because A and $A^{(n)}$ are inball vectors of their corresponding Elfving sets the supporting hyperplanes at these points are unique and given by $\frac{A}{\lambda_{\min}}$ and $\frac{A^{(n)}}{\lambda_{\min}^{(n)}}$, respectively. Observing Lemma 2.7 we thus obtain

$$\frac{A^{(n)}}{\lambda_{\min}^{(n)}} = \left(1 + \frac{1}{n} \frac{tr(A'RA)}{\lambda_{\min}^2} \right)^{-\frac{1}{2}} \frac{A}{\lambda_{\min}} \tag{3.2}$$

*

which implies $\rho_n = \left(1 + \frac{1}{n}\frac{tr(A'RA)}{\lambda^2_{\min}}\right)^{\frac{1}{2}}$, or equivalently

(3.3) $$\lambda^{(n)}_{\min} = \left(1 + \frac{1}{n}\frac{tr(A'RA)}{\lambda^2_{\min}}\right)\lambda_{\min} \ ,$$

(3.4) $$A^{(n)} = \left(1 + \frac{1}{n}\frac{tr(A'RA)}{\lambda^2_{\min}}\right)^{\frac{1}{2}} A.$$

From Theorem 3.4 of Dette and Studden (1992) we see that the classical E-optimal design ξ is also classical optimal for $A'\theta$, and the Elfving Theorem of Studden (1971) shows that there exist (unique) vectors $\varepsilon_1, \ldots, \varepsilon_k \in \mathbf{R}^m$ such that

(3.5) $$A = \sum_{\nu=1}^{k} p_\nu f(x_\nu)\varepsilon'_\nu \in \partial\mathcal{S}_m.$$

Using Theorem 2.5 it follows that the design $\xi_B = \left\{ \begin{smallmatrix} x_\nu \\ \hat{p}_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ is Bayesian optimal for $A'\theta$ and by (3.4) also Bayesian optimal for $A^{(n)\prime}\theta$. Thus observing Theorem 2.12 the design ξ_B can be considered as a candidate for a Bayesian E-optimal design.

Applying Theorem 2.3 we obtain a (unique) representation

(3.6) $$A^{(n)} = \left(1 + \frac{1}{n}tr(D'RD)\right)^{-\frac{1}{2}}\left(U + \frac{1}{n}RD\right) \in \bar{\mathcal{K}}^{(n)}_m$$

where $U \in \partial\mathcal{S}_k,\ D \in \mathcal{C}(U)$ satisfy

$$U = \sum_{\nu=1}^{k} \hat{p}_\nu f(x_\nu)\varepsilon'_\nu \quad \text{and} \quad D = \frac{A}{\lambda_{\min}}.$$

This follows from the proof of Theorem 2.5 (see Dette (1992)) and the fact that all supporting hyperplanes at the face

$$\mathcal{S}^{x_1,\ldots,x_k} := \left\{ \sum_{\nu=1}^{k} \alpha_\nu f(x_\nu)\varepsilon'_\nu \mid \alpha_\nu > 0,\ \sum_{\nu=1}^{k}\alpha_\nu = 1 \right\} \subseteq \partial\mathcal{S}_m$$

spanned by the points $\{f(x_\nu)\varepsilon'_\nu\}_{\nu=1}^k$ are unique and given by $A/\lambda_{\min}$ (note that by (3.5) the inball vector A of $\mathcal{S}_m$ satisfies $A \in \mathcal{S}_m^{x_1,\ldots,x_k}$). On the other hand El-Krunz and Studden (1991) showed that the supporting hyperplane to $\mathcal{S}_m$ at U in (3.6) is given by

$$D = \left(1 + \frac{1}{n}\frac{tr(A'RA)}{\lambda^2_{\min}}\right)^{\frac{1}{2}} M_B^{-1}(\xi_B)A^{(n)}$$

*

which implies (using (3.4))

$$\frac{A}{\lambda_{\min}} = M_B^{-1}(\xi_B)\left(1+\frac{1}{n}\frac{tr(A'RA)}{\lambda_{\min}^2}\right)A. \tag{3.7}$$

Observing (3.3) it thus follows that the columns of the matrix A (or $A^{(n)}$) are eigenvectors of the Bayesian information matrix $M_B(\xi_B)$ corresponding to $\lambda_{\min}^{(n)}$ and it only remains to show that $\lambda_{\min}^{(n)}$ is in fact the minimum eigenvalue of $M_B(\xi_B)$. But from (3.1) we have for this matrix (note that R^* is diagonal by assumption)

$$\begin{aligned} M_B(\xi_B) &= \left(1+\frac{1}{n}\sum_{i=1}^{n} r_i^*\right)\sum_{\nu=1}^{k} p_\nu f(x_\nu)f'(x_\nu) - \frac{1}{n}\sum_{\nu=1}^{k} r_\nu^* f(x_\nu)f'(x_\nu) + \frac{1}{n}R \\ &= \left(1+\frac{1}{n}\sum_{i=1}^{n} r_i^*\right)M(\xi) - \frac{1}{n}FR^*F' + \frac{1}{n}R = \left(1+\frac{1}{n}\sum_{i=1}^{n} r_i^*\right)M(\xi) \end{aligned}$$

and an application of (3.3) completes the proof of the theorem. ■

The calculation of the inball vectors of the sets $\mathcal{S}_k$ and $\bar{\mathcal{K}}_k^{(n)}$ is in general very difficult and some examples can be found in a recent paper of Dette and Studden (1992). However, the preceding theorem suggests the following procedure for the determination of the Bayesian E-optimal design when the classical E-optimal design supported at k points is known, $R^* = TRT'$ is diagonal and the assumptions of Theorem 3.1 are hard to verify.

(I) In a first step the normalized eigenvectors z_j corresponding to the minimum eigenvalue $\lambda_{\min}$ of the classical E-optimal information matrix $M(\xi)$ are calculated and a matrix

$$E = \sum_{i=1}^{k_0} \alpha_i z_i z_i' \qquad \left(\alpha_i > 0 \quad \sum_{i=1}^{k_0} \alpha_i = 1\right) \tag{3.8}$$

satisfying $f'(x)Ef(x) \le \lambda_{\min}$ $(\forall\ x \in \mathcal{X})$ is determined (note that this is the equivalence theorem for the classical E-optimality criterion, that is $R = 0$ in Theorem 2.1 (see Pukelsheim (1980) for more details)). By Theorem 3.3 of Dette and Studden (1992) it follows that the matrix $A = \sqrt{\lambda_{\min}}\,(\sqrt{\alpha_1}z_1,\ldots,\sqrt{\alpha_{k_0}}z_{k_0})$ defines an inball vector of the set $\mathcal{S}_{k_0}$.

(II) In a second step we calculate $A^{(n)} = \left(1+\frac{1}{n}\frac{tr(A'RA)}{\lambda_{\min}^2}\right)^{\frac{1}{2}}A$. If $A^{(n)}$ is an inball vector of the Bayesian Elfving set $\bar{\mathcal{K}}_{k_0}^{(n)}$ and $n \ge n_0$, then Theorem 3.1 shows that the design

$\xi_B = \left\{ \begin{smallmatrix} x_\nu \\ \hat{p}_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ with masses $\hat{p}_\nu$ (defined by (3.1)) at the points x_ν is a Bayesian E-optimal design.

(III) If it can not be proved (or disproved) that $A^{(n)}$ is an inball vector of $\bar{\mathcal{K}}_{k_0}^{(n)}$ there still is the possibility that ξ_B is a Bayesian E-optimal design. In this case we calculate by an application of (3.3)

$$\lambda = \left(1 + \frac{1}{n} \frac{tr(A'RA)}{\lambda_{\min}^2}\right) \lambda_{\min}$$

as a candidate for the minimum eigenvalue of the Bayesian E-optimal information matrix. From the proof of Theorem 3.1 we know that λ is the minimum eigenvalue of $M_B(\xi_B)$ and we finally apply the equivalence Theorem 3.1 to examine if ξ_B is a Bayesian E-optimal design. ■

The situation of Theorem 3.1 becomes more transparent if the minimum eigenvalue of the classical E-optimal information matrix has multiplicity 1. In this case there always exist inball radii of $\mathcal{S}_1$ and $\bar{\mathcal{K}}_1^{(n)}$ having the same direction.

Theorem 3.2. Let $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ denote a classical E-optimal design such that the minimum eigenvalue of the E-optimal information matrix $M(\xi)$ has multiplicity 1. There exists an $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$ we have the following. For every inball vector c_n of $\bar{\mathcal{K}}_1^{(n)}$ there exists an inball vector c of $\mathcal{S}_1$ with the same direction as c_n.

Proof. Without loss of generality we assume that the minimum eigenvalue of the classical E-optimal design is given by 1, and obtain for the inball radius s_m of $\mathcal{S}_m$ by Theorem 3.3 of Dette and Studden (1992) that $s_m = 1$ whenever $m \geq 1$. In the following let

$$\mathcal{A} = \{C \in \partial\mathcal{S}_1 \mid \|c\|_2 = 1\}$$

denote the set of all inball vectors of $\mathcal{S}_1$ and $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^{k}$ denote a classical E-optimal design such that the minimum eigenvalue of its information matrix has multiplicity 1. By Theorem 3.4 of Dette and Studden (1992) ξ is classical c-optimal for $c'\theta$ for all $c \in \mathcal{A}$ and by the famous Elfving Theorem (see Elfving (1952) or Pukelsheim (1981)), we have for every $c \in \mathcal{A}$ a representation

(3.9) $$c = \sum_{\nu=1}^{k} p_\nu f(x_\nu)\varepsilon_\nu(c)$$

where $\varepsilon_\nu(c) = c' f(x_\nu) = \mp 1$ $(\nu = 1, \dots, k)$ and the p_ν are positive because the E-optimal design has a nonsingular information matrix. In the following let

$$\mathcal{A}^{x_1,\dots,x_k}(c) := \left\{ \sum_{\nu=1}^{k} \alpha_\nu f(x_\nu) \varepsilon_\nu(c) \middle|\ \alpha_\nu > 0 \quad \sum_{\nu=1}^{k} \alpha_\nu = 1 \right\}$$

denote the corresponding boundary face of the inball vector $c \in \mathcal{A}$ (with a representation (3.9)) and define

$$\mathcal{A}^* := \bigcup_{c \in \mathcal{A}} \mathcal{A}^{x_1,\dots,x_k}(c)$$

as the union of all these faces. Because all weights p_ν in the representation (3.9) are positive, every inball vector $c \in \mathcal{A}$ is a point of the relative interior of its corresponding face $\mathcal{A}^{x_1,\dots,x_k}(c)$ and we obtain

$$t = \inf\{\|u\|_2 |\ u \in \partial \mathcal{S}_1 \backslash \mathcal{A}^*\} > 1. \tag{3.10}$$

In a first step we will now show that there exists an $n_0 \in \mathbb{N}$ such that whenever $n \geq n_0$ we can find for every inball vector c_n of $\bar{\mathcal{K}}_1^{(n)}$ an inball vector $c \in \mathcal{A}$ and a boundary point $u_n \in \mathcal{A}^{x_1,\dots,x_k}(\mathcal{C})$ satisfying

$$c_n = \frac{u_n + \frac{1}{n} Rc}{(1 + \frac{1}{n} c' Rc)^{\frac{1}{2}}} \tag{3.11}$$

(note that the unique supporting hyperplane at $u_n \in \mathcal{A}^{x_1,\dots,x_k}(c)$ is given by c). To do this we assume the contrary and obtain from the definition of $\bar{\mathcal{K}}_1^{(n)}$ and Lemma 2.7 a sequence of inball radii

$$c_{n_k} = \frac{u_{n_k} + \frac{1}{n_k} R d_{n_k}}{(1 + \frac{1}{n_k} d_{n_k} R d_{n_k})^{\frac{1}{2}}}$$

where $u_{n_k} \in \delta \mathcal{S}_1 \backslash \mathcal{A}^*$ and $d_{n_k} \in \mathcal{C}(u_{n_k})$. This sequence is bounded and contains a convergent subsequence also denoted by n_k. The inequality (3.10) now yields

$$\lim_{n_k \to \infty} \|c_{n_k}\|_2 \geq t > 1$$

which contradicts the fact that $\bar{\mathcal{K}}_1^{(n)}$ approximates the set $\mathcal{S}_1$ arbitrary close as $n \to \infty$ (note that we have assumed that the inball radius of $\mathcal{S}_1$ is 1). This shows that (3.11) holds for sufficiently large n, say $n \geq n_0$.

Because c_n is an inball vector of $\bar{\mathcal{K}}_1^{(n)}$, the unique supporting hyperplane d_n to $\bar{\mathcal{K}}_1^{(n)}$ at c_n has the same direction as c_n which yields $d_n = \frac{c_n}{c_n' c_n}$. But by Lemma 2.7 and its proof the vector $(1 + \frac{1}{n} c' R c)^{-\frac{1}{2}} c$ is a supporting hyperplane to $\bar{\mathcal{K}}_1^{(n)}$ at c_n (note that c is the unique supporting hyperplane to $\mathcal{S}_1$ at u_n) and we thus obtain

$$c_n = c_n' c_n \left(1 + \frac{1}{n} c' R c\right)^{-\frac{1}{2}} c.$$

This shows that whenever $n \geq n_0$ we can find for every inball vector c_n of $\bar{\mathcal{K}}_1^{(n)}$ an inball vector of $\mathcal{S}_1$ with the same direction as c_n, and completes the proof of Theorem 3.2. ■

Theorem 3.3. Let $\xi = \left\{ \begin{smallmatrix} x_\nu \\ p_\nu \end{smallmatrix} \right\}_{\nu=1}^k$ denote a classical E-optimal design supported at k points such that the E-optimal information matrix $M(\xi)$ is unique and the minimum eigenvalue $\lambda_{\min}$ has multiplicity 1 with corresponding normalized eigenvector c. Let $F = [f(x_1), \ldots, f(x_k)]$ be non singular, $T = F^{-1}$, $R^* = TRT' = (r_{ij}^*)_{i,j=1}^k$ and

$$\hat{p}_\nu = p_\nu \left(1 + \frac{1}{n} \sum_{i,j=1}^k \varepsilon_i \varepsilon_j r_{ij}^*\right) - \frac{1}{n} \sum_{j=1}^k \varepsilon_\nu \varepsilon_j r_{ij}^* \tag{3.12}$$

where $\varepsilon_1, \ldots, \varepsilon_k$ are the quantities of the Elfving representation (2.1) of the inball vector $\sqrt{\lambda_{\min}} c$. There exists an $n_0 \geq \min\{n | \hat{p}_\nu > 0 \;\; \nu = 1, \ldots, k\}$ such that whenever $n \geq n_0$ the design $\xi_B^{(n)} = \left\{ \begin{smallmatrix} x_\nu \\ \hat{p}_\nu \end{smallmatrix} \right\}_{\nu=1}^k$ with masses $\hat{p}_\nu$ at the points x_ν is Bayesian E-optimal. Moreover, the multiplicity of the minimum eigenvalue of $M_B(\xi_B^{(n)})$ is 1.

Proof. In what follows let $\xi^{(n)}$ denote a Bayesian E-optimal. In a first step we show that there exists an n_1 such that for all $n \geq n_1$ the minimum eigenvalue $\lambda_{\min}^{(n)}$ of $M_B(\xi^{(n)})$ has multiplicity 1. To this end we assume the contrary and obtain the existence of a subsequence $(n_k)_{k \in \mathbb{N}}$ such that the minimum eigenvalues $\lambda_{\min}^{(n_k)}$ of the Bayesian information matrices $M_B(\xi^{(n_k)})$ have multiplicity greater than or equal to two. The design space $\mathcal{X}$ is compact and thus the set of all probability measures on $\mathcal{X}$ is relatively compact (see Billingsley (1968)). This implies the existence of a weakly convergent subsequence $\xi^{(n_k)}$ (also denoted by n_k) with limit ξ^*. From the continuity of the regression functions we obtain

$$M(\xi^*) = \lim_{n_k \to \infty} M_B(\xi^{(n_k)})$$

and Theorem 2.11 and Theorem 3.4 of Dette and Studden (1992) imply

$$\lambda_{\min} = \lim_{n_k \to \infty} \lambda_{\min}^{(n_k)}. \tag{3.13}$$

Using (3.13) it is now straight forward to show that ξ^* is also a classical E-optimal design such that the minimum eigenvalue of the information matrix $M(\xi^*)$ has multiplicity greater than or equal to 2. This is a contradiction to the assumptions that the classical E-optimal information matrix is unique and its minimum eigenvalue $\lambda_{\min}$ has multiplicity 1.

From now on assume that $n \geq n_1$ such that all Bayesian E-optimal information matrices $M_B(\xi^{(n)})$ have a minimum eigenvalue with multiplicity 1. By Theorem 3.2 there exists an $n_2 \geq n_1$ such that for all $n \geq n_2$ we have the following: For every inball vector c_n of $\bar{\mathcal{K}}_1^{(n)}$ there exists an inball vector c of $\mathcal{S}_1$ with the same direction as c_n. Using similar arguments as in the proof of Theorem 3.1 it follows that for sufficiently large $n \geq n_0 \geq \max\{n|\hat{p}_\nu > 0$ $\nu = 1, \ldots, k\}$, c_n is an eigenvector of the Bayesian information matrix $M_B(\xi_B^{(n)})$ corresponding to $\lambda_{\min}^{(n)}$. The proof will now be completed by showing that for sufficiently large n, $\lambda_{\min}^{(n)}$ is in fact the minimum eigenvalue of $M_B(\xi_B^{(n)})$ (note that the arguments at the end of the proof of Theorem 3.1 do not apply because R^* is not necessarily diagonal).

For this final step we again assume the contrary which means the existence of a subsequence of Bayesian information matrices $M_B(\xi_B^{(n_k)})$ with minimum eigenvalues $\lambda^{(n_k)} < \lambda_{\min}^{(n_k)}$. From the weak convergence of $\xi_B^{(n_k)}$ to the classical E-optimal design ξ^* we conclude

$$\lim_{n_k \to \infty} M_B(\xi_B^{(n_k)}) = M(\xi^*). \tag{3.14}$$

Let b_{n_k} denote a sequence of normalized eigenvectors corresponding to $\lambda^{(n_k)}$ such that $b'_{n_k} c_{n_k} = 0$ (note that c_{n_k} is an eigenvector of $M_B(\xi_B^{(n)})$). Because $\lambda_{\min}^{(n_k)} = \|c_{n_k}\|_2^2$ is the squared inball radius of $\bar{\mathcal{K}}_1^{(n_k)}$ we obtain from (3.14), Theorem 2.11 and Theorem 3.4 of Dette and Studden (1992) that

$$\lambda_{\min} = \lim_{n_k \to \infty} \lambda^{(n_k)} \leq \lim_{n_k \to \infty} \lambda_{\min}^{(n_k)} = \lambda_{\min}. \tag{3.15}$$

Obviously, there exist convergent subsequences of $\{b_{n_k}\}$ and $\{c_{n_k}\}$ with respective limits b and c such that $b'c = 0$. Thus it follows from (3.14) and (3.15) that b and c are linearly independent eigenvectors of the classical E-optimal information matrix $M(\xi)$ corresponding to the minimum eigenvalue $\lambda_{\min}$. This is a contradiction to the assumption that this eigenvalue has multiplicity 1 and the assertion of Theorem 3.3 is proved. ■

Remark 3.4.

a) It is an immediate consequence of the proof of Theorem 3.2 and Theorem 3.3 that the number n_0 defined by Theorem 3.3 can be much larger than $\min\{n|\hat{p}_\nu > 0 \quad \nu = 1, \dots, k\}$ although it will be the same in the examples of Section 4.

b) The assumption that the classical E-optimal information matrix is unique can be replaced by the weaker assumption that the minimum eigenvalues of all classical E-optimal information matrices have the same multiplicity.

c) In general the assertion of Theorem 3.2 is not necessarily true for the sets $\mathcal{S}_m$ and $\bar{\mathcal{K}}_m^{(n)}$ if $m \geq k_0 > 1$ because in this case (3.10) can not be verified. However, in most cases there will exist inball radii of the sets $\mathcal{S}_m$ and $\bar{\mathcal{K}}_m^{(n)}$ with the same direction when n is sufficiently large. This is intuitively motivated from the definition of the set $\bar{\mathcal{K}}_m^{(n)}$.

The results stated so far investigate the case when the classical and the Bayesian E-optimal design are supported at the same set of points (for a sufficiently large number of observations). The following theorem gives a sufficient condition guaranteeing that the two designs are identical.

Theorem 3.5. Let ξ denote a classical E-optimal design, $m \geq k_0$ and $\sqrt{\lambda_{\min}}A$ an inball vector of the set $\mathcal{S}_m$. If $RA = \lambda A$ for some $\lambda > 0$ and $\lambda_{\min} + \frac{\lambda}{n}$ is the minimum eigenvalue of the Bayesian information matrix $M_B(\xi)$, then the design ξ is also Bayesian E-optimal.

Proof. Because $\sqrt{\lambda_{\min}}A$ is an inball radius of the set $\mathcal{S}_m$ we obtain for the supporting hyperplane D to $\mathcal{S}_m$ at $\sqrt{\lambda_{\min}}A$

$$\text{(3.16)} \qquad \frac{A}{\sqrt{\lambda_{\min}}} = D = M^{-1}(\xi)\sqrt{\lambda_{\min}}A$$

(see e.g. Dette and Studden (1992)), which shows that the columns of A are eigenvectors of the classical E-optimal information matrix $M(\xi)$ corresponding to its minimum eigenvalue $\lambda_{\min}$. Theorem 3.4 of Dette and Studden (1992) shows that the design ξ is also (classical) optimal for $A'\theta$ and we obtain from Theorem 2.2 (putting formally $R = 0$) that

$$f'(x)M^{-1}(\xi)AA'M^{-1}(\xi)f(x) \le tr\left(A'M^{-1}(\xi)A\right)$$

for all $x \in \mathcal{X}$ or equivalently (using (3.16) and $tr(A'A) = 1$)

$$f'(x)AA'f(x) \le \lambda_{\min}.$$

Defining $z_i = a_i/\|a_i\|_2$ and $\alpha_i = \|a_i\|_2^2$ where a_i denotes the i-th column of the inball vector A it is straightforward to show that the matrix $E = \sum_{i=1}^m \alpha_i z_i z_i'$ satisfies (3.8). By the assumptions and (3.16) we have for $i = 1, \dots, m$

$$M_B(\xi)z_i = \lambda_{\min} z_i + \frac{\lambda}{n} z_i = \left(\lambda_{\min} + \frac{\lambda}{n}\right) z_i$$

establishing that $z_1, \dots, z_m$ are normalized eigenvectors of the Bayesian information matrix $M_B(\xi)$ corresponding to its minimum eigenvalue $\lambda_{\min} + \frac{\lambda}{n}$. Thus we obtain from (3.8)

$$f'(x)Ef(x) \le \lambda_{\min} = \sum_{i=1}^m \alpha_i z_i' \left(M_B(\xi) - \frac{1}{n}R\right) z_i = \left(\lambda_{\min} + \frac{\lambda}{n}\right) - \frac{1}{n} tr(RE)$$

and the assertion follows from Theorem 2.1. ■

4. Applications and examples. In this section we consider the polynomial regression model on the interval $[-1,1]$, that is $f(x) = (1, x, \dots, x^k)'$ and $\mathcal{X} = [-1,1]$. The classical E-optimal design problem was recently solved by Pukelsheim and Studden (1992).

Corollary 4.1. (Bayesian E-optimal designs for polynomial regression) Let $k \ge 2$, $f(x) = (1, x, \dots, x^k)'$ and $x \in [-1,1]$. There exists an $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$ the design which puts masses

$$\hat{p}_\nu = \left(1 + \frac{1}{n} \sum_{i,j=0}^k (-1)^{i+j} r_{ij}^*\right) p_\nu - \frac{1}{n} \sum_{j=0}^k (-1)^{j+\nu} r_{\nu j}^* \tag{4.1}$$

at the support points $x_\nu = -\cos(\frac{\nu\pi}{k})$ $(\nu = 0, \dots, k)$ is Bayesian E-optimal. Here p_ν denotes the weight of the classical E-optimal design at the point x_ν. The minimum eigenvalue of the Bayesian information matrix is given by

$$\lambda_{\min}^{(n)} = \frac{1}{\|c\|_2^2} \left(1 + \frac{1}{n} c'Rc\right)$$

where c denotes the Chebyshev vector containing the coefficients of the k-th Chebyshev polynomial as coordinates (i.e. $c'f(x) = T_k(x) = \cos(k \arccos x)$).

Proof. Pukelsheim and Studden (1992) proved that the classical E-optimal design is unique and supported at the Chebyshev points $x_\nu = -\cos(\frac{\nu\pi}{k})$. The minimum eigenvalue of the E-optimal information matrix has multiplicity 1. Thus the assertion about the support and the weights follows directly from Theorem 3.3. For the second part we remark that the results of the same authors show that the Chebyshev vectors $\pm\frac{c}{\|c\|_2^2}$ are the only inball vectors of the set $\mathcal{S}_1$ and thus the minimum eigenvalue of the E-optimal information matrix is given by $\lambda_{\min} = \frac{1}{\|c\|_2^2}$. Observing (3.3) (note that A is unique up to the factor ± 1) we obtain for all $n \geq n_0$

$$\lambda_{\min}^{(n)} = \frac{1}{\|c\|_2^2}\left(1 + \frac{1}{n}c'Rc\right)$$

which completes the proof of Corollary 4.1. ■

Remark 4.2. Pukelsheim and Studden (1992) showed that the classical E-optimal design for polynomial regression of degree k on the interval $[-1, 1]$ puts masses

$$p_\nu = \frac{(-1)^{k-\nu}u_\nu}{\|c\|_2^2} \qquad \nu = 0, \ldots, k$$

at the Chebyshev points $x_\nu = -\cos(\frac{\nu\pi}{k})$ where the coefficients u_ν are determined from

$$\sum_{\nu=0}^{k} u_\nu f(x_\nu) = c$$

(note that this is essentially the Elfving representation of the inball vector $\frac{c}{\|c\|_2^2}$ of $\mathcal{S}_1$).

Example 4.3. (Quadratic regression) Let $f(x) = (1, x, x^2)'$ and

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{12} & r_{22} & r_{23} \\ r_{13} & r_{23} & r_{33} \end{pmatrix}$$

denote a positive definite precision matrix. From Remark 4.2 it follows that the classical E-optimal design puts masses $\frac{1}{5}, \frac{3}{5}, \frac{1}{5}$ at the points -1, 0, 1. The matrix F is given by $F = [f(-1), f(0), f(1)]$ and its inverse

$$T = F^{-1} = \frac{1}{2}\begin{pmatrix} 0 & -1 & 1 \\ 2 & 0 & -2 \\ 0 & 1 & 1 \end{pmatrix}$$

which yields

$$R^* = TRT' =$$

$$\frac{1}{4}\begin{pmatrix} r_{33}-2r_{23}+r_{22} & 2(-r_{33}+r_{13}+r_{23}-r_{12}) & r_{33}-r_{22} \\ 2(-r_{33}+r_{13}+r_{23}-r_{12}) & 4r_{33}-8r_{13}+4r_{11} & 2(-r_{33}-r_{23}+r_{13}+r_{12}) \\ r_{33}-r_{22} & 2(-r_{33}-r_{23}+r_{13}+r_{12}) & r_{33}+2r_{23}+r_{22} \end{pmatrix}.$$

By (4.1) and Corollary 4.1 it is now straight forward to show that for sufficiently large n the E-optimal Bayesian design puts masses

$$\begin{aligned}
\hat{p}_0 &= \frac{1}{10}\left(2+\frac{1}{n}\{2r_{11}-5r_{12}-3r_{13}+10r_{23}-2r_{33}\}\right) \\
\hat{p}_1 &= \frac{1}{10}\left(6+\frac{2}{n}\{3r_{13}-2r_{11}+2r_{33}\}\right) \\
\hat{p}_2 &= \frac{1}{10}\left(2+\frac{1}{n}\{2r_{11}+5r_{12}-3r_{13}-10r_{23}-2r_{33}\}\right)
\end{aligned}$$

at the points $x_0 = -1,\ x_1 = 0,\ x_2 = 1$. Using the equivalence Theorem 3.1 it is easy to see that the design

$$\hat{\xi} = \begin{pmatrix} -1 & 0 & 1 \\ \hat{p}_0 & \hat{p}_1 & \hat{p}_2 \end{pmatrix}$$

is Bayesian E-optimal whenever $n \geq n_0 = \max\{n|\hat{p}_\nu > 0\}$ (note that in general n_0 is not determined by the assumption that the weights $\hat{p}_\nu$ have to be positive). The minimum eigenvalue is given by (note that the Chebyshev vector is $(-1,0,2)'$)

$$\lambda_{\min}^{(n)} = \frac{1}{5}\left(1+\frac{1}{n}[4r_{33}-4r_{13}+r_{11}]\right).$$

<u>**Corollary 4.4.**</u> Let $k \geq 2,\ f(x) = (1,x,\ldots,x^k)',\ x \in \mathcal{X},\ c \in \mathbb{R}^{k+1}$ denote the Chebyshev vector (i.e. $c'f(x) = T_k(x)$) and ξ be the classical E-optimal design. If the precision matrix R satisfies $Rc = \lambda c$ for some $\lambda > 0$ and $\frac{1}{\|c\|_2^2} + \frac{\lambda}{n}$ is the minimum eigenvalue of the Bayesian information matrix $M_B(\xi)$, then the classical and Bayesian E-optimal design are identical.

<u>**Proof.**</u> It follows from the results of Pukelsheim and Studden (1992) that the minimum eigenvalue of the classical E-optimal information matrix is $\lambda_{\min} = \frac{1}{\|c\|_2^2}$ and has multiplicity 1. An inball vector of $\mathcal{S}_1$ is given by $\frac{c}{\|c\|_2^2}$ and the assertion now follows directly from Theorem 3.5. ∎

Example 4.5. (Quadratic regression) Let $k = 2$, $f(x) = (1, x, x^2)'$ and consider a precision matrix R of the form

$$R = \begin{pmatrix} 5 & 2 & 2 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \end{pmatrix}.$$

The Chebyshev vector $c = (-1, 0, 2)'$ is an eigenvector of R with eigenvalue 1 and thus Corollary 4.4 is applicable. From Example 4.3 we obtain that the classical E-optimal design puts masses $\frac{1}{5}, \frac{3}{5}, \frac{1}{5}$ at the points $-1, 0, 1$ which yields for the classical E-optimal information matrix and for the Bayesian information matrix of ξ

$$M(\xi) = \begin{pmatrix} 1 & 0 & \frac{2}{5} \\ 0 & \frac{2}{5} & 0 \\ \frac{2}{5} & 0 & \frac{2}{5} \end{pmatrix}, \quad M_B(\xi) = \frac{1}{5} \begin{pmatrix} 5 + \frac{25}{n} & \frac{10}{n} & 2 + \frac{10}{n} \\ \frac{10}{n} & 2 + \frac{5}{n} & \frac{5}{n} \\ 2 + \frac{10}{n} & \frac{5}{n} & 2 + \frac{10}{n} \end{pmatrix}.$$

Straight forward calculations show that the eigenvalues of $M_B(\xi)$ are given by

$$\lambda_1 = \frac{1}{5} + \frac{1}{n}, \quad \lambda_{2/3} = \frac{4}{5} + \frac{7}{2n} \pm \sqrt{\frac{4}{25} + \frac{2}{n} + \frac{45}{4n^2}}$$

where the eigenvector corresponding to λ_1 is the Chebyshev vector $c = (-1, 0, 2)'$. For $n \geq 4$, λ_1 is the minimum eigenvalue of $M_B(\xi)$ and the design ξ which puts masses $\frac{1}{5}, \frac{3}{5}, \frac{1}{5}$ at the points $-1, 0, 1$ is the Bayesian E-optimal design for all $n \geq 4$.

Remark 4.6. The results of Corollary 4.1 and 4.4 carry over for polynomial regression on the interval $[a, b]$, where $0 \leq a < b$ (see the discussion in Pukelsheim and Studden (1992)). Note also that these corollaries will also hold for the weighted polynomial regression models $f(x) = \sqrt{\lambda(x)}\,(1, x, \ldots, x^n)$ ($x \in [-1, 1]$) where the efficiency function $\lambda(x)$ is one of the functions $1 - x$, $1 + x$, $1 - x^2$. In these models the classical E-optimal designs were recently determined by Dette (1993).

Example 4.7. (Weighted linear regression) In our final example we will illustrate the application of Theorem 3.1. Let $f(x) = \sqrt{4 - x^2}(1, x)'$ and $x \in [-2, 2]$, then it is straight forward to show that the classical E-optimal design ξ puts masses at the two points -1 and 1 with information matrix

$$M(\xi) = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix},$$

and that a matrix E satisfying (3.8) is given by

$$E = \frac{1}{3} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} = \frac{2}{3} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (0, 1) + \frac{1}{3} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (0, 1).$$

From Theorem 3.3 of Dette and Studden we have that the matrix

$$A = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{pmatrix}$$

defines an inball vector of the Elfving set $\mathcal{S}_2$. For the matrix

$$R = \begin{pmatrix} r_1 & r_2 \\ r_2 & r_1 \end{pmatrix}$$

it can easily be shown that

$$R^* = TRT' = \begin{pmatrix} r_1 - r_2 & 0 \\ 0 & r_1 + r_2 \end{pmatrix}$$

is diagonal, and we obtain for the quantities defined by (3.1)

$$\hat{p}_1 = \frac{1}{2}\left(1 + \frac{r_2}{3n}\right), \quad \hat{p}_2 = \frac{1}{2}\left(1 - \frac{r_2}{3n}\right).$$

Observing the discussion following the proof of Theorem 3.1 it is straight forward to show (by an application of Theorem 2.1) that the design ξ_B, which puts masses $\hat{p}_1$ and $\hat{p}_2$ at the support points -1 and 1 of the classical E-optimal design, is Bayesian E-optimal whenever $n \geq \frac{r_2}{3}$ (note that this statement could also be obtained by a direct application of Theorem 3.1 showing that for $n \geq \frac{r_2}{3}$ there exists an inball vector $A^{(n)}$ of the Bayesian Elfving set $\bar{\mathcal{K}}_2^{(n)}$ with the same direction as the vector A).

*

References

P. Billingsley (1968), *Convergence of Probability Measures*, Wiley, New York.

K. Chaloner (1984), Optimal Bayesian experimental design for linear models, *Ann. Statist.* **12**, 283-300.

H. Dette (1992), Geometric characterizations of model robust designs, *Habilitationsschrift*, Universität Göttingen.

H. Dette (1993), E-optimal designs for weighted polynomial regression, *Ann. Statist.*, to appear.

H. Dette, W.J. Studden (1992), Geometry of E-optimality, *Ann. Statist.*, to appear.

Elfving (1952), Optimum allocation in linear regression, *Ann. Math. Statist.* **23**, 255-263.

S.M. El-Krunz, W.J. Studden (1991), Bayesian optimal designs for linear regression models, *Ann. Statist.* **19**, 2183-2208.

J. Fellman (1974), On the allocation of linear observations, *Commentationes Physico-Mathematicae (Finska Vetenskaps Societeten)*, **44** 27-77.

J. Pilz (1991), *Bayesian estimation and experimental design in linear regression models*, Wiley, New York.

F. Pukelsheim (1980), On linear regression designs which maximize information, *J. Statist. Plan. Inf.* **4**, 339-364.

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

F. Pukelsheim, W.J. Studden (1992), E-optimal designs for polynomial regression, *Ann. Statist.*, to appear.

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

From seele Thu Feb 11 15:59:45 1993
Received: by pop.stat.purdue.edu (4.1/Purdue_CC)
id AA07181; Thu, 11 Feb 93 15:59:42 EST
Date: Thu, 11 Feb 93 15:59:42 EST
From: seele (Teena Seele)
Message-Id: <9302112059.AA07181@pop.stat.purdue.edu>
To: nlucas@pop.stat.purdue.edu
Subject: studden
Status: R

He said to leave the tech report 92-27 files alone. He said that the symposium is a shorter version of the tech report and for now he would rather keep the files separate.

FYI

Teena