BAYESIAN OPTIMAL DESIGNS
IN
LINEAR REGRESSION MODELS

by

Sadi M. El–Krunz* and William J. Studden *

Technical Report #89-29

Department of Statistics
Purdue University

August 1989

* Research supported by National Science Foundation under Grant No. DMS 88–02535

⋆

1. Introduction

Consider the linear regression model

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

where $\boldsymbol{f}'(x) = (f_1(x), \ldots, f_k(x))$, x is the control variable, $\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_k)$ is the vector of unknown parameters, and ϵ is a normally distributed random variable with mean 0 and variance σ^2 independent of x. We assume that $\mathcal{X}$ is a compact set, containing at least k points, with Borel field containing all one point sets. The regression functions $f_1, f_2, \ldots, f_k$, are k linearly independent real valued continuous functions on the design space $\mathcal{X}$, which are assumed to be known to the experimenter. As usual, uncorrelated observations $y_1, y_2, \ldots, y_n$ on the dependent random variable y, are taken at levels $x_1, x_2, \ldots, x_n \in \mathcal{X}$, respectively, and the n–dimensional random vector $\boldsymbol{y} = (y_1, \ldots, y_n)'$ is assumed to have a normal distribution with mean vector $X\boldsymbol{\theta}$ and covariance matrix $\sigma^2 I$, where $X = (f(x_1), \ldots, f(x_n))'$ is the $n \times k$ design matrix and I is the $n \times n$ identity matrix. We also assume that a prior distribution $\pi(\boldsymbol{\theta}, \sigma^2)$ on $\boldsymbol{\theta}, \sigma^2$ is given such that the conditional prior distribution $\pi(\boldsymbol{\theta}|\sigma^2)$ of $\boldsymbol{\theta}$ given σ^2 is $N(\boldsymbol{\mu}, \sigma^2 R^{-1})$, where R is a given positive definite $k \times k$ "precision" matrix. Under the above assumptions the posterior conditional distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y}, \sigma^2)$ of $\boldsymbol{\theta}$ given $\boldsymbol{y}$, σ^2 is normal with mean vector

$$\hat{\boldsymbol{\theta}}_b = E(\boldsymbol{\theta}|\boldsymbol{y}, \sigma^2) = (X'X + R)^{-1}(X'\boldsymbol{y} + R\boldsymbol{\mu}) \tag{1.2}$$

and covariance matrix $\sigma^2(X'X + R)^{-1}$. Thus, if we are interested in estimating $A'\boldsymbol{\theta}$ for some $k \times s$ matrix A of full rank s, $1 \le s \le k$, then under squared error loss and with σ^2 either known or $E(\sigma^2)$ finite, the best estimator of $A'\boldsymbol{\theta}$ is $A'\hat{\boldsymbol{\theta}}_b$ and the expected posterior risk of $A'\hat{\boldsymbol{\theta}}_b$ is given by

$$E(\sigma^2) tr AA'(X'X + R)^{-1}. \tag{1.3}$$

Thus for the Bayes estimator $\hat{\boldsymbol{\theta}}_b$ (or any estimator with covariance structure proportional to $(X'X + R)^{-1}$ with a specified positive definite $k \times k$ matrix R), a reasonable criterion of optimality is then to choose X to minimize some appropriate functional Φ of the matrix $(X'X+R)$. For a more complete discussion of the above model and of the optimality and robustness of the Bayes estimator $\hat{\boldsymbol{\theta}}_b$ see Pilz (1983), Chaloner (1982,1984), Duncan and DeGroot (1976) and Sinha (1970).

We are concerned here with the approximate design theory wherein the designs are the class Ξ of probability measures ξ on $\mathcal{X}$ and the Bayesian information matrix (per unit observation) of the design ξ is $M_b(\xi) = M(\xi) + \frac{1}{n}R$, where $M(\xi) = \int_{\mathcal{X}} \boldsymbol{f}(x)\boldsymbol{f}'(x)\xi(dx)$ and n is the total number of observations. Thus $M_b(\xi)$ is a positive definite $k \times k$ matrix and $M_b^{-1}(\xi)$ is proportional to the posterior covariance matrix of the Bayes estimator of $\boldsymbol{\theta}$. We let $\mathcal{M}_b = \{M_b(\xi) \colon \xi \in \Xi\}$. Then the family $\mathcal{M}_b$ of all Bayesian information matrices is a convex compact set. It is the closed convex hull of the set $\{\boldsymbol{f}(x)\boldsymbol{f}'(x) + \frac{1}{n}R \colon x \in \mathcal{X}\}$ of Bayesian information matrices corresponding to one point designs. Moreover if h is the dimension of the vector space generated by the products $\{f_i f_j\}_{i,j=1}^{k}$ of the regression functions $f_1, f_2, \ldots, f_k$, then for any design $\xi \in \Xi$, the Bayesian information matrix $M_b(\xi)$ can be represented in the form $M_b(\xi) = \sum_{i=1}^{m} p_i(\frac{1}{n}R + \boldsymbol{f}(x_i)\boldsymbol{f}'(x_i))$, $m \le h+1 \le \frac{k(k+1)}{2}+1$, $0 < p_i < 1$, $\sum_{i=1}^{m} p_i = 1$. If $M_b(\xi)$ is a boundary point of $\mathcal{M}_b$, then $h+1$ and $\frac{k(k+1)}{2}+1$ can be replaced by h and $\frac{k(k+1)}{2}$ respectively. Usually the optimality criterion Φ (to be minimized) is finite on $\mathcal{M}_b$ and satisfies the following conditions:

1) ϕ is convex on $\mathcal{M}_b$, that is, if ξ and $\eta \in \Xi$ then for any $0 \le \alpha \le 1$, we have

$$\Phi((1-\alpha)M_b(\xi) + \alpha M_b(\eta)) \le (1-\alpha)\Phi(M_b(\xi) + \alpha M_b(\eta) \tag{1.4}$$

2) Φ is non–increasing in the sense that if $M_1 - M_2$ is non–negative definite, then $\Phi(M_1) \le \Phi(M_2)$.

3) Φ is homogeneous of positive degree p, that is, for any $a > 0$, $\Phi(aM) = a^{-p}\Phi(M)$.

4) Φ is continuous and differentiable everywhere on $\mathcal{M}_b$.

The Bayesian optimal design problem is then to characterize the designs ξ_0 which are Bayesian Φ–optimum; that is,

$$\Phi(M_b(\xi_0)) = \inf_{\xi \in \Xi} \Phi(M_b(\xi)) \tag{1.5}$$

The convexity of $\mathcal{M}_b$ ensure that there always exists a Bayesian Φ–optimal design supported by $m \leq h \leq \frac{k(k+1)}{2}$ points. The main purpose of this paper is to study a Bayesian version of Elfving's Theorem for the $\boldsymbol{c}$–optimality criterion and discuss some of its implications. This criterion is given by

$$\Phi(M_b(\xi)) = \boldsymbol{c}' M_b^{-1}(\xi)\boldsymbol{c}, \ \xi \in \Xi \tag{1.6}$$

and corresponds to the case where one is interested in estimating a linear combination of the form $\boldsymbol{c}'\boldsymbol{\theta}$ for some nonrandom $k \times 1$ vector $\boldsymbol{c}$. The famous Elfving's Theorem for classical (non–Bayesian) $\boldsymbol{c}$–optimal designs is the following:

<u>Theorem 1.1</u> (Elfving 1952)

Let $\mathcal{G}$ = the symmetric convex hull of $f(\mathcal{X})$. A design ξ^* is classical $\boldsymbol{c}$–optimum (in the sense that it minimizes $\boldsymbol{c}'M^-(\xi)\boldsymbol{c}$ over all designs ξ for which $\boldsymbol{c}'\theta$ is estimable) if and only if there exists a measurable real valued function $\epsilon^*(x)$ satisfying $|\epsilon^*(x)| = 1$ such that (i) $\int \epsilon^*(x)\boldsymbol{f}(x)\xi^*(dx) = \beta^*\boldsymbol{c}$ for some positive constant β^* and (ii) $\beta^*\boldsymbol{c}$ is a boundary point of $\mathcal{G}$. Moreover $\beta^*\boldsymbol{c}$ lies on the boundary of $\mathcal{G}$ if and only if $\inf_\xi \boldsymbol{c}'M^-(\xi)\boldsymbol{c} = {\beta^*}^{-2}$. ■

Definitions and some preliminary lemmas are given in Section 2. Duality theory is used to derive a Bayesian version of Elfving's Theorem for Bayesian $\boldsymbol{c}$–optimality in Section 3, where emphasis is laid on the geometry inherent in the Bayesian $\boldsymbol{c}$–optimal design problem and the parallelism between classical and Bayesian $\boldsymbol{c}$–optimal design theory is illustrated. Conditions under which a one point design is Bayesian $\boldsymbol{c}$–optimum are given in Section 4. In Section 5 the class of prior precision matrices R for which the Bayesian $\boldsymbol{c}$–optimal designs are supported by the points of the classical $\boldsymbol{c}$–optimal design is characterized. It is also proved that the Bayesian $\boldsymbol{c}$–optimal design, for n large enough, is always supported by the same support points of the classical $\boldsymbol{c}$–optimal design ξ^* if ξ^* is supported at exactly k distinct points and for a large class of prior precision matrices R if ξ^* is supported at

$1 \leq m < k$ points. In Section 6 the geometry–duality approach is extended for the Ψ–optimality criterion and a matrix analog of the geometric result of Elfving is derived and in Section 7 its applications are discussed.

2. Definitions and Preliminary Lemmas

Assume that we are interested in the estimation of parametric functions of the form $\boldsymbol{c}'\boldsymbol{\theta}$, where $\boldsymbol{c}$ is an arbitrary nonrandom $k \times 1$ vector. Let $\mathcal{G}$ be the convex closure of the set of all vectors $\epsilon \boldsymbol{f}(x)$, $x \in \mathcal{X}$, $\epsilon \in \{\pm 1\}$, that is, $\mathcal{G}$ is the symmetric convex hull of $f(\mathcal{X})$. Thus $\mathcal{G}$ is a symmetric convex compact subset of k–dimensional Euclidean space and every vector $\boldsymbol{a} \in \mathcal{G}$ has a representation

$$\boldsymbol{a} = \sum_{i=1}^{m} \epsilon_i p_i \boldsymbol{f}(x_i) \tag{2.1}$$

for some positive integer m, $p_i > 0$, $x_i \in \mathcal{X}$, $\epsilon_i \in \{\pm 1\}$ and $\sum_{i=1}^{m} p_i = 1$. From our assumptions on $\mathcal{X}$ and $\boldsymbol{f}$ it is evident that the set $\mathcal{G}$ has the origin in its interior and every half line through the origin intersects the boundary of $\mathcal{G}$ at exactly one point and so for any nonzero $k \times 1$ vector $\boldsymbol{c}$, there exists a unique positive constant β^* such that $\beta^* \boldsymbol{c} \in \partial\mathcal{G}$ = boundary of $\mathcal{G}$. The following simple lemma characterizes the boundary points of the set $\mathcal{G}$.

Lemma 2.1

A vector $\boldsymbol{a}$ of the form (2.1) is a boundary point of $\mathcal{G}$ if and only if there exists a $k \times 1$ vector $\boldsymbol{d}$ such that

$$|\boldsymbol{d}'\boldsymbol{f}(x)| \leq 1 \text{ for all } x \in \mathcal{X} \tag{2.2}$$

and equality holds for each x_i with $\epsilon_i = \boldsymbol{d}'\boldsymbol{f}(x_i)$, $i = 1, 2, \ldots, m$ and $\boldsymbol{a}'\boldsymbol{d} = 1$. ■

Proof: See Lemma 2.1 of Studden (1968).

The vector $\boldsymbol{d}$ given in Lemma 2.1 defines the hyperplane $\{\boldsymbol{u}: \boldsymbol{u} \in \mathbb{R}^k, \boldsymbol{d}'\boldsymbol{u} = 1\}$ supporting $\mathcal{G}$ at its boundary $\boldsymbol{a}$, that is, $\boldsymbol{d}'\boldsymbol{u} \leq 1 = \boldsymbol{d}'\boldsymbol{a}$ for all $\boldsymbol{u} \in \mathcal{G}$. Identifying a hyperplane with its inducing vector $\boldsymbol{d}$, we define

$$\mathcal{D} = \{\boldsymbol{d}: \boldsymbol{d}'\boldsymbol{u} \leq 1 \text{ for all } \boldsymbol{u} \in \mathcal{G} \text{ and } \boldsymbol{d}'\boldsymbol{u}_0 = 1 \text{ for some } \boldsymbol{u}_0 \in \partial\mathcal{G}\} \tag{2.3}$$

to be the set of all normalized supporting hyperplanes to the surface of $\mathcal{G}$. For every $\boldsymbol{d} \in \mathcal{D}$, define the contact set

$$\mathcal{C}(\boldsymbol{d}) = \{\boldsymbol{u}: \boldsymbol{u} \in \partial\mathcal{G},\ \boldsymbol{d}'\boldsymbol{u} = 1\} \tag{2.4}$$

to be the intersection of the hyperplane $\boldsymbol{d}$ with $\mathcal{G}$ and for any point $\boldsymbol{u}_0 \in \partial\mathcal{G}$, let $\mathcal{D}_{\boldsymbol{u}_0} = \{\boldsymbol{d}: \boldsymbol{d}'\boldsymbol{u} \leq 1 = \boldsymbol{d}'\boldsymbol{u}_0 \text{ for all } \boldsymbol{u} \in \mathcal{G}\}$ denote the set of all supporting hyperplanes to $\mathcal{G}$ at $\boldsymbol{u}_0$. The set $\mathcal{D}_{\boldsymbol{u}_0}$ is either a single point or a closed convex set and $\mathcal{D} = \bigcup_{\boldsymbol{u}\in\partial\mathcal{G}} \mathcal{D}_{\boldsymbol{u}}$. Now let R be a given $k \times k$ positive definite matrix, n be a given positive integer and let us define the following:

$$\mathcal{H} = \left\{\boldsymbol{z}: \boldsymbol{z} = \boldsymbol{u} + \tfrac{1}{n}R\boldsymbol{d},\ \boldsymbol{d} \in \mathcal{D} \text{ and } \boldsymbol{u} \in \mathcal{C}(d)\right\} \tag{2.5}$$

$$\mathcal{D}^* = \left\{\boldsymbol{d}^*: \boldsymbol{d}^* = \left(1 + \tfrac{1}{n}\boldsymbol{d}'R\boldsymbol{d}\right)^{-\frac{1}{2}} \boldsymbol{d},\ \boldsymbol{d} \in \mathcal{D}\right\} \tag{2.6}$$

$$\mathcal{V} = \left\{\boldsymbol{v}: \boldsymbol{v} = \left(1 + \tfrac{1}{n}\boldsymbol{d}'R\boldsymbol{d}\right)^{-\frac{1}{2}} (\boldsymbol{u} + \tfrac{1}{n}R\boldsymbol{d}),\ \boldsymbol{d} \in \mathcal{D} \text{ and } \boldsymbol{u} \in \mathcal{C}(\boldsymbol{d})\right\} \tag{2.7}$$

It is important to note the dependence of $\boldsymbol{u}$ and $\boldsymbol{d}$ in the definition of $\mathcal{H}$ and $\mathcal{V}$. In addition both sets depend on the precision matrix R. For any set $\mathcal{A}$, we shall use the notation $\overline{\mathcal{A}}$ to mean the set

$$\overline{\mathcal{A}} = \{t\boldsymbol{a}: \boldsymbol{a} \in \mathcal{A},\ 0 \leq t \leq 1\} \tag{2.8}$$

We show in Lemma 2.5 that $\mathcal{G} \subseteq \overline{\mathcal{V}} \subseteq \overline{\mathcal{H}}$. The sets $\overline{\mathcal{V}}$ and $\overline{\mathcal{H}}$ will serve as Bayesian analogs of $\mathcal{G}$. The set $\overline{\mathcal{V}}$ will be shown to be the convex hull of $\mathcal{V}$. The set $\mathcal{V}$ is just the "normalized" version of $\mathcal{H}$. The set $\overline{\mathcal{H}}$ is not convex in general and seems to be more useful in practice than $\overline{\mathcal{V}}$.

We let $\{x_i, p_i\}_{i=1}^m$ denote the design ξ which puts weights $p_i > 0$ at the points $x_i \in \mathcal{X}, i = 1, \ldots, m$. The following version of the equivalence theorem for Bayesian $\boldsymbol{c}$–optimal designs; see Chaloner (1984), Pilz (1981) or El–Krunz (1989), will be needed.

Lemma 2.2

The design $\xi_0 = \{x_i, p_i\}_{i=1}^m$ is Bayesian $\boldsymbol{c}$–optimum if and only if

$$|\boldsymbol{f}'(x)M_b^{-1}(\xi_0)\boldsymbol{c}|^2 \leq \boldsymbol{c}'M_b^{-1}(\xi_0)M(\xi_0)M_b^{-1}(\xi_0)\boldsymbol{c} \text{ for all } x \in \mathcal{X} \tag{2.9}$$

and equality holds for each $x_i,\ i = 1, 2, \ldots, m$, in the spectrum of the design ξ_0.

Lemma 2.3

For any nonzero $k \times 1$ vector $\boldsymbol{c}$, there exists a positive constant γ_0 such that

$$\sum_{i=1}^m \epsilon_i p_i \boldsymbol{f}(x_i) + \tfrac{1}{n}R\boldsymbol{d}_0 = \gamma_0\boldsymbol{c} \in \mathcal{H} \tag{2.10}$$

for some $\boldsymbol{d}_0 \in \mathcal{D}$ and some positive integer m, where $p_i > 0,\ \epsilon_i d_0' f(x_i) = 1$ and $\sum\limits_{i=1}^m p_i = 1$, $(u_0 = \Sigma\epsilon_i p_i \boldsymbol{f}(x_i) \in \partial\mathcal{G})$.

Proof: Let $\xi_0 = \{x_i, p_i\}_{i=1}^m$ be the Bayesian $\boldsymbol{c}$–optimal design. As indicated in Section 1, Bayesian $\boldsymbol{c}$–optimal designs always exist. Thus it follows from Lemma 2.2 that (2.9) holds for the design ξ_0. Let $\gamma_0^{-2} = \boldsymbol{c}'M_b^{-1}(\xi_0)M(\xi_0)M_b^{-1}(\xi_0)\boldsymbol{c}$ and let $\boldsymbol{d}_0 = \gamma_0 M_b^{-1}(\xi_0)\boldsymbol{c}$. Then it follows from (2.9) that $|\boldsymbol{d}_0'\boldsymbol{f}(x)| \leq 1$ for all $x \in \mathcal{X}$ and $|\boldsymbol{d}_0'\boldsymbol{f}(x_i)| = 1,\ i = 1, 2, \ldots, m$ and so $\boldsymbol{d}_0 \in \mathcal{D}$. Let $\epsilon_i = \boldsymbol{d}_0'\boldsymbol{f}(x_i),\ i = 1, 2, \ldots, m$, then $\epsilon_i \in \{\pm 1\},\ i = 1, 2, \ldots, m$. Since $M_b(\xi_0)\boldsymbol{d}_0 = \gamma_0\boldsymbol{c}$ and $M_b(\xi_0)\boldsymbol{d}_0 = M(\xi_0)\boldsymbol{d}_0 + \frac{1}{n}R\boldsymbol{d}_0 = \sum\limits_{i=1}^m p_i\boldsymbol{f}(x_i)\boldsymbol{f}'(x_i)\boldsymbol{d}_0 + \frac{1}{n}R\boldsymbol{d}_0 = \sum\limits_{i=1}^m \epsilon_i p_i \boldsymbol{f}(x_i) + \frac{1}{n}R\boldsymbol{d}_0$, then $\sum\limits_{i=1}^m \epsilon_i p_i \boldsymbol{f}(x_i) + \frac{1}{n}R\boldsymbol{d}_0 = \gamma_0\boldsymbol{c}$ and so $\gamma_0\boldsymbol{c} \in \mathcal{H}$ which completes the proof. ∎

If in (2.10), we let $\beta_0 = (\boldsymbol{d}_0'\boldsymbol{c})^{-1}$, then premultiplication of both sides of (2.10) by $\boldsymbol{d}_0'$ gives

$$\gamma_0 = \beta_0(1 + \tfrac{1}{n}\boldsymbol{d}_0'R\boldsymbol{d}_0) \tag{2.11}$$

Note that $\beta_0\boldsymbol{c}$ lies on the hyperplane $\boldsymbol{d}_0 \in \mathcal{D}$. Since $\boldsymbol{d}_0$ supports the convex set $\mathcal{G}$ at $\boldsymbol{u}_0 := \Sigma\epsilon_i p_i \boldsymbol{f}(x_i)$ then $\beta_0 \geq \beta^*$ if β^* is defined such that $\beta^*\boldsymbol{c} \in \partial\mathcal{G}$. Furthermore since $\gamma_0 > \beta_0$ then $\gamma_0\boldsymbol{c}$ and $\mathcal{G}$ lie on opposite sides of the support plane $\boldsymbol{d}_0$.

The set $\overline{\mathcal{D}}$ is a "convex dual" of the set $\mathcal{G}$ and its boundary $\mathcal{D}$ is a "convex dual" of $\partial\mathcal{G}$ as it is evident from the following lemma.

Lemma 2.4

The set $\overline{\mathcal{D}}$ is a compact symmetric convex set in R^k which has $\mathcal{D}$ as its boundary. Moreover for any nonzero $k \times 1$ vector $\boldsymbol{c}$, there exists a unique positive constant α such that $\alpha \boldsymbol{c} \epsilon \mathcal{D}$.

Proof: Let $\boldsymbol{d}$ be any element of $\mathcal{D}$. Then $\boldsymbol{d}$ is a supporting hyperplane to $\mathcal{G}$ at some point $\boldsymbol{u} \epsilon \partial \mathcal{G}$. Since G is symmetric, then $-\boldsymbol{u} \epsilon \partial \mathcal{G}$ and so $-\boldsymbol{d}$ is a supporting hyperplane to $\mathcal{G}$ at the point $-\boldsymbol{u} \epsilon \partial \mathcal{G}$ and so $\mathcal{D}$ is symmetric which implies that $\overline{\mathcal{D}}$ is symmetric. In fact from the definition of $\mathcal{D}$, we have

$$\overline{\mathcal{D}} = \{\boldsymbol{d} \colon |\boldsymbol{d}'\boldsymbol{u}| \leq 1 \text{ for all } \boldsymbol{u} \epsilon \partial \mathcal{G}\}. \tag{2.12}$$

Thus for any $\boldsymbol{d}_1, \boldsymbol{d}_2 \epsilon \overline{\mathcal{D}}$ and $0 < \lambda < 1$, we have

$$|(\lambda \boldsymbol{d}_1 + (1-\lambda)\boldsymbol{d}_2)'\boldsymbol{u}| \leq \lambda |\boldsymbol{d}_1'\boldsymbol{u}| + (1-\lambda)|\boldsymbol{d}_2'\boldsymbol{u}| \leq 1 \text{ for all } \boldsymbol{u} \epsilon \mathcal{G} \text{ and so } \lambda \boldsymbol{d}_1 + (1-\lambda)\boldsymbol{d}_2 \epsilon \overline{\mathcal{D}}$$

which implies that $\overline{\mathcal{D}}$ is convex. Since $\mathcal{G}$ is a compact subset of R^k with the origin in its interior, then it follows from (2.12) that $\overline{\mathcal{D}}$ is also a compact subset of R^k. For any $\boldsymbol{u}_0 \epsilon \partial \mathcal{G}$, let $\boldsymbol{a} = \sup\{\boldsymbol{u}'\boldsymbol{u}_0 \colon \boldsymbol{u} \epsilon \mathcal{G}\}$. Since $\mathcal{G}$ is compact, then this supremum exists and is attained at some point $\boldsymbol{u}_1 \epsilon \partial \mathcal{G}$ and so $\boldsymbol{d} = a^{-1}\boldsymbol{u}_0$ is a supporting hyperplane to $\mathcal{G}$ at the point $\boldsymbol{u}_1 \epsilon \partial \mathcal{G}$. Thus for every $\boldsymbol{u} \epsilon \partial \mathcal{G}$, there exists a positive constant a such that $a^{-1}\boldsymbol{u} \epsilon \mathcal{D}$ which implies that any half line through the origin intersects $\mathcal{D}$. Thus it follows from this and the convexity of $\overline{\mathcal{D}}$ that any half line through the origin intersects $\mathcal{D}$ at exactly one point and $\mathcal{D}$ is the boundary of $\overline{\mathcal{D}}$. Thus for any nonzero $k \times 1$ vector $\boldsymbol{c}$, there exists a unique positive constant α such that $\alpha \boldsymbol{c} \epsilon \mathcal{D}$ which completes the proof.

Lemma 2.5

The set $\overline{\mathcal{V}}$ a convex compact set and $\mathcal{G} \subset \overline{\mathcal{V}} \subset \overline{\mathcal{H}}$.

Proof: Let $\boldsymbol{c}$ be any nonzero $k \times 1$ vector. From the definition of $\mathcal{G}$, there exists a positive constant β^* such that $\beta^* \boldsymbol{c} \epsilon \partial \mathcal{G}$ and from Lemma 2.3, there exists a positive constant $\gamma_0 = \beta_0(1 + \frac{1}{n}\boldsymbol{d}_0' R \boldsymbol{d}_0)$ such that $\gamma_0 \boldsymbol{c} \epsilon \mathcal{H}$. From Lemma 2.3 and the definition of $\mathcal{V}$ it follows that the positive constant $\delta_0 = \beta_0(1 + \frac{1}{n}\boldsymbol{d}_0' R \boldsymbol{d}_0)^{\frac{1}{2}}$ is such that $\delta_0 \boldsymbol{c} \epsilon \mathcal{V}$. Since $\beta_0 \boldsymbol{c}$ lies on the supporting hyperplane to $\mathcal{G}$ at the point $\boldsymbol{u}_0$ for which $\boldsymbol{u}_0 + \frac{1}{n}\boldsymbol{R}\boldsymbol{d}_0 = \gamma_0\boldsymbol{c}$, then $\beta^* \leq \beta_0$ and since R is positive definite, then $\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0 > 0$ and so $1 + \frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0 > 1$ which implies that $\beta^* < \delta_0 < \gamma_0$ and so $\mathcal{G} \subset \overline{\mathcal{V}} \subset \overline{\mathcal{H}}$. Note that the set inclusion is actually "strict in every direction." Since $\mathcal{G}$ and $\overline{\mathcal{D}}$ are compact it follows that $\overline{\mathcal{V}}$ is bounded and closed and so compact. To prove that the set $\overline{\mathcal{V}}$ is convex it is enough to show that there exists a supporting hyperplane to $\overline{\mathcal{V}}$ at every point $\boldsymbol{v}\epsilon\mathcal{V}$. So let $\boldsymbol{v}_0 = \frac{\boldsymbol{u}_0 + \frac{1}{n}\boldsymbol{R}\boldsymbol{d}_0}{\sqrt{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0}}$ be any point on $\mathcal{V}$. We will show that $\boldsymbol{v}'\boldsymbol{d}_0^* \leq 1 = \boldsymbol{v}_0'\boldsymbol{d}_0^*$ for all $\boldsymbol{v} = \frac{\boldsymbol{u}+\frac{1}{n}\boldsymbol{R}\boldsymbol{d}}{\sqrt{1+\frac{1}{n}\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d}}}\epsilon\mathcal{V}$ where $\boldsymbol{d}_0^* = \frac{\boldsymbol{d}_0}{\sqrt{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0}}$. Since $\boldsymbol{v}_0'\boldsymbol{d}_0^* = \frac{\boldsymbol{d}_0'}{\sqrt{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0}} \cdot \frac{\boldsymbol{u}_0+\frac{1}{n}\boldsymbol{R}\boldsymbol{d}_0}{\sqrt{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0}} = \frac{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0}{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0} = 1$ and $\boldsymbol{v}'\boldsymbol{d}_0^* = \frac{\boldsymbol{d}_0'}{\sqrt{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0}} \cdot \frac{\boldsymbol{u}+\frac{1}{n}\boldsymbol{R}\boldsymbol{d}}{\sqrt{1+\frac{1}{n}\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d}}}$, then it is enough to show that $\boldsymbol{d}_0'\boldsymbol{u} + \frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d} \leq \sqrt{1+\frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0}\sqrt{1+\frac{1}{n}\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d}}$. Since R is positive definite, then $(\boldsymbol{d}-\boldsymbol{d}_0)'R(\boldsymbol{d}-\boldsymbol{d}_0) \geq 0$ with equality occurring if and only if $\boldsymbol{d} = \boldsymbol{d}_0$ and so

$$2\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d} \leq \boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0 + \boldsymbol{d}'\boldsymbol{R}\boldsymbol{d} \tag{2.13}$$

Also from Schwartz's inequality, it is immediate that

$$(\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d})^2 \leq (\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0)(\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d}) \tag{2.14}$$

with equality holding if and only if $\boldsymbol{d} = \boldsymbol{d}_0$. Combining (2.13) and (2.14) it follows that

$$1 + \tfrac{2}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d} + \tfrac{1}{n^2}(\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d})^2 \leq 1 + \tfrac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0 + \tfrac{1}{n}\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d} + \tfrac{1}{n^2}(\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0)(\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d})$$

which implies that

$$(1 + \tfrac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d})^2 \leq (1 + \tfrac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0)(1 + \tfrac{1}{n}\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d})$$

and so

$$1 + \tfrac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d} \leq \sqrt{1 + \tfrac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}_0} \cdot \sqrt{1 + \tfrac{1}{n}\boldsymbol{d}'\boldsymbol{R}\boldsymbol{d}} \tag{2.15}$$

Since $\boldsymbol{d}_0'(u + \frac{1}{n}\boldsymbol{R}\boldsymbol{d}) \leq 1 + \frac{1}{n}\boldsymbol{d}_0'\boldsymbol{R}\boldsymbol{d}$, the result follows.

Remark 2.1

The mapping from $\mathcal{D}$ to $\mathcal{D}^*$ defined by $\boldsymbol{d}^* = (1+\frac{1}{n}\boldsymbol{d}'R\boldsymbol{d})^{-\frac{1}{2}}\boldsymbol{d}$ is one to one and onto because $\boldsymbol{d}^* = (1+\frac{1}{n}\boldsymbol{d}'R\boldsymbol{d})^{-\frac{1}{2}}\boldsymbol{d}$ if and only if $\boldsymbol{d} = (1+\frac{1}{n}\boldsymbol{d}^{*\prime}R\boldsymbol{d}^*)^{-\frac{1}{2}}\boldsymbol{d}^*$. Also from Lemma 2.5 the set $\mathcal{D}^*$ is the set of all normalized supporting hyperplanes to $\mathcal{V}$. Note also that the symmetry of $\mathcal{G}$ and $\overline{\mathcal{D}}$ implies that both $\overline{\mathcal{V}}$ and $\overline{\mathcal{D}}^*$ are symmetric. Thus as in the case of $\mathcal{G}$ and $\overline{\mathcal{D}}$, the set $\overline{\mathcal{D}}^*$ is the convex dual of $\overline{\mathcal{V}}$, $\mathcal{D}^*$ is the boundary of $\overline{\mathcal{D}}^*$ and $\mathcal{V}$ is the boundary of $\overline{\mathcal{V}}$. Thus for any nonzero $k \times 1$ vector $\boldsymbol{c}$, there exists a unique positive constant α_0 and a unique positive constant δ_0 such that $\alpha_0 \boldsymbol{c} \epsilon \mathcal{D}^*$ and $\delta_0 \boldsymbol{c} \epsilon \mathcal{V}$.

Thus it follows from (2.10) of Lemma 2.3 that every non–zero $k \times 1$ vector $\boldsymbol{c}$ has the representation

$$(1+\tfrac{1}{n}\boldsymbol{d}_0'R\boldsymbol{d}_0)^{-\frac{1}{2}}\left(\sum_{i=1}^{m}\epsilon_i p_i \boldsymbol{f}(x_i) + \tfrac{1}{n}R\boldsymbol{d}_0\right) = \delta_0 \boldsymbol{c} \in \mathcal{V} \tag{2.16}$$

for some $\boldsymbol{d}_0 \in \mathcal{D}$ and $\boldsymbol{u}_0 = \sum_{i=1}^{m}\epsilon_i p_i \boldsymbol{f}(x_i) \in \mathcal{C}(\boldsymbol{d}_0)$ for some positive integer m, where $p_i > 0$, $\epsilon_i \boldsymbol{d}_0'\boldsymbol{f}(x_i) = 1$, $\sum_{i=1}^{m} p_i = 1$ and the unique positive constant δ_0 is given by

$$\delta_0 = \beta_0 \left(1+\tfrac{1}{n}\boldsymbol{d}_0'R\boldsymbol{d}_0\right)^{\frac{1}{2}} \tag{2.17}$$

As in the proof of Lemma 2.5, one can easily show that the set $\overline{\mathcal{H}}$ is a symmetric compact subset of R^k, however unlike $\overline{\mathcal{V}}$, the set $\overline{\mathcal{H}}$ (as mentioned earlier) is not convex in general. Example 3.1 illustrates this point. Nevertheless, the set $\overline{\mathcal{H}}$ satisfies other properties of $\mathcal{G}$ and $\overline{\mathcal{V}}$ as the following lemma indicates.

Lemma 2.6

Any half line through the origin intersects $\mathcal{H}$ at exactly one point and the representation (2.10) is unique.

Proof: Assume to the contrary that there exists $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \partial\mathcal{G}$, $\boldsymbol{d}_1 \in \mathcal{D}_{\boldsymbol{u}_1}$, $\boldsymbol{d}_2 \in \mathcal{D}_{\boldsymbol{u}_2}$ and $\gamma \geq 1$ such that $\boldsymbol{z} = \boldsymbol{u}_1 + \frac{1}{n}R\boldsymbol{d}_1$ and $\gamma\boldsymbol{z} = \boldsymbol{u}_2 + \frac{1}{n}R\boldsymbol{d}_2$ are elements of $\mathcal{H}$. Let $\beta_1 = (\boldsymbol{d}_1'\boldsymbol{z})^{-1}$ and $\beta_2 = (\boldsymbol{d}_2'\boldsymbol{z})^{-1}$. Then it follows from Lemma 2.5 and Remark 2.1 that there exists

$\delta > 0$ such that $\delta \boldsymbol{z} \in \mathcal{V}$ and $\delta = \sqrt{\beta_1} = \sqrt{\beta_2 \gamma}$. Premultiplication of $\boldsymbol{z}$ and $\gamma \boldsymbol{z}$ by $\boldsymbol{d}_1'$ and $\boldsymbol{d}_2'$ and using the fact that $\beta_1 = \gamma \beta_2$ we get

$$\beta_1(1 + \tfrac{1}{n}\boldsymbol{d}_1' R \boldsymbol{d}_1) = 1 \tag{2.18}$$

$$\beta_1(1 + \tfrac{1}{n}\boldsymbol{d}_2' R \boldsymbol{d}_2) = \gamma^2 \tag{2.19}$$

$$\beta_1(\boldsymbol{d}_2' \boldsymbol{u}_1 + \tfrac{1}{n}\boldsymbol{d}_1' R \boldsymbol{d}_2) = \gamma \tag{2.20}$$

$$\beta_1(\boldsymbol{d}_1' \boldsymbol{u}_2 + \tfrac{1}{n}\boldsymbol{d}_1' R \boldsymbol{d}_2) = \gamma \tag{2.21}$$

From (2.20) and (2.21), we have $\boldsymbol{d}_1' \boldsymbol{u}_2 = \boldsymbol{d}_2' \boldsymbol{u}_1$ and combining the above four equations we get

$$\tfrac{1}{n}(\gamma \boldsymbol{d}_1 - \boldsymbol{d}_2)' R (\gamma \boldsymbol{d}_1 - \boldsymbol{d}_2) = \tfrac{\gamma^2}{\beta_1} - \gamma^2 + \tfrac{\gamma^2}{\beta_1} + 2\gamma \boldsymbol{d}_2' \boldsymbol{u}_1 \leq -(\gamma - 1)^2 \leq 0 \tag{2.22}$$

with equality holding if and only if $\gamma = 1$ and $\boldsymbol{d}_1' \boldsymbol{u}_2 = \boldsymbol{d}_2' \boldsymbol{u}_1 = 1$. Since R is positive definite, then

$$\tfrac{1}{n}(\gamma \boldsymbol{d}_1 - \boldsymbol{d}_2)' R (\gamma \boldsymbol{d}_1 - \boldsymbol{d}_2) \geq 0 \tag{2.23}$$

with equality holding if and only if $\gamma \boldsymbol{d}_1 = \boldsymbol{d}_2$. Thus from (2.22) and (2.23) we get a contradiction unless $\boldsymbol{d}_1 = \boldsymbol{d}_2 = \boldsymbol{d}$ (say) and $\gamma = 1$. But if this is the case, then $\boldsymbol{u}_1 + \frac{1}{n} R \boldsymbol{d} = \boldsymbol{u}_2 + \frac{1}{n} R \boldsymbol{d}$ which implies that $\boldsymbol{u}_1 = \boldsymbol{u}_2$. This completes the proof. ■

Remark 2.2

What Lemma 2.6 says is that for any nonzero $k \times 1$ vector $\boldsymbol{c}$, there exist a unique triple $(\boldsymbol{u}_0, \boldsymbol{d}_0, \gamma_0)$, $\boldsymbol{u}_0 \in \partial \mathcal{G}$, $\boldsymbol{d}_0 \in \mathcal{D}_{\boldsymbol{u}_0}$ and $\gamma_0 > 0$ for which $\boldsymbol{u}_0 + \frac{1}{n} R \boldsymbol{d}_0 = \gamma_0 \boldsymbol{c} \in \mathcal{H}$. Thus the triple $(\boldsymbol{u}_0, \boldsymbol{d}_0, \boldsymbol{\delta}_0)$ in the representation (2.16) is also unique. That $\overline{\mathcal{H}}$ is a symmetric compact set which spans $\mathbb{R}^k$ follows from this and the symmetry and compactness of $\overline{\mathcal{V}}$. Thus $\overline{\mathcal{H}}$ is a "starlike" set in $\mathbb{R}^k$ with boundary $\mathcal{H}$.

3. Elfving's Theorem, Geometry and Duality Theory

The following result is the Bayesian version of Elfving's Theorem (mentioned in the introduction) for the $\boldsymbol{c}$–optimality criterion.

Theorem 3.1

Given a nonzero $k \times 1$ vector $\boldsymbol{c}$ and a $k \times k$ positive definite matrix R, the design ξ_0 is Bayesian $\boldsymbol{c}$–optimum if and only if $\boldsymbol{c}$ has the representation (2.10), or equivalently (2.16), with $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$. Bayesian $\boldsymbol{c}$–optimal designs always exist and $\inf_{\xi \in \Xi} \boldsymbol{c}' M_b^{-1}(\xi)\boldsymbol{c} = \rho(\boldsymbol{c}) = \delta_0^{-2} = (\beta_0 \gamma_0)^{-1}$.

Proof: Let $\xi = \{x_i, p_i\}_{i=1}^{m}$ be any design in Ξ and let $\boldsymbol{d}^*$ be any vector in $\mathcal{D}^*$. Then one has

$$\boldsymbol{d}^{*\prime} M_b(\xi)\boldsymbol{d}^* = \frac{\boldsymbol{d}' M_b(\xi)\boldsymbol{d}}{1 + \frac{1}{n}\boldsymbol{d}' R\boldsymbol{d}} = \frac{1}{1 + \frac{1}{n}\boldsymbol{d}' R\boldsymbol{d}} \left(\sum_{i=1}^{m} p_i (\boldsymbol{d}' \boldsymbol{f}(x_i))^2 + \tfrac{1}{n}\boldsymbol{d}' R\boldsymbol{d} \right) \tag{3.1}$$

Since $|\boldsymbol{d}' \boldsymbol{f}(x)| \le 1$ for all $x \in \mathcal{X}$, then $|\boldsymbol{d}' \boldsymbol{f}(x_i)|^2 \le 1$, $i = 1, 2, \ldots, m$ and so $\sum_{i=1}^{m} p_i (\boldsymbol{d}' \boldsymbol{f}(x_i))^2 \le \sum_{i=1}^{m} p_i = 1$. Thus it follows from this and (3.1) that

$$\boldsymbol{d}^{*\prime} M_b(\xi)\boldsymbol{d}^* \le 1 \text{ for all } \xi \in \Xi \text{ and all } \boldsymbol{d}^* \in \mathcal{D}^* \tag{3.2}$$

with equality holding if and only if $|\boldsymbol{d}' \boldsymbol{f}(x)| = 1$ for all x in the spectrum of the design ξ. In other words, equality holds in (3.2) if and only if $\boldsymbol{u} = \int_{\mathcal{X}} \epsilon(x) \boldsymbol{f}(x) \xi(dx) \in \partial \mathcal{G}$ for some function ϵ being defined on the support of the design ξ which takes values ± 1 and $\boldsymbol{d}$ is the supporting hyperplane to $\mathcal{G}$ at the boundary point $\boldsymbol{u}$. Let $\boldsymbol{c}$ be any nonzero $k \times 1$ vector. Then it follows from Schwarz's inequality that

$$\boldsymbol{c}' M_b^{-1}(\xi)\boldsymbol{c} = \sup_{\boldsymbol{d} \in \mathcal{D}} \frac{(\boldsymbol{d}'\boldsymbol{c})^2}{\boldsymbol{d}' M_b(\xi)\boldsymbol{d}} = \sup_{\boldsymbol{d}^* \in \mathcal{D}^*} \frac{(\boldsymbol{d}^{*\prime}\boldsymbol{c})^2}{\boldsymbol{d}^{*\prime} M_b(\xi)\boldsymbol{d}^*} \ge \frac{(\boldsymbol{d}^{*\prime}\boldsymbol{c})^2}{\boldsymbol{d}^{*\prime} M_b(\xi)\boldsymbol{d}^*} \tag{3.3}$$

which implies that

$$\boldsymbol{d}^{*\prime} M_b(\xi)\boldsymbol{d}^* \ge (\boldsymbol{c}' M_b^{-1}(\xi)\boldsymbol{c})^{-1} (\boldsymbol{d}^{*\prime}\boldsymbol{c})^2 \text{ for all } \xi \in \Xi \text{ and all } \boldsymbol{d}^* \in \mathcal{D}^* \tag{3.4}$$

From (3.2) and (3.4) it follows that

$$\inf_{\xi \in \Xi} \boldsymbol{c}' M_b^{-1}(\xi)\boldsymbol{c} \ge \sup_{\boldsymbol{d}^* \in \mathcal{D}^*} (\boldsymbol{d}^{*\prime}\boldsymbol{c})^2 \tag{3.5}$$

Since $\boldsymbol{c}$ has a unique representation

$$\boldsymbol{u}_0 + \tfrac{1}{n} R\boldsymbol{d}_0 = \gamma_0 \boldsymbol{c} \tag{3.6}$$

with $\boldsymbol{d}_0 \in \mathcal{D}$, $\boldsymbol{u}_0 \in \mathcal{C}(\boldsymbol{d}_0)$ and $\gamma_0 > 0$ is such that $\gamma_0 \boldsymbol{c} \in \mathcal{H}$, then it follows from Lemma 2.1 that $\boldsymbol{u}_0 = \sum_{i=1}^{m} \epsilon_i p_i \boldsymbol{f}(x_i)$ for some positive integer m, $p_i > 0$, $\epsilon_i = \boldsymbol{d}_0' \boldsymbol{f}(x_i) \in \{\pm 1\}$ and $\sum_{i=1}^{m} p_i = 1$. Thus it follows from (3.6) that

$$(\boldsymbol{d}_0^{*\prime}\boldsymbol{c})^2 = \frac{(\boldsymbol{d}_0'\boldsymbol{c})^2}{1 + \frac{1}{n}\boldsymbol{d}'R\boldsymbol{d}} = \frac{1}{\beta_0^2(1 + \frac{1}{n}\boldsymbol{d}_0'R\boldsymbol{d}_0)} = \frac{1}{\beta_0\gamma_0} = \frac{1}{\delta_0^2} \tag{3.7}$$

and

$$\boldsymbol{c}'M_b^{-1}(\xi_0)\boldsymbol{c} = \frac{1}{\gamma_0}\boldsymbol{c}'\boldsymbol{d}_0 = \frac{1}{\beta_0\gamma_0} = \frac{1}{\delta_0^2} \tag{3.8}$$

if and only if $\xi_0 = \{x_i, p_i\}_{i=1}^{m}$. Thus it follows from (3.7) and (3.8) that

$$\boldsymbol{c}'M_b^{-1}(\xi_0)\boldsymbol{c} = (\boldsymbol{d}_0^{*\prime}\boldsymbol{c})^2 = \frac{1}{\beta_0\gamma_0} = \frac{1}{\delta_0^2} \tag{3.9}$$

if and only if $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$ and so it follows from this and (3.5) that

$$\inf_{\xi \in \Xi} \boldsymbol{c}'M_b^{-1}(\xi)\boldsymbol{c} = \boldsymbol{c}'M_b^{-1}(\xi_0)\boldsymbol{c} = (\boldsymbol{d}_0^{*\prime}\boldsymbol{c})^2 = \sup_{\boldsymbol{d}^* \in \mathcal{D}^*} (\boldsymbol{d}^{*\prime}\boldsymbol{c})^2 = \frac{1}{\beta_0\gamma_0} = \frac{1}{\delta_0^2} \tag{3.10}$$

if and only if $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$. Thus ξ_0 is a Bayesian $\boldsymbol{c}$–optimal design if and only if $\boldsymbol{c}$ has the representation (2.10) with $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$, and $\inf_{\xi \in \Xi} \boldsymbol{c}'M_b^{-1}(\xi)\boldsymbol{c} = \rho(\boldsymbol{c}) = (\beta_0\gamma_0)^{-1} = \delta_0^{-2}$ which completes the proof. ■

The following result follows directly from (3.10).

Corollary 3.1

The Bayesian $\boldsymbol{c}$–optimal design problem

$$\text{Minimize } \boldsymbol{c}'M_b^{-1}(\xi)\boldsymbol{c} \text{ subject to } \xi \in \Xi \tag{3.11}$$

is the dual of the problem

$$\text{Maximize } (\boldsymbol{d}^{*\prime}\boldsymbol{c})^2 \text{ subject to } \boldsymbol{d}^* \in \mathcal{D}^* \tag{3.12}$$

and the two problem share a common extreme value $\rho(\boldsymbol{c}) = \delta_0^{-2} = (\beta_0\gamma_0)^{-1}$. ■

The parallelism between Theorem 3.1 and its classical analog of Elfving's Theorem is now evident. To see that, let ϵ denote the set of all functions ϵ defined on $\mathcal{X}$ which takes values ± 1. In Elfving's Theorem we find the unique positive constant β^* such that $\beta^* \boldsymbol{c} \in \partial\mathcal{G}$ and the classical $\boldsymbol{c}$–optimal design is then the design ξ^* for which

$$\int_{\mathcal{X}} \epsilon^*(x) f(x) \xi^*(dx) = \beta^* \boldsymbol{c} \text{ for some } \epsilon^* \in \mathcal{E} \tag{3.13}$$

in which case the infimum of $\boldsymbol{c}' M^-(\xi)\boldsymbol{c}$ among all designs ξ for which $\boldsymbol{c}'\boldsymbol{\theta}$ is estimable is equal to β^{*-2} and is attained at $\xi = \xi^*$. In Theorem 3.1 we find the unique positive constant δ_0 such that $\delta_0 \boldsymbol{c} \in \mathcal{V}$ (or equivalently the unique positive constant γ_0 such that $\gamma_0 \boldsymbol{c} \in \mathcal{H}$) and the Bayesian $\boldsymbol{c}$–optimal design is the design ξ_0 for which

$$\int_{\mathcal{X}} \epsilon(x) \boldsymbol{f}(x) \xi_0(dx) = \boldsymbol{u}_0 \text{ for some } \epsilon \in \mathcal{E} \tag{3.14}$$

where $\boldsymbol{u}_0 \epsilon \partial\mathcal{G}$ is uniquely determined by the representation $\boldsymbol{u}_0 + \frac{1}{n} R\boldsymbol{d}_0 = \gamma_0 \boldsymbol{c}$ (or equivalently the representation $\frac{\boldsymbol{u}_0 + \frac{1}{n} R\boldsymbol{d}_0}{\sqrt{1 + \frac{1}{n}\boldsymbol{d}_0' R\boldsymbol{d}_0}} = \delta_0 \boldsymbol{c}$) of $\boldsymbol{c}$ and $\boldsymbol{d}_0$ is the supporting hyperplane to $\mathcal{G}$ at $\boldsymbol{u}_0 \in \partial\mathcal{G}$ normalized so that $\boldsymbol{d}_0' \boldsymbol{u}_0 = 1$. The infimum of $\boldsymbol{c}' M_b^{-1}(\xi)\boldsymbol{c}$ among all designs ξ is equal to δ_0^{-2} and is attained at $\xi = \xi_0$.

Following similar steps to those by which Corollary 3.1 is derived, one can easily see that the classical $\boldsymbol{c}$–optimal design problem is the dual of the problem

$$\text{Maximize } (\boldsymbol{d}'\boldsymbol{c})^2 \text{ subject to } \boldsymbol{d} \in \mathcal{D} \tag{3.15}$$

and that the two problems share a common extreme value. Thus finding the classical and the Bayesian $\boldsymbol{c}$–optimal designs can be achieved geometrically by visualizing the sets $\partial\mathcal{G}$ and $\mathcal{V}$ or equivalently the sets $\partial\mathcal{G}$ and $\mathcal{H}$. Also the design problem and its dual problem are clearly equivalent in the sense that by solving any one of them one can, with the aid of the unique representation (3.6), obtain a solution of the other and so in addition to the intuitive appeal of the above geometrical approach, it may be possible to solve certain covering or dual problems both in theory and in practice. For more discussion of this and the approximation theory interpretation of the above results, see El–Krunz (1989).

We now prove the following simple result which gives the condition under which the Bayesian and the classical $\boldsymbol{c}$–optimal designs coincide. Further applications are given in later sections.

Corollary 3.2

Let $\xi_0 = \{x_i, p_i\}_{i=1}^{m}$ be a classical $\boldsymbol{c}$–optimal design and let β_0 be such that $\beta_0 \boldsymbol{c} \in \partial\mathcal{G}$. Then ξ_0 is a Bayesian $\boldsymbol{c}$–optimal design if and only if

$$R\boldsymbol{d}_0 = \alpha_0 \boldsymbol{c} \text{ for some } \alpha_0 > 0 \text{ and some } \boldsymbol{d}_0 \in \mathcal{D}_{\beta_0 \boldsymbol{c}} \tag{3.16}$$

Proof: Since ξ_0 is a classical $\boldsymbol{c}$–optimal design, then it follows from Elfving's Theorem that there exists $\epsilon_i \in \{\pm 1\}$, $i = 1, 2, \ldots, m$ such that $\sum_{i=1}^{m} \epsilon_i p_i f(x_i) = \beta_0 \boldsymbol{c}$. If ξ_0 is also a Bayesian $\boldsymbol{c}$–optimal design, then it follows from Theorem 3.1 that $\beta_0 \boldsymbol{c} + \frac{1}{n} R\boldsymbol{d}_0 = \gamma_0 \boldsymbol{c}$, where $\boldsymbol{d}_0 \in \mathcal{D}_{\beta_0 \boldsymbol{c}}$ and $\gamma_0 = \beta_0(1 + \frac{1}{n}\boldsymbol{d}' R\boldsymbol{d})$ which implies that $R\boldsymbol{d}_0 = \alpha_0 \boldsymbol{c}$, where $\alpha_0 = n(\gamma_0 - \beta_0) = \beta_0 \boldsymbol{d}_0' R\boldsymbol{d}_0 > 0$. On the other hand if $R\boldsymbol{d}_0 = \alpha_0 \boldsymbol{c}$ for some $\alpha_0 > 0$ and some $\boldsymbol{d}_0 \in \mathcal{D}_{\beta_0 \boldsymbol{c}}$, then $\beta_0 \boldsymbol{c} + \frac{1}{n} R\boldsymbol{d}_0 = (\beta_0 + \frac{\alpha_0}{n})\boldsymbol{c} = \alpha_0 \boldsymbol{c}$ which implies that $\sum_{i=1}^{m} \epsilon_i p_i \boldsymbol{f}(x_i) + \frac{1}{n} R\boldsymbol{d}_0 = \gamma_0 \boldsymbol{c}$ and so ξ_0 is a Bayesian $\boldsymbol{c}$–optimal design. This completes the proof.

Example 3.1 In this example we consider the simple linear regression to illustrate the sets $\mathcal{G}, \mathcal{V}$ and $\mathcal{H}$ and also Corollary 3.2. Let $\boldsymbol{f}(x) = (1, x)'$, $x \epsilon [-1, 1]$, $R = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 2/3 \end{pmatrix}$ and take $n = 1$. The sets $\mathcal{G}, \mathcal{V}$ and $\mathcal{H}$ are depicted in Figure 3.1. The set $\mathcal{G}$ is readily seen to be the square with side of length 2. To draw the set $\mathcal{H}$ we simply take each point $\boldsymbol{u} \epsilon \partial\mathcal{G}$ and transform to the point $\boldsymbol{h} = \boldsymbol{u} + R\boldsymbol{d}$ where $\boldsymbol{d}$ supports $\mathcal{G}$ at $\boldsymbol{u}$. Note that the representation $\boldsymbol{h} = \boldsymbol{u} + R\boldsymbol{d}$ is linear in both $\boldsymbol{u}$ and $\boldsymbol{d}$; however they depend on one another. Thus the four sides of the square transform to line segments since each corresponds to the same $\boldsymbol{d}$. The right vertical face has $\boldsymbol{d} = (1, 0)$ so it transforms to $\boldsymbol{u} + R\boldsymbol{d} = \boldsymbol{u} + (1, 1/2)'$. In particular (1,1) corresponds to (2, 3/2) as indicated on the figure. Note that $\mathcal{G} \subset \overline{\mathcal{V}} \subset \overline{\mathcal{H}}$; that $\mathcal{G}$ and $\overline{\mathcal{V}}$ are convex while $\overline{\mathcal{H}}$ is not. The set $\mathcal{V}$ is just $\mathcal{H}$ pulled toward the origin by the factor $(1 + \frac{1}{n}\boldsymbol{d}' R\boldsymbol{d})^{-\frac{1}{2}}$. In most of our examples we have found the set $\mathcal{H}$ to be easier to work with than $\mathcal{V}$; in fact in most cases we do not even consider $\mathcal{V}$.

In extrapolating, say to $x_0 > 1$, we take $\boldsymbol{c} = (1, x_0)'$. One can readily check that the classical design puts weight α and $1 - \alpha$ at -1 and 1 where $\alpha = 2^{-1}(1 - x_0^{-1})$ and the minimum variance is x_0^2. In the Bayesian case we use the same two points with $\alpha = \frac{2}{3}(\frac{9}{8} - x_0^{-1})$ and the Bayes risk is $(\frac{3}{4})^2 x_0^2$.

This same example illustrates corollary 3.2. Thus if in the extrapolation case we take R so that $R\binom{0}{1} = \alpha_0\binom{1}{x_0}$ the design stays exactly the same. This is the case if $R = k\begin{pmatrix}\rho & 1\\ 1 & x_0\end{pmatrix}$ where $k > 0,\ \rho > 0$ and $\rho x_0 > 1$.

Example 3.2

Assume that $\boldsymbol{f}(x) = \left(\frac{1}{\sqrt{1+x^2}}, \frac{x}{\sqrt{1+x^2}}\right)'$, $x \in \mathcal{X} = [-1,1]$. This model actually arises from the simpler standard linear regression model $(1,x)$ except that the variances are not assumed constant. More details can be found in DasGupta and Studden (1989). One can easily see that $\boldsymbol{f}(\mathcal{X}) = \left\{(a, \pm\sqrt{1-a^2})' : a \in \left[\frac{1}{\sqrt{2}}, 1\right]\right\}$ and that the boundary of $\mathcal{G}$ consists of $\boldsymbol{f}(\mathcal{X})$, $-\boldsymbol{f}(\mathcal{X})$, the line segment joining the two points $\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ and the line segment joining the two points $\left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$, $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$. Note also that $\boldsymbol{f}(\mathcal{X}) \cup (-\boldsymbol{f}(\mathcal{X}))$ is the part of the circle $a^2 + b^2 = 1$ for which $|b| \leq a$. For any nonzero $k \times 1$ vector $\boldsymbol{c}$, we want to characterize the entire class of Bayesian $\boldsymbol{c}$–optimal designs. So for each $x \in (-1,1)$, let $\boldsymbol{u} = (u_1, u_2)' = \boldsymbol{f}(x)$. Then $\frac{1}{\sqrt{2}} < u_1 < 1$ and the supporting hyperplane to $\mathcal{G}$ at $\boldsymbol{u}$ is $\boldsymbol{u}$ itself. Thus it follows from Theorem 3.1 that the one point design ξ_x is a Bayesian $\boldsymbol{c}$–optimal design if and only if

$$\boldsymbol{u} = \gamma_0(I + \tfrac{1}{n}R)^{-1}\boldsymbol{c} \tag{3.17}$$

where γ_0 is chosen such that $\gamma_0(I + \frac{1}{n}R)^{-1}\boldsymbol{c} \in \boldsymbol{f}(\mathcal{X})$, that is, $\gamma_0 = \pm\|(I + \frac{1}{n}R)^{-1}\boldsymbol{c}\|$. Let $R = ((r_{ij}))_{i,j=1}^{2}$ and let $z_1 = (1 + \frac{1}{n}r_{22})c_1 - \frac{1}{n}r_{12}c_2$ and $z_2 = -\frac{1}{n}r_{12}c_1 + (1 + \frac{1}{n}r_{11})c_2$. Then it follows from (3.17) that the one point design ξ_x, $x \in (-1,1)$ is a Bayesian $\boldsymbol{c}$–optimal design if and only if $|z_1| > |z_2|$ and $x = \frac{z_2}{z_1}$.

It also follows from Theorem 3.1 that the Bayesian $\boldsymbol{c}$–optimal design puts weights $1-p,\ p,\ 0 < p < 1$ at the two points $-1, 1$ respectively if and only if

$$\gamma_0\boldsymbol{c} = p\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} + (1-p)\begin{pmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix} + \tfrac{1}{n}R\boldsymbol{d}_0 \tag{3.18}$$

where $\boldsymbol{d}_0 = (0, \sqrt{2})'$, $\gamma_0 = \beta_0(1 + \frac{1}{n}\boldsymbol{d}_0'R\boldsymbol{d}_0)$ and $\beta_0 = (\boldsymbol{d}_0'\boldsymbol{c})^{-1}$. Since $\boldsymbol{d}_0'\boldsymbol{c} = \sqrt{2}\,c_2$, where $\boldsymbol{c} = (c_1, c_2)'$, then for the Bayesian $\boldsymbol{c}$–optimal design to be supported at the two points $-1,\ 1$, we must have $c_2 \neq 0$. Thus it follows from (3.18) that the Bayesian $\boldsymbol{c}$–optimal design puts weights $1-p,\ p,\ 0 < p < 1$ at the two points

$-1, 1$ respectively if and only if

$$-\tfrac{n}{2}\left(1+\tfrac{c_1}{c_2}\right) < \left(\tfrac{c_1}{c_2}r_{22} - r_{12}\right) < \tfrac{n}{2}\left(1-\tfrac{c_1}{c_2}\right) \tag{3.19}$$

in which case

$$p = \tfrac{1}{n}\left(\tfrac{c_1}{c_2}r_{22} - r_{12}\right) + \tfrac{1}{2}\left(\tfrac{c_1}{c_2}+1\right) \tag{3.20}$$

Finally, using the same theorem, one can easily demonstrate that the design ξ_1 which puts all of its weight at the point $x = 1$ is a Bayesian $\boldsymbol{c}$–optimal design if and only if $0 \leq z_1 \leq z_2$ or $z_2 \leq z_1 \leq 0$, and, $\left(\tfrac{c_1}{c_2}r_{22} - r_{12}\right) \geq \tfrac{n}{2}\left(1-\tfrac{c_1}{c_2}\right)$ or $c_2 = 0$; the design ξ_{-1} which puts all of its weight at the point $x = -1$ is a Bayesian $\boldsymbol{c}$–optimal design if and only if $0 \leq z_1 \leq -z_2$ or $-z_2 \leq z_1 \leq 0$, and, $\left(\tfrac{c_1}{c_2}r_{22} - r_{12}\right) \leq -\tfrac{n}{2}\left(1+\tfrac{c_1}{c_2}\right)$ or $c_2 = 0$.

4. One Point Designs and Alternative Formulations

In the classical theory of optimal design, a one point design ξ_{x_0} cannot be a classical $\boldsymbol{c}$–optimal design unless $\beta^*\boldsymbol{c} = \pm\boldsymbol{f}(x_0) \in \partial\mathcal{G}$ for some $x_0 \in \mathcal{X}$ and some constant $\beta^* \neq 0$. For the Bayesian theory of optimal design it follows from the unique representation (3.6) of $\boldsymbol{c}$ that the one point design ξ_{x_0} is a Bayesian $\boldsymbol{c}$–optimal design if and only if $\boldsymbol{u}_0 = \pm\boldsymbol{f}(x_0)$ for some $x_0 \in \mathcal{X}$. This will always be the case if $\partial\mathcal{G} \subseteq \boldsymbol{f}(\mathcal{X}) \cup (-\boldsymbol{f}(\mathcal{X}))$. The following result is given in Chaloner (1984) and Pilz (1983).

Lemma 4.1

If the design space $\mathcal{X}$ and $\boldsymbol{f}$ are such that $\partial\mathcal{G} \subseteq \boldsymbol{f}(\mathcal{X}) \cup (-\boldsymbol{f}(\mathcal{X}))$ then every Bayesian $\boldsymbol{c}$–optimum design is a one point design for some $x_0 \in \mathcal{X}$. ■

The above Lemma states a simple sufficient condition for the existence of Bayesian $\boldsymbol{c}$–optimal one point designs. A necessary and sufficient condition however is that the point $\boldsymbol{u}_0 \in \partial\mathcal{G}$ in the unique representation (3.6) of $\boldsymbol{c}$ be an element of $\pm\boldsymbol{f}(\mathcal{X})$ and so if the prior precision matrix R has some convenient structure, one point designs can often be Bayesian $\boldsymbol{c}$–optimum. One point designs are of special interest because they are exact designs which are easy to implement and which keep the experimental effort minimal. What we are interested in here is to characterize the set of precision matrices R for which the one point design ξ_{x_0} is Bayesian $\boldsymbol{c}$–optimal for a given nonzero $k \times 1$ vector $\boldsymbol{c}$ and a point $x_0 \in \mathcal{X}$ for which $\boldsymbol{f}(x_0) \in \partial\mathcal{G}$. Consideration of this problem led to an alternate formulation of the design problem given in Theorem 4.1. The following lemma will be needed. For a proof, see El–Krunz (1989).

Lemma 4.2

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be given $k \times 1$ nonzero vectors and let R be an unknown $k \times k$ positive definite matrix. Then a positive definite solution in R to the matrix equation $R\boldsymbol{x} = \boldsymbol{y}$ exists if and only if $\boldsymbol{x}'\boldsymbol{y} > 0$. The general solution is

$$R = \frac{\boldsymbol{y}\boldsymbol{y}'}{\boldsymbol{x}'\boldsymbol{y}} + U\Lambda U' \tag{4.1}$$

where $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$ is an arbitrary orthogonal matrix for which $\boldsymbol{u}_1 = \boldsymbol{x}/|\boldsymbol{x}|$ and $\Lambda = \text{diag}\,(0, \lambda_2, \ldots, \lambda_k)$, where $\lambda_2, \ldots, \lambda_k$ are arbitrary positive real numbers. ■

Let $\boldsymbol{c}$ be any nonzero $k \times 1$ vector and let $\boldsymbol{u}_0 \in \partial\mathcal{G}$ with corresponding design ξ_0. For each $\boldsymbol{d}_0 \in \mathcal{D}_{u_0}$, let $\beta_0 = (\boldsymbol{d}_0'\boldsymbol{c})^{-1}$. Without loss of generality, let us assume that $\beta_0 > 0$. Let us define $\mathcal{R}$ to be the set of precision matrices R for which the design ξ_0 is Bayesian $\boldsymbol{c}$–optimal, that is, $\mathcal{R}$ is the set of all positive definite matrices R for which

$$\boldsymbol{u}_0 + \tfrac{1}{n}R\boldsymbol{d}_0 = \gamma_0\boldsymbol{c} \text{ for some } \boldsymbol{d}_0 \in \mathcal{D}_{u_0} \text{ and some } \gamma_0 > \beta_0 \tag{4.2}$$

For every $\boldsymbol{d}_0 \in \mathcal{D}_{u_0}$, let $\mathcal{R}_{d_0}$ denote the set of all positive definite matrices R for which

$$\boldsymbol{u}_0 + \tfrac{1}{n}R\boldsymbol{d}_0 = \gamma_0\boldsymbol{c} \text{ for some } \gamma_0 > \beta_0 \tag{4.3}$$

or equivalently

$$\tfrac{1}{n}R\boldsymbol{d}_0 = (\gamma_0\boldsymbol{c} - \boldsymbol{u}_0) \text{ for some } \gamma_0 > \beta_0 \tag{4.4}$$

Let $\mathcal{N}_{d_0}$ be the set of all matrices $U\Lambda U'$, where $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$ is an arbitrary orthogonal matrix for which $u_1 = \frac{d_0}{\sqrt{d_0'd_0}}$ and $\Lambda = \text{diag}\,(0, \lambda_2, \ldots, \lambda_k)$, where $\lambda_2, \ldots, \lambda_k$ are arbitrary positive real numbers. Since $(\gamma_0\boldsymbol{c} - \boldsymbol{u}_0)'\boldsymbol{d}_0 = \frac{\gamma_0}{\beta_0} - 1 > 0$, then it follows from Lemma 4.2 that

$$\mathcal{R}_{d_0} = \left\{ R\colon R = \frac{n(\gamma_0\boldsymbol{c} - \boldsymbol{u}_0)(\gamma_0\boldsymbol{c} - \boldsymbol{u}_0)'}{\left(\frac{\gamma_0}{\beta_0} - 1\right)} + U\Lambda U',\ \gamma_0 > \beta_0 \text{ and } U\Lambda U' \in \mathcal{N}_{d_0} \right\} \tag{4.5}$$

and $\mathcal{R} = \bigcup_{d_0 \in \mathcal{D}_{u_0}} \mathcal{R}_{d_0}$. Thus, we have the following result.

Theorem 4.1

Let $\boldsymbol{c}$ be any nonzero $k \times 1$ vector and $\boldsymbol{u}_0 \in \partial\mathcal{G}$ have corresponding design ξ_0. The design ξ_0 in Bayesian $\boldsymbol{c}$–optimal if and only if $R \in \mathcal{R}$. ■

Special Case

Assume that $\boldsymbol{c} = \boldsymbol{u}_0 = \boldsymbol{f}(x_0)$, that is, we are interested in the estimation of $\boldsymbol{f}'(x_0)\boldsymbol{\theta}$ for some $x_0 \in \mathcal{X}$ and that $\boldsymbol{f}(x_0) \in \partial\mathcal{G}$. Then $\beta_0 = 1$ and (4.5) becomes

$$\begin{aligned}\mathcal{R}_{d_0} &= \{R\colon\ R = n(\gamma_0 - 1)\boldsymbol{c}\boldsymbol{c}' + U\Lambda U',\ \gamma_0 > 1 \text{ and } U\Lambda U' \in \mathcal{N}_{d_0}\} \\ &= \{R\colon\ R = \alpha_0 \boldsymbol{c}\boldsymbol{c}' + U\Lambda U',\ \alpha_0 > 0 \text{ and } U\Lambda U' \in \mathcal{N}_{d_0}\} \qquad (4.6)\end{aligned}$$

which is independent of n and so we have the following result.

Corollary 4.1

If $\boldsymbol{c} = \boldsymbol{u}_0 = \boldsymbol{f}(x_0) \in \partial\mathcal{G}$ for some $x_0 \in \mathcal{X}$, then the one point design ξ_{x_0} is a Bayesian $\boldsymbol{c}$–optimal design if and only if $R \in \mathcal{R} = \bigcup_{d_0 \in \mathcal{D}_c} \mathcal{R}_{d_0}$, where $\mathcal{R}_{d_0}$ is given by (4.6). Moreover $\mathcal{R}$ does not depend on n, i.e. $\mathcal{R}$ is a "cone".

Example 4.1

This example is described in Chaloner (1984) and Pilz (1983). Assume that the design space $\mathcal{X}$ is such that $f(\mathcal{X})$ is the unit ball, that is,

$$f(\mathcal{X}) = \partial\mathcal{G} = \{\boldsymbol{u} \in \mathsf{R}^k\colon \boldsymbol{u}'\boldsymbol{u} = 1\}. \qquad (4.7)$$

Then it follows readily that all the Bayesian $\boldsymbol{c}$–optimal designs are one point designs. Since the supporting hyperplane at any point $\boldsymbol{u}_0 \in \partial\mathcal{G}$ is $\boldsymbol{u}_0$ itself, then it follows from Theorem 3.1 that the one point design ξ_{x_0} is Bayesian $\boldsymbol{c}$–optimum if and only if

$$(I + \tfrac{1}{n}R)\boldsymbol{f}(x_0) = \gamma_0 \boldsymbol{c} \qquad (4.8)$$

or equivalently

$$\boldsymbol{f}(x_0) = \gamma_0 (I + \tfrac{1}{n}R)^{-1}\boldsymbol{c} \qquad (4.9)$$

where γ_0 is chosen such that $\gamma_0(I + \frac{1}{n}R)^{-1}\boldsymbol{c} \in \partial\mathcal{G}$, i.e., $\gamma_0^{-1} = \pm||(I + \frac{1}{n}R)^{-1}\boldsymbol{c}||$. If $\boldsymbol{c} = \boldsymbol{f}(x_0) \in \partial\mathcal{G}$ for some $x_0 \in \mathcal{X}$, then it follows from Corollary 4.1 that the one point design ξ_{x_0} is Bayesian $\boldsymbol{c}$–optimum if and only if $R\boldsymbol{c} = \alpha_0\boldsymbol{c}$ for some $\alpha_0 > 0$, i.e., ξ_{x_0} is a Bayesian $\boldsymbol{c}$–optimal design for all prior precision matrices R for which $\boldsymbol{c}$ is an eigenvector.

This suggests a more general simple result. The proof is straightforward.

Lemma 4.3

If the vector $\boldsymbol{c}$ is such that the support plane to $\mathcal{G}$, at $\beta_0\boldsymbol{c} \in \partial\mathcal{G}$, is proportional to $\boldsymbol{c}$ then the classical design ξ_0 is Bayesian $\boldsymbol{c}$–optimal for all prior precision matrices R for which $\boldsymbol{c}$ is an eigenvector. If $R\boldsymbol{c} = \lambda\boldsymbol{c}$ then

$$\inf_{\xi} \boldsymbol{c}' M_b^{-1}(\xi)\boldsymbol{c} = \left(\beta_0^2 + \frac{\lambda}{n||\boldsymbol{c}||^2}\right)^{-1} \tag{4.10}$$

Remark 4.1

Take the set of points of contact of $\mathcal{G}$ with either the sphere inscribed in or circumscribing $\mathcal{G}$. If $\beta_0\boldsymbol{c}$ is any of these points the conditions of Lemma 4.3 hold. If b denotes either radius then (4.10) can be rewritten as $\beta_0^{-2}(1 + \lambda/nb^2)^{-1}$.

Example 4.2

Consider the quadratic polynomial regression model for which $\boldsymbol{f}(x) = (1, x, x^2)'$, $x \in \mathcal{X} = [-1, 1]$. Then the set $\mathcal{G}$ is the convex hull of the parabolic arcs $\pm\boldsymbol{f}(x) = \pm(1, x, x^2)$, $x \in \mathcal{X}$. The "upper face" of $\mathcal{G}$ is the 2–dimensional convex set

$$\mathcal{C} = \left\{\boldsymbol{u}_0\colon\ \boldsymbol{u}_0 = p_1\boldsymbol{f}(-1) - p_2\boldsymbol{f}(0) + p_3\boldsymbol{f}(1),\ p_i \geq 0,\ \sum_{i=1}^{3} p_i = 1\right\} \tag{4.11}$$

and $\boldsymbol{d}_0 = (-1, 0, 2)'$ is the hyperplane supporting $\mathcal{G}$ at the whole face $\mathcal{C}$. Thus the sphere of radius b inscribed in $\mathcal{G}$ touches $\mathcal{G}$ at exactly one point $c \in \mathcal{C}$ and $\boldsymbol{d}_0 = \frac{c}{b^2}$. We shall assume without loss of generality that $\beta_0 = 1$. Since $\boldsymbol{d}_0'\boldsymbol{c} = 1$, then $b^2\boldsymbol{d}_0'\boldsymbol{d}_0 = 1$ which implies that $b = \frac{1}{\sqrt{5}}$ and so $\boldsymbol{c} = \boldsymbol{u}_0 = b^2\boldsymbol{d}_0 = \frac{1}{5}(-1, 0, 2)'$. From (4.11), it follows that $p_1 = p_3 = \frac{1}{5}$ and $p_2 = \frac{3}{5}$ and so it follows from Lemma 4.3 that the design ξ_0 which puts weights $p_1 = \frac{1}{5}$, $p_2 = \frac{3}{5}$ and $p_3 = \frac{1}{5}$ at the three points $-1, 0$ and 1 respectively is

Bayesian c–optimum for all prior precision matrices R for which c is an eigenvector and $c'M_b^{-1}(\xi_0)c = \inf_{\xi \in \Xi} c'M_b^{-1}(\xi)c = \left(1 + \frac{5\lambda}{n}\right)^{-1}$, where λ is the eigenvalue of R corresponding to the eigenvector c. The sphere circumscribing $\mathcal{G}$ touches the boundary of $\mathcal{G}$ at the four points $c_1 = (1,1,1)$, $c_2 = (1,-1,1)$, $c_3 = -c_1 = (-1,-1,-1)$ and $c_4 = -c_2 = (-1,1,-1)$, and has radius $\sqrt{3}$. Thus for $c = \pm c_1$, the design ξ_1 which puts all of its weight at the point $x = 1$ is a Bayesian c–optimal design for all prior precision matrices R for which c is an eigenvector and $\inf_{\xi \in \Xi} c'M_b^{-1}(\xi)c = \left(1 + \frac{\lambda_1}{3n}\right)^{-1}$, where λ_1 is the eigenvalue corresponding to the eigenvector c of R. For $c = \pm c_2$, the design ξ_{-1} which puts all of its weight at the point $x = -1$ is a Bayesian c–optimal design for all prior precision matrices R for which c is an eigenvector and $\inf_{\xi} c'M_b^{-1}(\xi)c = \left(1 + \frac{\lambda_2}{3n}\right)^{-1}$, where λ_2 is the eigenvalue corresponding to the eigenvector c of R.

One can use Corollary 3.2 to characterize the set of all prior precision matrices R for which the Bayesian c–optimal design and the classical c–optimal design coincide. For example, if we are interested in estimating the highest coefficient in this example , i.e., $c'\theta$ for $c = (0,0,1)'$, then the classical c –optimal design ξ^* puts weights $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$ at the points $-1, 0, 1$ respectively and $d_0 = (-1,0,2)'$. Thus it follows from Corollary 3.2 that ξ^* is also a Bayesian c–optimal design if and only if

$$Rd_0 = \alpha_0 c \text{ for some } \alpha_0 > 0 \qquad (4.12)$$

If we let $R = \|r_{ij}\|_{i,j=1}^{3}$, then it follows from (4.12) that $r_{11} = 2r_{13}$, $r_{12} = 2r_{23}$ and $r_{13} < 2r_{33}$. Since R is positive definite, then if $r_{11} = 2r_{13}$ and $r_{12} = 2r_{23}$, then the condition $r_{13} < 2r_{33}$ is trivially satisfied and so the Bayesian and the classical c–optimal designs coincide for all prior precision matrices R for which $r_{11} = 2r_{13}$ and $r_{12} = 2r_{23}$.

5. Bayesian c–Optimal Designs on the Support of Classical c–Optimal Designs

In the last section conditions on $\boldsymbol{c}$ and R were given so that the classical and Bayesian $\boldsymbol{c}$–optimal designs coincided. In this section we consider the more general problem of when the support points of the designs are the same. It was noticed in Chaloner (1984) that this happened in certain polynomial examples for large n. If n is large one expects the designs to be close. The fact that the supports are identical for large n is not entirely clear. We show this to be the case for any c when the classical design is on a "full set" of k points. Recall k is the number of regression functions. The general result is in Theorem 5.1.

Assume that the design $\xi^* = \{x_i^*, p_i^*\}_{i=1}^m$ is a classical $\boldsymbol{c}$–optimal design. Then it follows from Elfving's Theorem that there exists $\epsilon_i^* \in \{\pm 1\}$, $i = 1, 2, \ldots, m$ and a positive constant $\beta^*(= \beta_0)$ such that

$$\sum_{i=1}^{m} \epsilon_i^* p_i^* f(x_i^*) = \beta^* \boldsymbol{c} \in \partial \mathcal{G} \tag{5.1}$$

Thus it follows from Lemma 2.1 that $\sum_{i=1}^{m} \epsilon_i^* p_i \boldsymbol{f}(x_i) \in \partial \mathcal{G}$ for any set of non–negative weights $p_1, p_2, \ldots, p_m$ for which $\sum_{i=1}^{m} p_i = 1$. Thus it follows from Theorem 3.1 that the design $\xi_0 = \{x_j^*, p_j\}_{i=1}^m$ which puts weight $p_i > 0$ at the points $x_i^* \in \mathcal{X}$, $i = 1, 2, \ldots, m$ is a Bayesian $\boldsymbol{c}$–optimal design if and only if

$$\sum_{i=1}^{m} \epsilon_i^* p_i \boldsymbol{f}(x_i^*) = \left(1 + \tfrac{1}{n} \boldsymbol{d}_0' R \boldsymbol{d}_0\right) \sum_{i=1}^{m} \epsilon_i^* p_i^* \boldsymbol{f}(x_i^*) - \tfrac{1}{n} R \boldsymbol{d}_0 \tag{5.2}$$

for some $\boldsymbol{d}_0 \in \mathcal{D}_{\beta^* c}$. Let $F_1 = [\epsilon_1^* f(x_1^*), \ldots, \epsilon_m^* f(x_m^*)]$ be a $k \times m$ matrix of full rank m and let $F_2 = \left[\epsilon_{m+1}^* f(x_{m+1}^*), \ldots, \epsilon_k^* f(x_k^*)\right]$ be such that $F = [F_1, F_2]$ is a nonsingular $k \times k$ matrix, that is, if $m < k$ we add $k - m$ arbitrary points x_i^* with corresponding weights $p_i^* = 0$, $i = m+1, \ldots, k$ so that $F = [\epsilon_1^* f(x_1^*), \ldots, \epsilon_k^* f(x_k^*)]$ is nonsingular. We also let $\boldsymbol{p}^* = (p_1^*, \ldots, p_m^*)'$, $\boldsymbol{p} = (p_1, \ldots, p_m)'$, $\gamma^* = \beta^*(1 + \frac{1}{n} \boldsymbol{d}_0' R \boldsymbol{d}_0)$, $F^{-1} = \begin{bmatrix} F^{(1)} \\ F^{(2)} \end{bmatrix}$, where $F^{(1)}$ is an $m \times k$ matrix and $\boldsymbol{b} = (F_1' R^{-1} F_1)^{-1} \mathbf{1}$, where $\mathbf{1}$ is the $m \times 1$ vector of ones. It was shown in El–Krunz (1989) that for (5.2) to holds, the prior precision matrix R must satisfy the condition

$$F^{(2)} R \boldsymbol{d}_0 = 0 \text{ for some } \boldsymbol{d}_0 \in \mathcal{D}_{\beta^* c} \tag{5.3}$$

or equivalently

$$R^{-1}F_1(F_1'R^{-1}F_1)^{-1}\mathbf{1} = \boldsymbol{d}_0 \text{ for some } \boldsymbol{d}_0 \in \mathcal{D}_{\beta^* c} \tag{5.4}$$

in which case (5.2) becomes

$$\boldsymbol{p} = \left(1 + \tfrac{1}{n}\mathbf{1}'\boldsymbol{b}\right)\boldsymbol{p}^* - \tfrac{1}{n}\boldsymbol{b} \tag{5.5}$$

Thus it follows from the equivalence of (5.3) and (5.4) that the choice of F_2 is irrelevant. Note also that if $m = k$, then Condition (5.4) becomes $F'^{-1}\mathbf{1} = \boldsymbol{d}_0$ which trivially holds because $\epsilon_i^* \boldsymbol{d}_0' \boldsymbol{f}(x_i^*) = 1, \; i = 1, 2, \ldots, k$ and $\boldsymbol{d}_0$ is the unique supporting hyperplane to $\mathcal{G}$ at the point $\beta^*\boldsymbol{c}$. Thus we have the following result.

Theorem 5.1

Let $\xi^* = \{x_i^*, p_i^*\}_{i=1}^{m}$ be the classical $\boldsymbol{c}$–optimal design and let β^* be such that $\beta^*\boldsymbol{c} \in \partial\mathcal{G}$. Then the design $\xi_0 = \{x_i^*, p_i\}_{i=1}^{m}$ is Bayesian $\boldsymbol{c}$–optimal if and only if (5.4) and (5.5) hold.

Corollary 5.1

Let $\xi^* = \{x_i^*, p_i^*\}_{i=1}^{k}$ be the classical $\boldsymbol{c}$–optimal design, β^* be such that $\beta^*\boldsymbol{c} \in \partial\mathcal{G}, d_0$ be the unique supporting hyperplane to $\mathcal{G}$ at the point $\beta^*\boldsymbol{c} \in \partial\mathcal{G}$, and $\epsilon_i^* = \boldsymbol{d}_0'\boldsymbol{f}(x_i), \; i = 1, 2, \ldots, k$. Then the design $\xi_0 = \{x_i, p_i\}_{i=1}^{k}$ is Bayesian $\boldsymbol{c}$–optimal if and only if

$$\boldsymbol{p} = \frac{\gamma^*}{\beta^*}\boldsymbol{p}^* - \tfrac{1}{n}F^{-1}R\boldsymbol{d}_0 \tag{5.6}$$

■

Let us define the set $\mathcal{R}$ to be

$$\mathcal{R} = \begin{cases} \mathbb{R}^+_{k\times k} & \text{if } m = k \\ \{R: R \in \mathbb{R}^+_{k\times k}, R^{-1}F_1(F_1'R^{-1}F_1)^{-1}1 = \boldsymbol{d}_0 \text{ for some } \boldsymbol{d}_0 \in \mathcal{D}_{\beta^* c}\} & \text{if } m < k \end{cases} \tag{5.7}$$

that is, $\mathcal{R}$ is the set of all positive definite $k \times k$ matrices if $m = k$ and $\mathcal{R}$ is the set of all positive definite matrices for which (5.4) holds if $m < k$.

Also define

$$R^* = \begin{cases} F^{-1}R(F')^{-1} & \text{if } m = k \\ (F_1'R^{-1}F_1)^{-1} & \text{if } m < k \end{cases} \tag{5.8}$$

Then condition (5.5) becomes

$$\boldsymbol{p} = \left(1 + \tfrac{1}{n}\mathbf{1}'R^*\mathbf{1}\right)\boldsymbol{p}^* - \tfrac{1}{n}R^*\mathbf{1} \tag{5.9}$$

which can be written as

$$p_i = p_i^* \left(1 + \tfrac{1}{n}\sum_{i,j=1}^{m} r_{ij}^*\right) - \tfrac{1}{n}\sum_{j=1}^{m} r_{ij}^*,\ i = 1, 2, \ldots, m \tag{5.10}$$

We then have the following result which is very useful in characterizing the set of all matrices R for which the Bayesian $\boldsymbol{c}$–optimal design ξ_0 is supported at the same support points as the classical $\boldsymbol{c}$–optimal design.

Corollary 5.2

Let $\xi^* = \{x_i^*, p_i^*\}_{i=1}^{m}$ be the classical $\boldsymbol{c}$–optimal design. If $R \in \mathcal{R}$, and

$$\sum_{j=1}^{m} r_{ij}^* < p_i^* \left(n + \sum_{i,j=1}^{m} r_{ij}^*\right),\ i = 1, 2, \ldots, m \tag{5.11}$$

then the Bayesian $\boldsymbol{c}$–optimal design puts weights $p_i > 0$ at the points $x_j^* \in \mathcal{X}$, $i = 1, 2, \ldots, m$ and the p_i are given by (5.10), $i = 1, 2, \ldots, m$.

Corollary 5.3

For any $R \in \mathcal{R}$, the Bayesian and the classical $\boldsymbol{c}$–optimal design coincide, i.e., $\xi_0 = \xi^*$, if and only if $\sum\limits_{j=1}^{m} r_{ij}^* = p_i^* \sum\limits_{i,j=1}^{m} r_{ij}^*$.

Proof: If $\xi_0 = \xi^*$, then the result follows from (5.10) of Corollary 5.2. On the other hand if $\sum\limits_{j=1}^{m} r_{ij}^* = p_i^* \sum\limits_{i,j=1}^{m} r_{ij}^*$, $i = 1, 2, \ldots, m$, then it follows from (5.10) that $p_i = p_i^*$, $i = 1, 2, \ldots, m$ and condition (5.11) becomes $np_i^* > 0$ which trivially holds and so it follows from Corollary 5.2 that the classical $\boldsymbol{c}$–optimal design ξ^* is also the Bayesian $\boldsymbol{c}$–optimal design which completes the proof.

Corollary 5.4

For every $R \in \mathcal{R}$, there exists a positive integer $n_0 = n_0(R)$ such that the Bayesian $\boldsymbol{c}$–optimal design ξ_0 is supported on the points of the classical $\boldsymbol{c}$–optimal design ξ^* and the weights of ξ_0 are given by (5.10) for all $n \geq n_0$.

Proof: Assume that $R \in \mathcal{R}$ and R^* is defined as in (5.8). Since $p_i^* > 0$, $i = 1, 2, \ldots, m$, then there exists positive integers n_i, $i = 1, 2, \ldots, m$ such that

$$\sum_{j=1}^{m} r_{ij}^* < p_i^* \left(n + \sum_{i,j=1}^{m} r_{ij}^* \right) \text{ for all } n > n_j, \ i = 1, 2, \ldots, m \tag{5.12}$$

Since m is finite, then we can take $n_0 = \max\{n_1, \ldots, n_m\}$ and so (5.11) holds for all $n \geq n_0$. Therefore the result follows from Corollary 5.2 ■

Corollary 5.4 is of special importance. For example if the classical $\boldsymbol{c}$–optimal design is supported at exactly k distinct points as in the case of extrapolation or estimating the highest coefficient in polynomial regression, then the Bayesian $\boldsymbol{c}$–optimal design is supported at the same points of the classical $\boldsymbol{c}$–optimal design for n large enough. The same is true for any $R \in \mathcal{R}$ if $m < k$, where $\mathcal{R}$ is expected, in general, to be a very large set. In fact, it was shown in El–Krunz (1989) that $\mathcal{R}$ is a nonempty, unbounded set which is the union of closed convex sets with respect to the usual topology defined on the set of all positive definite $k \times k$ matrices. If for any given positive integer n, we define

$$\mathcal{R}^{(n)} = \{R: R \in \mathcal{R} \text{ and support } (\xi_0) = \text{ support } (\xi^*)\} \tag{5.13}$$

to be the set of all positive definite matrices R for which the Bayesian $\boldsymbol{c}$–optimal design ξ_0 is supported at the same points as the classical $\boldsymbol{c}$–optimal design, then $\mathcal{R}^{(n)}$ is also a non–empty unbounded set which is the union of convex sets and the sequence $\{\mathcal{R}^{(n)}\}$ is an increasing sequence in n and $\lim_{n \to \infty} \mathcal{R}^{(n)} = \mathcal{R}$. For moderate values of n, however, the Bayesian $\boldsymbol{c}$–optimal design is not necessarily supported at the same points of the classical $\boldsymbol{c}$–optimal design. This should be clear; the following is an example.

Example 5.1

Assume that the design space $\mathcal{X}$ consists of three points x_1, x_2, x_3, where $\boldsymbol{f}(x_1) = (1,0)'$, $\boldsymbol{f}(x_2) = (1,1)'$, $\boldsymbol{f}(x_3) = (0,2)'$ and assume that $\boldsymbol{c} = (1,3)'$. Thus $\boldsymbol{f}(\mathcal{X})$ is given by

$$\boldsymbol{f}(\mathcal{X}) = \left\{ \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\2\end{pmatrix} \right\} \tag{5.14}$$

Since $\frac{\boldsymbol{c}}{2} = \frac{1}{2}\binom{0}{2} + \frac{1}{2}\binom{1}{1}$, then the classical $\boldsymbol{c}$–optimal design puts equal weights at the two points $(0,2)'$ and $(1,1)'$. Also since $\mathcal{R} = \mathbb{R}^{+}_{2\times 2}$, that is, the set $\mathcal{R}$ is the set of all positive definite 2×2 matrices, then for any positive definite 2×2 matrix $R = ((r_{ij}))^{2}_{i,j=1}$, it follows from Corollary 5.2 that the Bayesian $\boldsymbol{c}$–optimal design puts weights p_1, p_2 at the two points $(0,2)'$ and $(1,1)'$ respectively if and only if

$$-4n < 3r_{11} + 2r_{12} - r_{22} < 4n \tag{5.15}$$

in which case

$$\begin{aligned} p_1 &= \tfrac{1}{2} + \tfrac{1}{8n}(3r_{11} + 2r_{12} - r_{22}) \\ p_2 &= \tfrac{1}{2} - \tfrac{1}{8n}(3r_{11} + 2r_{12} - r_{22}) \end{aligned} \tag{5.16}$$

Thus $\mathcal{R}^{(n)} = \{R\colon R \in \mathbb{R}^{+}_{2\times 2},\ -4n < 3r_{11} + 2r_{12} - r_{22} < 4n\}$ and so for any prior precision matrix $R \in \mathbb{R}^{+}_{2\times 2}$, one can choose n large enough to force condition (5.15) to hold. However, if n is fixed, then for those matrices $R \in \mathbb{R}^{+}_{2\times 2}$ for which condition (5.15) does not hold, the Bayesian $\boldsymbol{c}$–optimal design is no longer supported at the two points $(0,2)'$ and $(1,1)'$. If we define $a = 3r_{11} - r_{12}$ and $b = r_{22} - 3r_{12}$, then using Theorem 3.1, it follows that the Bayesian $\boldsymbol{c}$–optimal design puts weights at the two points $(0,2)'$ and $(1,1)'$ if and only if $a \in (-4n + b,\ 4n + b)$; it puts weights at the two points $(1,0)'$ and $(1,1)'$ if and only if $a \in (-3n, -2n)$; it puts weights at the two points $(0,2)'$ and $(1,0)'$ if and only if $a \in (-3n - \frac{1}{2}b,\ -2n - \frac{1}{2}b)$; it puts all of its weight at the point $(1,1)'$ if and only if $a \in [-2n, -4n + b]$, it puts all of its weight at the point $(0,2)'$ if and only if $a \geq \max\{4n + b, -2n - \frac{1}{2}b\}$, and it puts all of its weight at the point $(1,0)'$ if and only if $a \leq \max\{-3n - \frac{1}{2}b, -3n\}$.

Example 5.2

Consider the cubic polynomial regression model, where $f(x) = (1, x, x^2, x^3)$, $|x| \leq 1$ and assume that $\boldsymbol{c} = (0,0,0,1)$, that is, we are interested in the estimation of θ_4, the coefficient of x^3. Since

$$\tfrac{1}{4}\boldsymbol{c} = -\tfrac{1}{6}\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix} + \tfrac{1}{3}\begin{pmatrix}1\\-\frac{1}{2}\\\frac{1}{4}\\-\frac{1}{3}\end{pmatrix} - \tfrac{1}{3}\begin{pmatrix}1\\\frac{1}{2}\\\frac{1}{4}\\\frac{1}{8}\end{pmatrix} + \tfrac{1}{6}\begin{pmatrix}1\\1\\1\\1\end{pmatrix} \tag{5.17}$$

then $\beta^*\boldsymbol{c} = \sum_{i=1}^{4} \epsilon_i^* p_i^* \boldsymbol{f}(x_i^*)$, where $\beta^* = \frac{1}{4}$, $\epsilon_1^* = -1$, $\epsilon_2^* = 1$, $\epsilon_3^* = -1$, $\epsilon_4^* = 1$, $p_1^* = p_4^* = \frac{1}{6}$, $p_2^* = p_3^* = \frac{1}{3}$ and the x_i^*'s are the "Chebychev" points -1, $-\frac{1}{2}$, $\frac{1}{2}$ and 1. Thus the classical $\boldsymbol{c}$–optimal design ξ^* puts weights $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{3}$ and $\frac{1}{6}$ at the points -1, $-\frac{1}{2}$, $\frac{1}{2}$ and 1 respectively, $\boldsymbol{d}_0 = (0, -3, 0, 4)'$ and $\mathcal{R} = \mathbb{R}^+_{4\times 4}$ and so for any positive definite 4×4 matrix $R = ((r_{ij}))_{i,j=1}^{4}$, there exists a positive integer n_0 such that $R \in \mathcal{R}^{(n)}$ for all $n \geq n_0$. From (5.6) of Corollary (5.1), it follows that

$$\begin{aligned} n_1 &= \tfrac{n}{6} + \tfrac{4}{3}\left(\tfrac{3r_{22}}{4} - r_{24}\right) + \tfrac{2}{3}\left[\left(\tfrac{3r_{12}}{4} - r_{14}\right) - (3r_{23} - 4r_{34})\right] \\ n_2 &= \tfrac{n}{3} - \tfrac{4}{3}\left(\tfrac{3r_{22}}{4} - r_{24}\right) + \tfrac{2}{3}\left[4\left(\tfrac{3r_{12}}{4} - r_{14}\right) - (3r_{23} - 4r_{34})\right] \\ n_3 &= \tfrac{n}{3} - \tfrac{4}{3}\left(\tfrac{3r_{22}}{4} - r_{24}\right) - \tfrac{2}{3}\left[4\left(\tfrac{3r_{12}}{4} - r_{14}\right) - (3r_{23} - 4r_{34})\right] \\ n_4 &= \tfrac{n}{6} + \tfrac{4}{3}\left(\tfrac{3r_{22}}{4} - r_{24}\right) - \tfrac{2}{3}\left[\left(\tfrac{3r_{12}}{4} - r_{14}\right) - (3r_{23} - 4r_{34})\right] \end{aligned} \tag{5.18}$$

where $n_i = np_i$, $i = 1, 2, 3, 4$. Thus if all the n_i's in (5.18) are positive, then $R \in \mathcal{R}^{(n)}$ and the Bayesian $\boldsymbol{c}$–optimal design ξ_0 puts weights p_1, p_2, p_3 and p_4 at the Chebychev points $-1, -\frac{1}{2}, \frac{1}{2}$ and 1 respectively. Moreover it follows from Corollary 5.3 and (5.18) that the Bayesian and the classical $\boldsymbol{c}$–optimal designs coincide if and only if $R \in \mathbb{R}^+_{4\times4}$, $3r_{12} = 4r_{14}$, $3r_{22} = 4r_{24}$ and $3r_{23} = 4r_{34}$.

6. Bayesian Ψ-Optimal Designs

In the previous sections, we considered the case where one is interested in the estimation of a single parametric function of the form $\boldsymbol{c}'\boldsymbol{\theta}$ for some nonrandom $k \times 1$ vector $\boldsymbol{c}$. The generalization of this is the estimation of a linear combination $A'\boldsymbol{\theta}$ for some $k \times s$ matrix A of rank $s \leq k$. Under squared error loss, the linear Bayes estimator for $A'\boldsymbol{\theta}$ is $A'\hat{\boldsymbol{\theta}}_b$, where $\hat{\boldsymbol{\theta}}_b$ is given by (1.2) and the Bayes risk is proportional to tr $\Psi(R + X'X)^{-1}$, where $\Psi = AA'$ is a $k \times k$ matrix of rank $s \leq k$ and R is the prior precision matrix. Thus in terms of the Bayes information matrix, we are interested in minimizing the optimality criterion functional

$$\phi(M_b(\xi)) = \text{ tr } \Psi M_b^{-1}(\xi) \tag{6.1}$$

over the set Ξ of all approximate designs. This criterion is what we called the Ψ–optimality. The main purpose in this section is to extend the results of the previous sections on $\boldsymbol{c}$–optimality for Ψ–optimality and to give a matrix analog of the Elfving's Theorem for Bayesian Ψ–optimal designs. The treatment in this section is similar to that of $\boldsymbol{c}$–optimality and so some of the details will be omitted since the results will be direct extensions of that in the previous section. So let $\epsilon(x) = (\epsilon_1(x), \ldots, \epsilon_s(x))'$ be a vector of s real valued functions defined on the design space $\mathcal{X}$ and define $\mathcal{G}$ as the smallest convex set of $k \times s$ matrices which contains the matrices $f(x)\epsilon'(x)$ for all $x \in \mathcal{X}$ and all functions ϵ for which $|\epsilon(x)| \leq 1$ for all $x \in \mathcal{X}$, where by $|\cdot|$, we mean the usual Euclidean norm. Treating the matrices in $\mathcal{G}$ as vectors in the ks–dimensional Euclidean space, it is not hard to see that $\mathcal{G}$ is a symmetric convex compact subset in the ks–dimensional Euclidean space and that any half line through the origin intersects $\partial\mathcal{G}$ at exactly one point. Thus for any nonzero $k \times s$ matrix A, there exists a unique positive constant β^* such that $\beta^* A \in \partial\mathcal{G}$. Let $\mathbb{R}_{k\times s}$ denote the set of all $k \times s$ matrices and define

$$\mathcal{D} = \{D \in \mathbb{R}_{k\times s}\colon\ \text{tr } D'U \leq 1 \text{ for all } U \in \mathcal{G} \text{ and tr } D'U_0 = 1 \text{ for some } U_0 \in \partial\mathcal{G}\} \tag{6.2}$$

to be the set of all normalized supporting hyperplanes to the surface of $\mathcal{G}$, where here we again identify the hyperplane $\{U \in \mathbb{R}_{k\times s}\colon\ \text{tr } D'U = 1\}$ with its inducing $k \times s$ matrix D. For every $D \in \mathcal{D}$, define the contact set

$$\mathcal{C}(D) = \{U\colon\ U \in \partial\mathcal{G} \text{ and tr } D'U = 1\} \tag{6.3}$$

to be the intersection of the hyperplane D with $\mathcal{G}$. For any point $U_0 \in \partial\mathcal{G}$, let $\mathcal{D}_{U_0} = \{D \in \mathcal{D}\colon\ \text{tr } D'U \leq 1 = \text{ tr } D'U_0 \text{ for all } U \in \mathcal{D}\}$ denote the set of all supporting hyperplanes to $\mathcal{G}$ at U_0. The set $\mathcal{D}_{U_0}$ is either single point or a closed convex set. Now let R be a given $k \times k$ positive definite matrix, n be a given positive integer and as in section 2, let us define the following

$$\mathcal{H} = \{Z \in \mathbb{R}_{k\times s}\colon\ Z = U + \tfrac{1}{n}RD,\ D \in \mathcal{D} \text{ and } U \in \mathcal{C}(D)\} \tag{6.4}$$

$$\mathcal{D}^* = \{D^* \in \mathsf{R}_{k\times s}\colon\ D^* = (1 + \tfrac{1}{n}\ \text{tr}\ D'RD)^{-\frac{1}{2}}D, \quad D \in \mathcal{D}\} \tag{6.5}$$

$$\mathcal{V} = \{V \in \mathsf{R}_{k\times s}\colon\ V = (1 + \tfrac{1}{n}\ \text{tr}\ D'RD)^{-\frac{1}{2}}(U + \tfrac{1}{n}RD), D \in \mathcal{D} \text{ and } U \in \mathcal{C}(D)\} \tag{6.6}$$

and for any set $\mathcal{A}$, we shall use the notation $\overline{\mathcal{A}}$ to mean the set

$$\overline{\mathcal{A}} = \{tA\colon\ a \in \mathcal{A},\ 0 \le t \le 1\}. \tag{6.7}$$

The following results follow directly from the corresponding results in Section 2 by considering every $k \times s$ matrix A as a vector a in the ks–dimensional Euclidean space and replacing k in Section 2 by ks.

<u>Lemma 6.1</u>

The set $\overline{\mathcal{D}}$ is a compact symmetric convex set in $\mathsf{R}_{k\times s}$ which has $\mathcal{D}$ as its boundary. Moreover for any nonzero $k \times s$ matrix A, there exists a unique positive constant α such that $\alpha A \in \mathcal{D}$.

<u>Lemma 6.2</u>

The set $\overline{\mathcal{D}^*}$ is a compact symmetric convex set in $\mathsf{R}_{k\times s}$ which has $\mathcal{D}^*$ as its boundary. Moreover for any nonzero $k \times s$ matrix A, there exists a unique positive constant α_0 such that $\alpha_0\ A \in \mathcal{D}^*$.

<u>Lemma 6.3</u>

The set $\overline{\mathcal{V}}$ is a compact symmetric convex set in $\mathsf{R}_{k\times s}$ which has $\mathcal{V}$ as its boundary. Moreover for any nonzero $k \times s$ matrix A, there exists a unique positive constant δ_0 such that $\delta_0 A \in \mathcal{V}$.

<u>Lemma 6.4</u>

The set $\overline{\mathcal{H}}$ is a compact symmetric set $\mathsf{R}_{k\times s}$ which has $\mathcal{H}$ as its boundary and $\mathcal{G} \subseteq \overline{\mathcal{V}} \subseteq \overline{\mathcal{H}}$. Moreover for any nonzero $k \times s$ matrix A, there exists a unique positive constant γ_0 such that $\gamma_0 A \in \mathcal{H}$ and the representation

$$U_0 + \tfrac{1}{n}RD_0 = \gamma_0 A \tag{6.8}$$

is unique, that is, there exists a unique pair (U_0, D_0), $D_0 \in \mathcal{D}$ and $U_0 \in \mathcal{C}(D_0)$ such that (6.8) holds.

If we define $\beta_0^{-1} = \text{ tr } D_0'A$, then premultiplication of both sides of (6.8) by D_0' and taking the trace gives

$$\gamma_0 = \beta_0(1 + \tfrac{1}{n} \text{ tr } D_0'RD_0). \tag{6.9}$$

Note here that $\beta_0 A$ lies on the hyperplane D_0 but not necessarily on $\partial\mathcal{G}$ and so if β^* is such that $\beta^* A \in \partial\mathcal{G}$, then $\beta_0 \geq \beta^*$. From the definition of $\mathcal{G}$, it also follows that every matrix $U \in \mathcal{G}$ has a representation $U = \sum_{i=1}^{m} p_i f(x_i)\epsilon'(x_i)$ for some positive integer m, $p_i > 0$, $\sum_{i=1}^{m} p_i = 1$, $|\epsilon(x_i)| \leq 1$ and $x_i \in \mathcal{X}$, $i = 1, 2, \ldots, m$ and that every matrix $U_0 \in \partial\mathcal{G}$ has a representation

$$U_0 = \sum_{i=1}^{m} p_i f(x_i)\epsilon'(x_i) \tag{6.10}$$

for some positive integer m, $p_i > 0$, $\sum_{i=1}^{m} p_i = 1$, $|\epsilon(x_i)| = 1$ and $x_i \in \mathcal{X}$, $i = 1, 2, \ldots, m$. The following lemma, which is the matrix analog of Lemma 2.1, characterizes the boundary points of the symmetric convex compact set $\mathcal{G}$.

Lemma 6.5

A matrix U_0 of the form (6.10) is a boundary point of $\mathcal{G}$ if and only if there exists a $k \times s$ matrix D_0 such that

$$|D_0' f(x)| \leq 1 \text{ for all } x \in \mathcal{X} \tag{6.11}$$

and equality holds for each x_i with $\epsilon(x_i) = D_0' f(x_i)$, $i = 1, 2, \ldots, m$ and tr $D_0' U_0 = 1$. ■

The $k \times s$ matrix D_0 in Lemma 6.5 defines the hyperplanes supporting $\mathcal{G}$ at its boundary point U_0. From (6.10), it follows that the unique representation (6.8) can be written as

$$\sum_{i=1}^{m} p_i f(x_i)\epsilon'(x_i) + \tfrac{1}{n} RD_0 = \gamma_0 A \tag{6.12}$$

for some positive integer m, $p_i > 0$, $\sum_{i=1}^{m} p_i = 1$, $|\epsilon(x_i)| = 1$ and $x_i \in \mathcal{X}$, $i = 1, 2, \ldots, m$, where γ_0 is the unique positive constant for which $\gamma_0 A \in \mathcal{H}$ and D_0 is a supporting hyperplane to $\mathcal{G}$ at the point $U_0 = \sum_{i=1}^{m} p_i f(x_i)\epsilon'(x_i)$ with $\epsilon(x_i) = D_0' f(x_i)$, $i = 1, 2, \ldots, m$. Let $\xi = \{x_i, p_i\}_{i=1}^{m}$ be any design in Ξ and D be any $k \times s$ matrix in $\mathcal{D}$. Then

$$\text{tr } D^{*'} M_b(\xi) D^* = \left(1 + \tfrac{1}{n} \text{ tr } D'RD\right)^{-1} \text{ tr } \left(\sum_{i=1}^{m} p_i D' f(x_i) f'(x_i) D + \tfrac{1}{n} D'RD\right)$$

$$= \left(1 + \tfrac{1}{n} \text{ tr } D'RD\right)^{-1} \left(\sum_{i=1}^{m} p_i |D'f(x_i)|^2 + \tfrac{1}{n} \text{ tr } D'RD\right) \tag{6.13}$$

Since $|D'f(x)| \leq 1$ for all $x \in \mathcal{X}$, then $|D'f(x_i)| \leq 1$, $i = 1, 2, \ldots, m$ and so $\sum_{i=1}^{m} p_i |D'f(x_i)|^2 \leq \sum_{i=1}^{n} p_i = 1$. Thus it follows from this and (6.13) that

$$\text{tr } D^{*'} M_b(\xi) D^* \leq 1 \text{ for all } \xi \in \Xi \text{ and all } D \in \mathcal{D} \tag{6.14}$$

with equality holding if and only $|D'f(x)| = 1$ for all x in the support of the design ξ. In other words, equality holds in (6.14) if and only if $U = \int_{\mathcal{X}} f(x)\epsilon'(x)\xi(dx) \in \partial\mathcal{G}$ for some function $\epsilon(x)$ satisfying $|\epsilon(x)| = 1$ for all x in the support of the design ξ and D is the supporting hyperplane to the surface of $\mathcal{G}$ at the point U.

We now prove the main result of this section which is the matrix analog for Bayesian Ψ–optimal designs of Theorem 3.1.

Theorem 6.1

Given a nonzero $k \times s$ matrix A and a $k \times k$ positive definite matrix R, the design ξ_0 is Bayesian Ψ–optimum if and only if A has the representation (6.12) with $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$. Bayesian Ψ–optimal designs always exist and

$$\inf_{\xi \in \Xi} \text{ tr } A'M_b^{-1}(\xi)A = \text{ tr } A'M_b^{-1}(\xi_0)A = \rho(A) = \tfrac{1}{\delta_0^2} = \tfrac{1}{\beta_0\gamma_0}.$$

Proof: First from Schwarz's inequality, it follows that

$$\text{tr } A'M_b^{-1}(\xi)A = \sup_{D \in \mathcal{D}} \frac{(\text{tr } D'A)^2}{\text{tr } D'M_b(\xi)D} = \sup_{D^* \in \mathcal{D}^*} \frac{(\text{tr } D^{*'}A)^2}{\text{tr } D^{*'}M_b(\xi)D^*} \geq \frac{(\text{tr } D^{*'}A)^2}{\text{tr } D^{*'}M_b(\xi)D^*} \tag{6.15}$$

which implies that

$$\text{tr } D^{*'} M_b(\xi) D^* \geq (\text{tr } A'M_b^{-1}(\xi)A)^{-1} (\text{tr } D^{*'}A)^2 \text{ for all } \xi \in \Xi \text{ and all } D^* \in \mathcal{D}^*. \tag{6.16}$$

From (6.14) and (6.16), it follows that

$$\inf_{\xi \in \Xi} \text{ tr } A'M_b^{-1}(\xi)A \geq \sup_{D^* \in \mathcal{D}^*} (\text{tr } D^{*'}A)^2. \tag{6.17}$$

Since A has the representation

$$\sum_{i=1}^{m} p_i f(x_i)\epsilon'(x_i) + \tfrac{1}{n} R D_0 = \gamma_0 A \tag{6.18}$$

where γ_0, p_i, x_i, $\epsilon(x_i)$, $i = 1, 2, \ldots, m$ and D_0 are as in (6.9) and (6.12), then

$$(\text{tr } D_0^{*'} A)^2 = (1 + \tfrac{1}{n} \text{ tr } D_0' R D_0)^{-1} (\text{tr } D_0' A)^2 = \beta_0^{-2} (1 + \tfrac{1}{n} \text{ tr } D_0' R D_0)^{-1} = \frac{1}{\beta_0 \gamma_0} = \frac{1}{\delta_0^2}. \tag{6.19}$$

It also follows from (6.18) that

$$\text{tr } A' M_b^{-1}(\xi_0) A = \gamma_0^{-1} \text{ tr } D_0' A = \frac{1}{\beta_0 \gamma_0} = \frac{1}{\delta_0^2} \tag{6.20}$$

if and only if $\xi_0 = \{x_i, p_i\}_{i=1}^{m}$ or equivalently $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$. Thus it follows from (6.19) and (6.20) that

$$\text{tr } A' M_b^{-1}(\xi_0) A = (\text{tr } D_0^{*'} A)^2 = \frac{1}{\beta_0 \gamma_0} = \frac{1}{\delta_0^2} \tag{6.21}$$

if and only if $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$ and so it follows from this and (6.17) that

$$\inf_{\xi \in \Xi} \text{ tr } A' M_b^{-1}(\xi) A = \text{ tr } A' M_b^{-1}(\xi_0) A = (\text{tr } D_0^{*'} A)^2 = \sup_{D^* \in \mathcal{D}^*} (\text{tr } D^{*'} A)^2 = \frac{1}{\beta_0 \gamma_0} = \frac{1}{\delta_0^2} \tag{6.22}$$

if and only if $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$. Thus ξ_0 is a Bayesian $\boldsymbol{c}$–optimal design if and only if A has the representation (6.12) with $\xi_0(x_i) = p_i$, $i = 1, 2, \ldots, m$ and $\inf_{\xi \in \Xi}$ tr $A' M_b^{-1}(\xi) A = \text{tr } A' M_b^{-1}(\xi_0) A = \rho(A) = \frac{1}{\beta_0 \gamma_0} = \frac{1}{\delta_0^2}$ which completes the proof. ■

The following result follows directly from (6.22).

<u>Corollary 6.1</u>

The Bayesian Ψ–optimal design problem

$$\text{Minimize tr } A' M_b^{-1}(\xi) A \text{ subject to } \xi \in \Xi \tag{6.23}$$

is the dual of the problem

$$\text{Maximize } (\text{tr } D^{*'} A)^2 \text{ subject to } D^* \in \mathcal{D}^* \tag{6.24}$$

and the two problems share a common extreme value. ■

Theorem 6.1 is the Bayesian analog of a result of Studden (1971) for classical Ψ–optimal designs which is stated in the following lemma

Lemma 6.6

The design ξ^* is a classical Ψ–optimal design if and only if there exists a function $\epsilon(x)$ satisfying $|\epsilon(x)| = 1$ such that (i) $\int_{\mathcal{X}} f(x)\epsilon'(x)\xi^*(dx) = \beta^* A$ for some scalar β^* and (ii) $\beta^* A \in \partial\mathcal{G}$. Moreover $\beta^* A \in \partial\mathcal{G}$ if and only if $\inf_{\xi} \text{tr } A'M^{-1}(\xi)A = \beta^{*-2}$. ■

To see the parallelism between Theorem 6.1 and its classical analog Lemma 6.1, let $\mathcal{E}$ denote the set of all functions $\epsilon = (\epsilon_1, \ldots, \epsilon_s)'$ defined on $\mathcal{X}$ and satisfying $|\epsilon(x)| = 1$. In Lemma 6.6 we find the unique positive constant β^* such that $\beta^* A \in \partial\mathcal{G}$ and the classical Ψ–optimal design is then the design ξ^* for which

$$\int_{\mathcal{X}} f(x)\epsilon'(x)\xi^*(dx) = \beta^* A \text{ for some } \epsilon \in \mathcal{E} \tag{6.25}$$

in which case the infimum of tr $A'M^{-1}(\xi)A$ among all design ξ for which $A'\theta$ is estimable is equal to β^{*-2} and is attained at $\xi = \xi^*$. In Theorem 6.1 we find the unique positive constant δ_0 such that $\delta_0 A \in \mathcal{V}$, or, equivalently the unique positive constant γ_0 such that $\gamma_0 A \in \mathcal{H}$ and the Bayesian Ψ–optimal design is then the design ξ_0 for which

$$\int_{\mathcal{X}} f(x)\epsilon'(x)\xi_0(dx) = U_0 \text{ for some } \epsilon \in \mathcal{E} \tag{6.26}$$

where $U_0 \in \partial\mathcal{G}$ is uniquely determined by the representation $U_0 + \frac{1}{n}RD_0 = \gamma_0 A$ or equivalently the representation $(1 + \frac{1}{n} \text{ tr } D_0'RD_0)^{-\frac{1}{2}}(U_0 + \frac{1}{n}RD_0) = \delta_0 A$ of A and D_0 is the supporting hyperplane to $\mathcal{G}$ at U_0 normalized so that tr $D_0'U_0 = 1$. The infimum of tr $A'M_b^{-1}(\xi)A$ among all designs ξ is equal to δ_0^{-2} and is attained at $\xi = \xi_0$. Following similar steps to those by which Corollary 6.1 was derived, one can easily see that the classical Ψ–optimal design problem is the dual of the problem

$$\text{Maximize } (\text{tr } D'A)^2 \text{ subject to } D \in \mathcal{D} \tag{6.27}$$

and that the two problems share a common extreme value β^{*-2}. Thus finding the classical and the Ψ–optimal designs can be achieved geometrically by visualizing the convex sets

$\mathcal{G}$ and $\mathcal{V}$ or equivalently the set $\mathcal{G}$ and $\mathcal{H}$ although this, in general, is not an easy task. Although Theorem 6.1 is mathematically attractive, the application of this theorem, as well as Lemma 6.6 is, at present, somewhat limited, however, the above theorem can be very useful in the characterization of those R's, for a given value of n, for which the Bayesian Ψ–optimal design ξ_0 is supported on the same support points of the classical Ψ–optimal design ξ^*.

7. Bayesian Ψ–Optimal Designs on the Support of Classical Ψ–Optimal Designs

Assume that the boundary representation

$$\beta^* A = \sum_{i=1}^{m} p_i^* \boldsymbol{f}(x_i^*)\epsilon'(x_i^*), \ |\epsilon(x_i^*)| = 1, \ p_i^* > 0, \ \sum_{i=1}^{m} p_i^* = 1 \tag{7.1}$$

holds with $m \leq k$, that is, the classical Ψ–optimal design ξ^* is supported at $m \leq k$ distinct points $x_1^*, \ldots, x_m^*$. If $m < k$ we add $k - m$ arbitrary points x_i^* with corresponding weights $p_i^* = 0$, $i = m+1, \ldots, k$ so that $F = [\boldsymbol{f}(x_1^*), \ldots, \boldsymbol{f}(x_k^*)]$ is nonsingular. Let $T = F^{-1}$ and let $\boldsymbol{\ell}(x) = T\boldsymbol{f}(x)$ denote the vector of Lagrange functions for the points $x_1^*, \ldots, x_k^*$. If we multiply (7.1) by T and let $TA = B$, we get

$$\beta^* B = \sum_{i=1}^{k} p_i^* \boldsymbol{\ell}(x_i^*)\epsilon'(x_i^*). \tag{7.2}$$

Since $\ell_i(x_j^*) = \delta_{ij}$, $i, j = 1, 2, \ldots, k$, then it follows from (7.2) that $\beta^* \boldsymbol{b}_i = p_i^* \epsilon(x_i^*)$, $i = 1, 2, \ldots, k$, where b_i' denotes the ith row of B. Thus it follows that

$$\beta^* = \left(\sum_{i=1}^{k} |\boldsymbol{b}_i| \right)^{-1}, \ p_i^* = \beta^* |\boldsymbol{b}_i| \text{ and } \epsilon(x_i^*) = \boldsymbol{b}_i |\boldsymbol{b}_i|^{-1}. \tag{7.3}$$

Note here that if m in (7.1) is less than k, then $\boldsymbol{b}_i = 0$, $i = m+1, \ldots, k$. In this case we let $|\boldsymbol{b}_i|^{-1} = 0$ and $\epsilon(x_i^*) = 0$ whenever $|\boldsymbol{b}_i| = 0$. Let us also define $B_0 = B_d^{-1} B$ where B_d^{-1} is the diagonal matrix with $|\boldsymbol{b}_i|^{-1}$ as its ith diagonal element, $i = 1, 2, \ldots, k$.

The following result characterizes the matrices A with a classical Ψ–optimal design supported on a given set of points $x_1^*, x_2^*, \ldots, x_k^*$.

Lemma 7.1 (Studden (1977))

If F is nonsingular, then a classical Ψ–optimal design ξ^* is supported on $x_1^*, x_2^*, \ldots, x_k^*$ if and only if there exists a $k \times s$ matrix B such that

(i) $\ell'(x)B_0B_0'\ell(x) \leq 1$ for all $x \in \mathcal{X}$

(ii) $A = FB$ ■

Let $\Delta_0 = (\epsilon(x_1^*), \ldots, \epsilon(x_k^*))$ and let B be the $k \times s$ matrix for which (i) and (ii) of Lemma 7.1 hold. Then it follows from (7.3) that $\Delta_0 = B_0' = D_0'F$. Let P and P^* be the diagonal matrices with nonnegative diagonal elements p_i and p_i^* respectively, $i = 1, 2, \ldots, k$. Since the classical Ψ–optimal design ξ^* puts weights p_i^* at the points x_i^*, $i = 1, 2, \ldots, k$, then it follows from Lemma 6.6 that $FP^*\Delta_0' = \sum_{i=1}^{k} p_i^* f(x_i^*)\epsilon'(x_i^*) = \beta^* A \in \partial\mathcal{G}$ and so $FP\Delta_0' = \sum_{i=1}^{k} p_i f(x_i^*)\epsilon'(x_i^*) \in \partial\mathcal{G}$ for all diagonal matrices P with nonnegative diagonal elements p_i, $i = 1, 2, \ldots, k$ for which $\sum_{i=1}^{k} p_i = 1$, and D_0 is a supporting hyperplane to $\mathcal{G}$ at the points $U_0 = FP\Delta_0' \in \partial\mathcal{G}$. Thus the following result is direct consequence of Theorem 6.1.

Corollary 7.1

The design ξ_0 which puts weights $p_i \geq 0$ at the support points x_i^*, $i = 1, 2, \ldots, k$ of the classical Ψ–optimal design ξ^* is Bayesian Ψ–optimum if and only if

$$P\Delta_0' = (\mathbf{1} + \tfrac{1}{n} \text{ tr } \Delta_0 R^* \Delta_0')P^*\Delta_0' - \tfrac{1}{n}R^*\Delta_0' \tag{7.4}$$

where $R^* = TRT'$. ■

Remark 7.1

Note that the matrix R^* is defined slightly different than in section 5. If $s = 1$, then $\epsilon(x_i^*) = \pm 1$, $i = 1, 2, \ldots, m$ and $\epsilon(x_i^*) = 0$, $i = m+1, \ldots, k$ and so Corollary 7.1 reduces to Corollary 5.2 of Section 5 and (7.4) becomes

$$p_i = p_i^* \left(1 + \tfrac{1}{n} \sum_{i,j=1}^{m} \epsilon(x_i^*)\epsilon(x_j^*)r_{ij}^*\right) - \tfrac{1}{n}\sum_{j=1}^{m} \epsilon(x_j^*)r_{ij}^*, \; i = 1, 2, \ldots, m \tag{7.5}$$

which is equation (5.10) of Section 5.

Remark 7.2

If $s = k$, then it follows from (7.4) that R^* must be diagonal and so tr $\Delta_0 R^* \Delta_0' =$ tr $R^* \Delta_0' \Delta_0 =$ tr R^*. Thus we have the following result.

Corollary 7.2

The design ξ_0 which puts weights $p_i \geq 0$ at the support points x_i^*, $i = 1, 2, \ldots, k$ of the classical Ψ–optimal design ξ^* is Bayesian Ψ–optimum if and only if

$$\begin{aligned} &\text{(i) } R^* = TRT' \text{ is diagonal} \\ &\text{(ii) } P = (1 + \tfrac{1}{n} \text{ tr } R^*)P^* - \tfrac{1}{n}R^* \end{aligned} \tag{7.6}$$

Example 7.1

Assume that the design space $\mathcal{X}$ is the k dimensional unit ball

$$\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^k : \boldsymbol{x}'\boldsymbol{x} \leq 1\} \tag{7.7}$$

and consider the multiple linear regression model

$$E(y) = \theta'\boldsymbol{x}, \ \boldsymbol{x} \in \mathcal{X} \tag{7.8}$$

and assume that A is a $k \times k$ matrix of full rank k. From the equivalence theorem for classical Ψ–optimal designs, we know that ξ^* is a classical Ψ–optimal design if and only if

$$\max_{\xi \in \Xi} \text{ tr } M^{-1}(\xi^*)\Psi M^{-1}(\xi^*)M(\xi) = \text{ tr } \Psi M^{-1}(\xi^*). \tag{7.9}$$

Since $\Psi = AA'$ is a positive definite $k \times k$ matrix, then there exists an orthogonal matrix $U = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k)$ such that $U\Lambda U' = \Psi$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)$. Let $F = U$ and assume that $M(\xi^*) = FP^*F'$. Then tr $\Psi M^{-1}(\xi^*) = \sum_{i=1}^{k} \frac{\lambda_i}{p_i^*}$ and so if we choose p_i^* to be proportional to $\sqrt{\lambda_i}$, it follows that $cp_i^* = \sqrt{\lambda_i}$, $c = \sum_{i=1}^{k} \sqrt{\lambda_i}$,

$$\text{tr } \Psi M^{-1}(\xi^*) = \sum_{i=1}^{k} c^2 p_i^* = c^2 \tag{7.10}$$

and

$$\text{tr } M^{-1}(\xi^*)\Psi M^{-1}(\xi^*)M(\xi) = c^2 \text{ tr } M(\xi) \leq c^2 \text{ for all } \xi \in \Xi \tag{7.11}$$

with equality holding if $\xi = \xi^*$. Thus it follows from (7.9) that ξ^* is a classical Ψ–optimal design. Now for the design ξ_0 which puts weights $p_i \geq 0$ at the points $\boldsymbol{x}_i^* = \boldsymbol{u}_i$, $i = 1, 2, \ldots, k$ to be a Bayesian Ψ–optimal design, the precision matrix R must satisfy the conditions

(i) $R^* = U'RU$ is diagonal and R must have the same eigenvector as Ψ and $R^* = \text{diag }(r_1^*, \ldots, r_k^*)$.

(ii) $p_i = (1 + \frac{1}{n}\sum_{i=1}^{k} r_i^*)\frac{\sqrt{\lambda_i}}{\sum_{i=1}^{k}\sqrt{\lambda_i}} - \frac{1}{n}r_i^*,\ i = 1, 2, \ldots, k.$

In other words, we must choose the p_i's in such a way that $p_i \geq 0$, $\sum_{i=1}^{k} p_i = 1$ and $p_i + \frac{1}{n}r_i^*$ is proportional to $\sqrt{\lambda_i}$, $i = 1, 2, \ldots, k$.

The above approach is similar to the one adopted by Pilz (1983). His approach is based on the idea of maximum compactness of the eigenvectors of the Bayesian information matrix. For instance in the case $A = I$, he assumed the existence of an optimal design whose information matrix has the same eigenvectors as the prior precision matrix R and chooses the p_i's in such a way to make a maximum number of the smallest values $p_i + \frac{1}{n}r_i^*$ get equal, where r_i^*, $i = 1, 2, \ldots, k$ are the eigenvalues of R.

In the case of polynomial regression with A being a $k \times k$ matrix of full rank, it is well known that the classical Ψ–optimal design ξ^* puts weights at k distinct points and $p_i^* \propto \sqrt{k_{ii}}$, where $K = ((k_{ij}))_{i,j=1}^{k} = T\Psi T'$. Thus if $R^* = TRT'$ is diagonal, then the design ξ_0, supported at the same support points of the classical Ψ–optimal design ξ^*, is Bayesian Ψ–optimal if and only if (7.6) holds.

Example 7.2

Consider the quadratic regression model with $f'(x) = (1, x, x^2)'$, $x \in [-1, 1]$ and assume that $A = I$. The classical Ψ–optimal design ξ^* puts weights $p_1^* = \frac{1}{4}$, $p_2^* = \frac{1}{2}$ and

$p_3^* = \frac{1}{4}$ at the points $x_1^* = -1,\ x_2^* = 0$ and $x_3^* = 1$ respectively and we have

$$B = F^{-1} = T = \begin{bmatrix} 0 & -\frac{1}{2} & \frac{1}{2} \\ 1 & 0 & -1 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix},\ \Delta_0' = B_0 = \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \text{ and } \beta^* = \frac{1}{2\sqrt{2}}.$$

From Corollary 7.2, it follows that

$$\mathcal{R}^{(n)} = \left\{ R \in \mathbb{R}^{+}_{3\times 3}\colon\ TRT' = \text{ diag } (r_1^*, r_2^*, r_3^*) \text{ and } r_i^* \leq \left(n + \sum_{i=1}^{3} r_i^* \right) p_i^*,\ i = 1,2,3 \right\} \tag{7.12}$$

is the set of all prior precision matrices R for which the Bayesian Ψ–optimal design ξ_0 puts weights p_1, p_2 and p_3 at the points $x_1^* = -1,\ x_2^* = 0$ and $x_3^* = 1$ respectively, and

$$p_i = (1 + \tfrac{1}{n} \sum_{i=1}^{3} r_i^*) p_i^* - \tfrac{1}{n} r_i^*,\ i = 1,2,3 \tag{7.13}$$

Remark 7.3

Assume that A is a $k \times k$ matrix of full rank k and that the classical Ψ–optimal design ξ^* puts weights $p_i^* > 0,\ i = 1,2,\ldots,k$ at exactly k distinct points $x_1^*, x_2^*, \ldots, x_k^*$. Let $\mathcal{R}$ denote the set of all positive definite matrices R for which $R^* = TRT'$ is diagonal and let $R^* = \text{diag } (r_1^*, \ldots, r_k^*)$. Then it follows from Corollary 7.2 that if

$$p_i = (1 + \frac{1}{n} \sum_{i=1}^{k} r_i^*) p_i^* - \frac{1}{n} r_i^* \geq 0,\ i = 1,2,\ldots,k \tag{7.14}$$

then the design ξ_0 which puts weights p_i at the points $x_i^*,\ i = 1,2,\ldots,k$ is a Bayesian Ψ–optimal design. Since k is finite, then it follows from (7.14) that for any $R \in \mathcal{R}$, there exists n_0 which depends on R such that (7.14) holds for all $n \geq n_0$. Thus if $R \in \mathcal{R}$ and n is large enough, there exists a Bayesian Ψ–optimal design on the support of the classical Ψ–optimal design and the optimal weights of the Bayesian Ψ–optimal design are given by (7.14).

References

[1] Bourbaki, K. (1966). *General Topology*. Part 2. Addison-Wesley, Massachusetts.

[2] Chaloner, K. (1982). *Optimal Bayesian experimental design for linear models.* Ph.D. Thesis. Carnegie-Mellon Technical Report No. 238.

[3] Chaloner, K. (1984). *Optimal Bayesian experimental design for linear models.* Ann. Statist. 12:283–300.

[4] Duncan, G. and DeGroot, M. H. (1976). *A mean squared error approach to optimal design theory.* Proceeding of 1976 conference on information, sciences and systems. The John Hopkins University. 217–221.

[5] Elfving, G. (1952). *Optimum allocation in linear regression theory.* Ann. Math. Statist. 23:255–262.

[6] El Krunz, S. (1989). *Bayesian Optimal Designs for Linear regression Models.* Ph.D. Thesis. Purdue University.

[7] Karlin, S. and Studden, W. J. (1966). *Optimal experimental designs.* Ann. Math. Statist. 37:783–815.

[8] Kiefer, J. (1959). *Optimal experimental designs.* J. Roy. Statist. Soc. Ser. B 21:272–319.

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

[10] Kiefer, J. and Wolfowitz, J. (1959). *Optimum designs in regression problems.* Ann. Math. Statist. 30:271–294.

[11] Pilz, J. (1983). *Bayesian Estimation and Experimental Designs in Linear Regression Models.* Teubner-Texte, Leipzig.

[12] Sinha, B. K. (1970). *A Bayesian approach to optimum allocation in regression problems.* Calcutta Statistical Association. Bulletin 19:45–52.

[13] Studden, W. J. (1968). *Optimal designs on Tchebyscheff points.* Ann. Math. Statist. 39:1435–1447.

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