

**A Predictive Approach to the Bayesian Design Problem  
with Application to Normal Regression Models**

by

Morris L. Eaton<sup>1</sup>, Alessandra Giovagnoli, and Paola Sebastiani  
Univ. of Minnesota, Univ. of Udine and Univ. of Perugia

Technical Report #598  
April 1994

---

<sup>1</sup>Research supported in part by the National Science Foundation Grant DMS-93-00973  
and a grant from the Italian National Research Council (CNR)

## **Abstract**

A predictive-decision theoretic approach is developed for the Bayesian design problem. The loss functions used are Fair Bayes (proper scoring rules) and are quadratic measures of distance between probability measures. The results are applied to certain normal regression models where explicit optimal designs are constructed.

AMS 1980 Subject Classifications: Primary 62C10; Secondary 62K05, 62F15

Key words and phrases: Bayesian optimal design, prediction, decision theory, normal theory regression, Fair Bayes loss functions.

# 1 Introduction

Wald's decision theory provides a convenient framework in which to formulate some problems of experimental design—especially those with a Bayesian bent when a prior distribution is specified. Although the Bayesian design problem is easy to formulate in this way, the resulting minimization problem is often intractable so optimal designs cannot be obtained for concrete problems. In this paper, we present a Bayesian design problem in a decision theoretic framework for which optimal design can be found in some important cases. In essence, we consider a prediction problem which has a design aspect to it, and then cast the problem in a decision theoretic model. Incorporating the prior distribution into the problem leads to an objective function (namely, the Bayes risk) which is a function of the design. Finally, selecting the design to minimize the objective function leads to a solution to the design problem.

Before introducing the prediction aspect of the problem, it is useful to first review a standard decision theoretic formulation of the Bayesian design problem. Suppose  $Y$  is an observable random vector whose distribution belongs to a family

$$(1.1) \quad \{P(\cdot|\theta, x)|\theta \in \Theta, x \in \mathcal{X}\}$$

where  $\theta$  is an unknown parameter of the model, and  $x$  represents a design which the experimenter can choose from the design set  $\mathcal{X}$ . Next, consider a particular decision problem with an action space  $\mathcal{A}$  and a loss function  $L$  defined on  $\mathcal{A} \times \Theta \times \mathcal{X}$ . A decision rule  $\delta$  is a function of  $Y$  and  $x$  which takes values in  $\mathcal{A}$ . The risk function of a decision rule  $\delta$  is

$$(1.2) \quad R(\delta, \theta, x) = \mathcal{E}_{\theta, x} L(\delta(Y, x), \theta, x)$$

where  $\mathcal{E}_{\theta, x}$  denotes expectation under the distribution  $P(\cdot|\theta, x)$ . Given a proper prior distribution  $\pi$  on  $\Theta$ , the Bayes risk is

$$(1.3) \quad r_\pi(x) = \inf_{\delta} \int_{\Theta} R(\delta, \theta, x) \pi(d\theta)$$

where the inf ranges over all decision rules. Finally, any  $x_0 \in \mathcal{X}$  which satisfies

$$(1.4) \quad r_\pi(x_0) = \inf_{x \in \mathcal{X}} r_\pi(x)$$

is a *Bayesian optimal design* for the decision problem at hand when the prior distribution is  $\pi$ . In other words,  $x_0$  is optimal if it minimizes the Bayes risk (1.3).

In this paper, the decision problem is a prediction problem, although our formulation of the prediction problem is somewhat unusual. It is based on work in Eaton (1982, 1992). This formulation is described in detail in Sections 2 and 3, and leads to a useful expression for the Bayes risk  $r_\pi(x)$  defined at (1.3).

In Section 4, a prediction version of the normal linear regression model (with known covariance) is treated in detail. This leads to a criterion function (see 4.15) whose minimization yields an optimal design. Various specializations of this criterion give some well-known classical and Bayesian design criteria.

The prediction viewpoint in experimental design goes back to Lindley (1968), who assumes an ignorance prior, and is also employed in a series of papers by Brooks (1972, 1974, 1976) who allows for use of prior knowledge. Piccinato (1980) sketches a useful outline. An important contribution to Bayes designs for linear models is Chaloner (1984), who deals also with designs for prediction at a point. The book by Pilz (1991) adopts a decision theoretic approach to designs for estimation, and employs several Bayesian optimality criteria, some of which can be thought of as specializations of the criterion of this paper. The same applies to some of the well-known classical design criteria for linear models that can be found in the design literature, see for instance the books by Fedorov (1972), Silvey (1980), and Pukelsheim (1993). An overview of modern trends in Bayesian designs is given by Verdinelli (1992).

## 2 Quadratic Forms in Probability Measures

The loss functions for the prediction problem described in the next section are reviewed below. In essence, these loss functions are quadratic forms in bounded signed measures, and are natural analogues of quadratic forms on Euclidean spaces. Recall that for any  $n \times n$  non-negative definite matrix  $A$ , the quadratic form

$$(2.1) \quad Q(x) = x'Ax, x \in R^n$$

can be used to define a pseudo-norm and hence a pseudo-metric on  $R^n$ . Namely, for  $x$  and  $y$  in  $R^n$ , set

$$(2.2) \quad \|x - y\|^2 = (x - y)'A(x - y).$$

Then  $\|x - y\| \equiv d(x, y)$  is a pseudo-metric on  $R^n$ —but, not necessarily a metric since  $d(x, y) = 0$  does not necessarily imply that  $x = y$ .

Now, let  $p$  and  $q$  be probability vectors in  $R^n$  so these specify probability measures (p.m.) on the finite set  $\{1, 2, \dots, n\}$ . Further,  $\|p - q\|$  is a “distance” between any two such p.m.’s. The extension of this construction to p.m.’s on infinite sets follows.

Consider a measurable space  $(\mathcal{U}, \mathcal{B})$  and let  $\mathcal{M}(\mathcal{U})$  be the set of all bounded signed measures defined on  $\mathcal{B}$ . In particular, if  $P_1$  and  $P_2$  p.m.’s defined on  $\mathcal{B}$ , then  $P_1 - P_2 \in \mathcal{M}(\mathcal{U})$ .

**Definition 2.1** Let  $K(u, v)$  be a bounded measurable complex valued function defined on  $\mathcal{U} \times \mathcal{U}$ . Then  $K$  is *non-negative definite* (n.n.d.) if

(i)  $K(u, v) = \overline{K(v, u)}$  where “ $\overline{\quad}$ ” denotes complex conjugate

(ii) for all  $\alpha \in \mathcal{M}(\mathcal{U})$ .

$$(2.3) \quad Q(\alpha) = \int_{\mathcal{U}} \int_{\mathcal{U}} K(u, v) \alpha(du) \alpha(dv) \geq 0.$$

The quadratic form  $Q$  defined by such a kernel  $K$  defines a pseudo-norm

$$\|\alpha\|^2 = Q(\alpha)$$

and the pseudo-metric

$$(2.4) \quad d(\alpha, \beta) = \|\alpha - \beta\|, \alpha, \beta \in \mathcal{M}(\mathcal{U}).$$

In particular, when  $\alpha = P_1$  and  $\beta = P_2$  are p.m.'s,  $\|P_1 - P_2\|$  is a "distance" between  $P_1$  and  $P_2$ .

For our purposes, the following example is of special interest.

**Example 2.1** Consider  $\mathcal{U} = R^m$  and for each  $t \in R^m$ , let

$$(2.5) \quad k_t(u) = \exp[it'u], u \in R^m.$$

For any probability distribution  $H$  on  $R^m$ , it is easy to check that

$$(2.6) \quad K(u, v) = \int_{R^m} k_t(u) \overline{k_t(v)} H(dt)$$

is n.n.d. Obviously  $K(u, v)$  is just the characteristic function of  $H$  evaluated at  $u - v$ . Furthermore, kernels of the form (2.6) define "translation invariant" distances between p.m.'s on  $R^n$ , as the following discussion shows.

Given a p.m.  $P$  on  $R^m$  and a vector  $g \in R^m$ , define the p.m.  $gP$  by

$$(2.7) \quad (gP)(B) = P(B - g)$$

so  $gP$  is a "translate" of  $P$ .

**Lemma 2.1** When the kernel  $K$  in (2.6) is used to define the pseudo-norm  $\|\cdot\|$ , then

$$(2.8) \quad \|P_1 - P_2\| = \|gP_1 - gP_2\|, g \in R^m$$

for p.m.'s  $P_1$  and  $P_2$ .

**Proof:** This follows easily from the relation

$$K(u, v) = K(u + g, v + g). \quad \square$$

When  $K$  is given by (2.6), it is easily verified that

$$\|P_1 - P_2\|^2 = \int_{R^m} |\phi_1(t) - \phi_2(t)|^2 H(dt)$$

where  $\phi_i$  is the characteristic function of  $P_i$ ,  $i = 1, 2$ . This completes Example 2.1.  $\square$

Now, return to the general case of an n.n.d. kernel  $K$  and its resulting pseudo-norm  $\|\cdot\|$  on  $\mathcal{M}(\mathcal{U})$ . Consider a family of p.m.'s  $\{P(\cdot|\theta)|\theta \in \Theta\}$  defined on  $(\mathcal{U}, \mathcal{B})$  where  $(\Theta, \zeta)$  is a measurable space and  $P(B|\cdot)$  is a measurable function for each  $B \in \mathcal{B}$ .

**Lemma 2.2** For each p.m.  $\pi$  on  $(\Theta, \zeta)$ , let

$$(2.9) \quad \hat{P}_\pi = \int_{\Theta} P(\cdot|\theta) \pi(d\theta).$$

For each p.m.  $P$ ,

$$(2.10) \quad \begin{aligned} \int \|P - P(\cdot|\theta)\|^2 \pi(d\theta) &\geq \\ \int \|\hat{P}_\pi - P(\cdot|\theta)\|^2 \pi(d\theta) &= \\ \int \|P(\cdot|\theta)\|^2 \pi(d\theta) - \|\hat{P}_\pi\|^2. \end{aligned}$$

**Proof:** The p.m.  $\hat{P}_\pi$  is the mean of the family  $P(\cdot|\theta)$  when  $\theta$  has the distribution  $\pi$ . Relation (2.10) simply asserts that

- (i) the mean minimizes mean squared error, and
- (ii) “variance” equals the second moment minus the square of the mean.

The rigorous verification of (2.10) is little more than the proof when  $U$  is a finite set. The details are omitted.  $\square$

### 3 Design and Prediction

Consider again the decision theoretic description of the design problem described in Section 1. The data available is a random vector  $Y \in R^n$  whose distribution is  $P(\cdot|\theta, x)$  where  $\theta$  is an unknown parameter and  $x \in \mathcal{X}$  is a design variable.

In order to employ the design criterion introduced in Section 1, we now formulate a prediction problem. Suppose  $Z$  is a random vector in  $R^m$  to be predicted. It is assumed that given  $\theta$  and the design  $x$ ,  $Z$  has a distribution  $S(\cdot|\theta, x)$  and  $Z$  is conditionally (given  $\theta, x$ ) independent of  $Y$ . The action space for this problem is taken to be  $M_1(R^m)$ —the set of all probability distributions on  $R^m$ . Thus, given a design  $x$  and an observation  $Y = y$ , a decision rule  $\delta(\cdot|y, x)$  is a probability distribution on  $R^m$ —namely,  $\delta(\cdot|y, x)$  is our guess at the distribution of  $Z$  when the design is  $x$  and  $Y = y$  is observed.

To introduce a loss structure for this problem, let  $\|\cdot\|$  be a pseudo-norm on  $\mathcal{M}(R^m)$  defined by an n.n.d. kernel as described in Section 2. Consider a loss function given by

$$(3.1) \quad L(\alpha, \theta, x) = \|\alpha - S(\cdot|\theta, x)\|^2$$

where  $\alpha \in M_1(R^m)$ . An action  $\alpha$  is a p.m. on  $R^m$  and the loss for action  $\alpha$  is the “distance” squared of  $\alpha$  from the true distribution of  $Z$  when the parameter value is  $\theta$  and the design is  $x$ —namely,  $S(\cdot|\theta, x)$ .

Now suppose  $\pi$  is a proper prior distribution for  $\theta$ . To derive a Bayes rule for our decision problem when the prior is  $\pi$ , we proceed as follows. The model assumptions for  $Y$  and  $Z$  together with the prior  $\pi$  determine a joint distribution for  $Y, Z$  and  $\theta$ . (Assume the design used is  $x \in \mathcal{X}$ .) From this joint distribution, we can determine the conditional distribution of  $Z$  given  $Y = y$  when the design is  $x$ —say

$$(3.2) \quad Q_\pi(\cdot|y, x).$$

**Theorem 3.1** *Given the prior  $\pi$  and the design  $x$ , the conditional distribution  $Q_\pi(\cdot|y, x)$  for  $Z$  given  $Y = y$  is a Bayes solution to the prediction problem for any loss function given by (3.1).*

**Proof:** A Bayes solution to the prediction problem is found by minimizing the posterior expected loss. This is

$$(3.3) \quad \psi_0(\alpha) = \int_{\Theta} \|\alpha - S(\cdot|\theta, x)\|^2 \tilde{Q}_\pi(d\theta|y, x)$$

where  $\tilde{Q}_\pi(\cdot|y, x)$  denotes the conditional distribution of  $\theta$  given  $Y = y$  and the design  $x$ . A direct application of Lemma 2.3 shows that  $\psi_0(\alpha)$  is minimized by

$$(3.4) \quad \hat{\alpha}_\pi(\cdot|y, x) = \int_{\Theta} S(\cdot|\theta, x) \tilde{Q}_\pi(d\theta|y, x).$$

A moment's reflection shows that (3.4) is just the conditional distribution (3.2).  $\square$

**Remark 3.1** There are other interesting loss functions in the prediction problem for which  $Q_\pi(\cdot|y, x)$  is also a Bayes solution to the prediction problem. For example, if the loss function is

$$L_1(\alpha, z) = \|\alpha - \epsilon_z\|^2$$

where  $\|\cdot\|^2$  is a non-negative definite quadratic form and  $\epsilon_z$  is the p.m. with mass 1 at  $z$ , then  $Q_\pi(\cdot|y, x)$  remains a Bayes solution to the prediction problem. See Eaton (1986) for the details of this argument and other examples of loss functions for which  $Q_\pi(\cdot|y, x)$  is a Bayes solution.  $\square$

**Theorem 3.2** *When the prior is  $\pi$  and the design is  $x$ , let  $M_\pi(\cdot|x)$  denote the marginal distribution of  $Y$ . A  $\pi$ -optimal design for the prediction problem is found by minimizing the Bayes risk*

$$(3.5) \quad r_\pi(x) = \int_{\Theta} \|S(\cdot|\theta, x)\|^2 \pi(d\theta) - \int_{R^n} \|Q_\pi(\cdot|y, x)\|^2 M_\pi(dy|x).$$

**Proof:** A direct application of the equality portion of equation (2.10) and the definition of the Bayes risk gives

$$(3.6) \quad \begin{aligned} r_\pi(x) &= \int_{R^n} \int_{\Theta} \|Q_\pi(\cdot|y, x) - S(\cdot|\theta, x)\|^2 \tilde{Q}_\pi(d\theta|y, x) M_\pi(dy|x) \\ &= \int_{R^n} \int_{\Theta} \|S(\cdot|\theta, x)\|^2 \tilde{Q}_\pi(d\theta|y, x) M_\pi(dy|x) \\ &\quad - \int_{R^n} \|Q_\pi(\cdot|y, x)\|^2 M_\pi(dy|x) \end{aligned}$$

which is just (3.5).  $\square$

In some situations, the first term in the final expression for  $r_\pi(x)$  in (3.6) does not depend on  $x$ . In this case, the criterion for finding a  $\pi$ -optimal design for the prediction problem is the maximization of

$$(3.7) \quad s_\pi(x) = \int_{R^n} \|Q_\pi(\cdot|y, x)\|^2 M_\pi(dy|x).$$

This will be the case in the normal regression example considered in the next section.

**Remark 3.2** In the prediction literature (see e.g. Aitchison and Dunsmore (1975)),  $Q_\pi(\cdot|y, x)$  is called the predictive distribution of  $Z$  given  $Y = y$ . On a priori grounds,  $Q_\pi$  is the Bayes solution to the prediction problem since it incorporates all we know about  $Z$  after seeing  $Y = y$  (given the model assumptions and the priori). The decision theoretic formulation of the prediction problem provides another method for finding a Bayes solution—namely one

minimizes the integrated risk. From the point of view of inferential consistency, it seems essential to require that the two arguments yield the same solution to the prediction problem. It is this desired consistency (established in Theorem 3.1) that prompted our formulation of the prediction problem, and in particular, the choice of loss function (3.1). Loss functions which yield such consistencies were introduced in Eaton (1982)—called *Fair Bayes Loss Functions*, and are obviously related to proper scoring rules (see Savage (1971) for example). Further discussion and related issues can be found in Eaton (1982, 1992).

## 4 A Normal Regression Example

### 4.1 Basic Theory

In this section, we analyze the Bayesian design problem for a normal regression model with known covariance. It is assumed that the data vector  $Y \in R^n$  satisfies

$$(4.1) \quad Y = X\theta + \epsilon$$

where  $X$  is an  $n \times p$  design matrix of rank  $p$ , the parameter vector  $\theta$  is in  $R^p$  and  $\epsilon$  has an  $N_n(0, \Sigma_1)$  where  $\Sigma_1$  is a full rank known covariance matrix. Therefore, the conditional distribution of  $Y$  given  $\theta$  is

$$(4.2) \quad \mathcal{L}(Y|\theta) = N_n(X\theta, \Sigma_1).$$

For this problem, the set of designs is just the set of all possible  $X$  matrices under consideration. In the calculations that follow,  $X$  remains fixed. Particular design sets are introduced in the next section where some special linear models are considered.

**Remark 4.1** The rows of the matrix  $X$  will typically be row vectors of the form  $f(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_p(\mathbf{x}))$ , where  $\mathbf{x}$  is a point in a suitable design space,  $\mathcal{S}$  say, usually a subset of  $R^k$ , and the  $f_j$  are known functions. A design is given by a finite collection of such points  $\mathbf{x}_i$  (called the support points), with respective multiplicities  $n_i$  such that  $\sum_i n_i = n$ . When observations are uncorrelated (i.e.  $\Sigma_1 = \sigma_1^2 I$ ) it is common practice not to restrict the  $n_i$ 's to be integers, but to approximate them by non-negative real numbers. The non-integer design can then be rounded to an integer design in some systematic way (Pilz, 1991, page 183).  $\square$

The variable  $Z \in R^m$  to be predicted is assumed to have a normal distribution

$$(4.3) \quad \mathcal{L}(Z|\theta) = N_m(T\theta, \Sigma_2)$$

where  $T$  is a known  $m \times p$  matrix,  $\theta$  is the vector of unknown parameters and  $\Sigma_2$  is a known  $m \times m$  covariance matrix (possibly singular). The prior distribution for  $\theta$  is assumed to be

$$(4.4) \quad \pi = \mathcal{L}(\theta) = N_p(\theta_0, \Sigma_3)$$

where  $\theta_0$  is known and  $\Sigma_3$  is a known  $p \times p$  non-singular covariance matrix. The variables  $Y$  and  $Z$  are assumed to be conditionally independent, given  $\theta$ .



Given the model assumptions (4.2) and (4.3) and the prior distribution (4.4), the conditional distribution of  $Z$  given  $Y = y$  (and the design  $X$ ) is needed. To this end, let  $\Sigma$  be the  $(n + m) \times (n + m)$  matrix

$$(4.5) \quad \Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} + \begin{pmatrix} X \\ T \end{pmatrix} \Sigma_3 \begin{pmatrix} X \\ T \end{pmatrix}'.$$

Next partition  $\Sigma$  as

$$(4.6) \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where

$$(4.7) \quad \begin{cases} \Sigma_{11} = \Sigma_1 + X\Sigma_3X' \\ \Sigma_{22} = \Sigma_2 + T\Sigma_3T' \\ \Sigma_{12} = X\Sigma_3T' = \Sigma'_{21}. \end{cases}$$

**Lemma 4.1** *The conditional distribution of  $Z$  given  $Y = y$  is  $m$ -dimensional multivariate normal with mean vector  $\nu(y)$  and covariance matrix  $\Sigma_{22.1}$  where*

$$(4.8) \quad \begin{cases} \nu(y) = T\theta_0 + \Sigma_{21}\Sigma_{11}^{-1}(y - X\theta_0) \\ \Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}. \end{cases}$$

**Proof:** This is a standard multivariate calculation whose details are omitted.  $\square$

**Lemma 4.2** *In terms of  $\Sigma_1, \Sigma_2, \Sigma_3, X$  and  $T$ , the conditional covariance matrix of  $Z$  given  $Y = y$  is*

$$(4.9) \quad \Sigma_{22.1} = \Sigma_2 + T(\Sigma_3^{-1} + X'\Sigma_1^{-1}X)^{-1}T'.$$

**Proof:** Routine matrix calculations using (4.7) and (4.8) establish (4.9).  $\square$

Finally, the loss function for the prediction problem is assumed to be of the translation invariant type discussed in Example 2.1. In other words, the pseudo-norm  $\|\cdot\|$  defined on  $\mathcal{M}(R^m)$  is generated by a kernel  $K$  of the form

$$(4.10) \quad K(u - v) = \int_{R^m} \exp[it'(u - v)]H(dt)$$

where  $H$  is some fixed distribution on  $R^m$ . For  $\xi \in \mathcal{M}(R^m)$ ,

$$(4.11) \quad \|\xi\|^2 = \int \int K(u - v)\xi(du)\xi(dv)$$

which in turn defines the loss function for the prediction problem via (3.1). With this loss function, it is now possible to describe explicitly the optimal design criteria.

**Theorem 4.1** *Consider the normal regression model and the prediction problem described above. When the loss function is specified by (3.1) via the pseudo-norm (4.11), the Bayesian optimal design is that  $X_0$  in the design space which maximizes*

$$(4.12) \quad C(X) = \int_{R^m} \exp[-t'\Sigma_{22.1}t]H(dt).$$

Here,  $\Sigma_{22.1}$  (as a function of  $X$ ) is given in (4.9) and  $H$  is the distribution which defines  $\|\cdot\|$ .

**Proof:** Since the conditional distribution of  $Z$  given  $\theta$  does not depend on the design, an optimal design is found by maximizing  $s_\pi$  in (3.7). From Lemma 4.1, the conditional distribution of  $Z$  given  $Y = y$  is

$$(4.13) \quad Q_\pi(\cdot|y, X) = N_m(\nu(y), \Sigma_{22.1}).$$

Because  $\|\cdot\|$  in (4.11) is translation invariant,

$$(4.14) \quad \|Q_\pi(\cdot|y, X)\|^2 = \|N_m(\nu(y), \Sigma_{22.1})\|^2 = \|N_m(0, \Sigma_{22.1})\|^2 = C(X).$$

The final equality is established from Fubini's Theorem and the form of the kernel in (4.10) which defined  $\|\cdot\|$ . Since  $C(X)$  does not depend on  $y$ , the integral in (3.7) defining  $s_\pi$  is equal to  $C(X)$ . Thus, the Bayesian optimal design is the one which maximizes  $C(X)$ .  $\square$

A particularly convenient choice for  $H$  in (4.12) corresponds to a  $N_m(0, \Sigma_4)$  distribution where  $\Sigma_4$  is a known non-singular covariance matrix.

**Theorem 4.2** *When  $H$  in (4.12) is the  $N(0, \Sigma_4)$  distribution, the Bayesian optimal design is obtained by minimizing (over  $X$ 's)*

$$(4.15) \quad B(X) = \det[\frac{1}{2}\Sigma_4^{-1} + \Sigma_2 + T(\Sigma_3^{-1} + X'\Sigma_1^{-1}X)^{-1}T']$$

**Proof:** For  $H$  equal to the  $N(0, \Sigma_4)$  distribution, the function  $C(X)$  in (4.12) is given by

$$(4.16) \quad C(X) = [\det(\Sigma_4) \det(\Sigma_4^{-1} + 2\Sigma_{22.1})]^{-1/2}.$$

From equation (4.9) for  $\Sigma_{22.1}$ , maximizing (4.16) is clearly equivalent to minimizing (4.15).  $\square$

For the remainder of this section, we focus on the criterion function (4.15) for various special cases.

**Case 1** By taking  $p = m, T = I_p$  and  $\Sigma_2 = 0$ , we see that  $Z = \theta$  so the prediction problem becomes an estimation problem. In this case, the Bayesian optimal design is found by minimizing

$$(4.17) \quad B_e(X) = \det[\frac{1}{2}\Sigma_4^{-1} + (\Sigma_3^{-1} + X'\Sigma_1^{-1}X)^{-1}].$$

Formally setting  $\Sigma_4^{-1} = \Sigma_3^{-1} = 0$ , (4.17) yields the classical  $D$ -optimality criterion in a linear regression problem.

**Case 2** By taking  $m = 1$  and  $\Sigma_3^{-1} = 0$ , the criterion becomes the minimization of

$$(4.18) \quad B_T(X) = T(X'\Sigma_1^{-1}X)^{-1}T' \in R^1.$$

This is the classical  $c$ -optimality criterion (with  $T$  playing the role of  $c$ ).

**Remark 4.2** With  $M = X'\Sigma_1^{-1}X$  (the so-called moment matrix of the design), the function

$$(4.19) \quad \Psi(M) = \det[\frac{1}{2}\Sigma_4^{-1} + \Sigma_2^{-1} + T(\Sigma_3^{-1} + M)^{-1}T']$$

is just the criterion function (4.15). In Appendix I, the convexity of  $\Psi$  and some related results are established.  $\square$

**Remark 4.3** The formal device of setting  $\Sigma_4^{-1}$  and/or  $\Sigma_3^{-1}$  to 0 is just that—a formal device. Of course, the criteria obtained in Cases 1 and 2 are continuous functions of  $\Sigma_3^{-1}$  and  $\Sigma_4^{-1}$  so one can certainly argue that (4.17) and (4.18) are limits of criteria obtained by proper Bayes arguments. However, it appears to not be possible to take either  $H(dt)$  or the prior for  $\theta$  to be “flat” (i.e. Lebesgue measure) at the outset, since some of the relevant expressions are not well defined.  $\square$

**Remark 4.4** The classical  $D$ -optimality criterion is invariant under non-singular transformations. However our criterion of minimizing (4.15) is not invariant since it involves a prior distribution (via  $\Sigma_3$ ). Of course, non-trivial criteria involving proper prior distributions cannot be invariant under all non-singular transformations since prior assessments are necessarily coordinate dependent.  $\square$

## 4.2 Moment Matrix Considerations

We now return to a discussion of (4.19). Observe that in working with  $\Psi(M)$  there is no need to invert the matrix  $\Sigma_3^{-1} + X'\Sigma_1^{-1}X$  since, writing for simplicity  $W = \frac{1}{2}\Sigma_4^{-1} + \Sigma_2$  and  $R = \Sigma_3^{-1}$ ,

$$(4.20) \quad \begin{aligned} \Psi(M) &= \det(W + T(M + R)^{-1}T') \\ &= \det(W) \det(M + R + T'W^{-1}T) / \det(M + R) \end{aligned}$$

so minimization of  $\Psi(M)$  means finding

$$(4.21) \quad \min_{M \in \mathbf{M}} \det(M + R + T'W^{-1}T) / \det(M + R).$$

where  $\mathbf{M}$  indicates the set of all moment matrices.

The optimization problem is in general too hard to be solved directly, but the search for optimal designs can be simplified using properties of  $\Psi$ . In Appendix I the convexity of  $\Psi$  in  $M$  and some related results are established. This enables us to apply the well-known theory of convex design criteria (see Pukelsheim, 1993), and in particular the Equivalence Theorem (Whittle, 1973).

**Theorem 4.3** *Assume  $\mathbf{M}$  is convex. For  $M_1, M_2 \in \mathbf{M}$  let  $F_\Psi(M_1, M_2)$  be the directional derivative of  $\Psi$  at  $M_1$  in the direction of  $M_2$ . The following conditions are equivalent:*

$$(i) \quad \Psi(M^*) = \min_{M \in \mathbf{M}} \Psi(M);$$

$$(ii) \quad \inf_{M \in \mathbf{M}_c} F_\Psi(M^*, M) = 0.$$

where  $M_e$  denotes the extreme points of the set  $M$ .

**Proof:** This is a consequence of Corollary A.2 and Theorem A.2 of Appendix I and the Equivalence Theorem of Whittle (1973).  $\square$

Condition (ii) above simplifies when  $\Sigma_1 = \sigma_1^2 I$  since the extreme points of  $M$  are the moment matrices of one point designs, i.e.  $M = n\mathbf{f}(\mathbf{x})'\mathbf{f}(\mathbf{x})$ . Then (ii) of Theorem 4.5 becomes

$$(4.22) \quad \text{If } \Sigma_1 = \sigma_1^2 I \text{ the points of support of an optimal design with moment matrix } M^* \text{ are among those points } \mathbf{x} \text{ for which } F_\Psi[M^*, \mathbf{f}(\mathbf{x})'\mathbf{f}(\mathbf{x})] = 0.$$

This is a useful tool for finding the design since it restricts the choice of potential design points in the design region. Observe that

$$(4.23) \quad \begin{aligned} F_\Psi[M, \mathbf{f}(\mathbf{x})'\mathbf{f}(\mathbf{x})] &= \det(W + T(M + R)^{-1}T') \\ &\times \text{trace}[(M + R)^{-1}T'(W + T(M + R)^{-1}T')^{-1} \\ &T(M + R)^{-1}(M - \mathbf{f}(\mathbf{x})'\mathbf{f}(\mathbf{x}))]. \end{aligned}$$

Together with convexity, invariance w.r.t. a group  $\mathcal{G}$  of transformations of the design space  $\mathcal{S}$  can be used to find optimal designs. The group  $\mathcal{G}$  induces a transformation of designs over  $\mathcal{S}$  and consequently of the moment matrices:  $\forall g \in \mathcal{G}$  denote by  $M_g$  the transformed matrix. We shall be interested in cases for which the following is true.

$$(4.24) \quad \forall g \in \mathcal{G} \exists \text{ a non-singular } p \times p \text{ matrix } Q_g \text{ such that } M_g = Q_g' M Q_g \forall M.$$

A typical case of (4.24) is with  $M_g$  a permutation of the rows and columns of  $M$ . Sometimes also the following can be assumed

$$(4.25) \quad Q_g' R Q_g = R \text{ and } Q_g' T' W^{-1} T Q_g = T' W^{-1} T \forall g \in \mathcal{G}.$$

Then

**Theorem 4.4** *If there is a group  $\mathcal{G}$  such that (4.24) and (4.25) hold, the search for an optimal design can be restricted to designs which are invariant w.r.t. that group.*

**Proof:** (4.20), (4.24) and (4.25) imply that  $\Psi$  is  $\mathcal{G}$ -invariant i.e.  $\Psi(M_g) = \Psi(M) \forall g \in \mathcal{G}, \forall M$ . The result follows from combining invariance with convexity; the idea is essentially Kiefer's (1959, §2E): see also Giovagnoli, Pukelsheim and Wynn (1987) for more detailed applications of those ideas to designs.  $\square$

In the next section these results will be applied to finding designs for particular regression problems.

## 5 Some special linear models

**Case 1: Multiple Linear Regression** Assume a normal model satisfying (4.2) with

$$(5.1) \quad E(Y_x | \theta) = \theta_0 + \theta_1 x_1 + \dots + \theta_k x_k$$

where  $\mathbf{x} = (x_1, \dots, x_k)$  is a point in a suitable design space, usually assumed to be a closed rectangle of  $R^k$ . It is easy to see

**Theorem 5.1** *If  $\Sigma_1 = \sigma_1^2 I$  there is an optimal design for (5.1) whose support points are among the extreme points of the design region.*

**Proof:** The rows of the matrix  $X$  are row vectors of the form  $\mathbf{f}(\mathbf{x}) = (1, x_1, \dots, x_k) = (1, \mathbf{x})$ . Simple algebra shows that (4.23) can be rewritten as:

$$(5.2) \quad F_\Psi[M, \mathbf{f}(\mathbf{x})' \mathbf{f}(\mathbf{x})] = c_1 - c_2(1, \mathbf{x})[(M + R)^{-1} T'(W + T(M + R)^{-1} T')^{-1} T(M + R)^{-1}](1, \mathbf{x})'$$

where  $c_1, c_2$  do not depend on  $\mathbf{x}$ , and this is a concave function of  $\mathbf{x}$  (see Marshall and Olkin, 1979, Ch.16, E.7.a). The proof follows from (4.22).  $\square$

If the design region is a hyperrectangle, w.l.o.g. we can take it to be the cube  $[-1, 1]^k$ , since a non-singular linear transformation  $\tilde{\theta} = B^{-1}\theta$  of the parameter space leads to:

$$\begin{aligned} \tilde{X} &= XB \\ \tilde{T} &= TB \\ \tilde{M} &= B'MB \\ \tilde{R} &= B'RB \end{aligned}$$

and

$$W + T(M + R)^{-1} T' = W + \tilde{T}(\tilde{M} + \tilde{R})^{-1} \tilde{T}'.$$

Furthermore observe that for estimating the parameters we can let  $\tilde{T} = I$  w.l.o.g. by changing the  $\Sigma_4$  matrix to  $B'\Sigma_4 B$ , since  $\det(\cdot)$  is invariant w.r.t. non-singular linear transformations.

The extreme points of the cube  $[-1, 1]^k$  are the  $2^k$  vertices, which give rise to the row vectors  $\mathbf{v}_i = (1, \pm 1, \pm 1, \dots, \pm 1)$  for the  $X$  matrix. The problem reduces to calculating the optimal number of observations  $n_i$  ( $n_i \geq 0$ ) to take at each vertex. Let  $M = \sum_i n_i \mathbf{v}_i' \mathbf{v}_i / \sigma_1^2$  and write for simplicity  $\sigma_1^2 R = \tilde{R} = (r_{ij})$  and  $\sigma_1^2 T' W^{-1} T = A = (a_{ij})$ ,  $i, j = 0, \dots, k$ . Minimization of  $\Psi(M)$  in this case becomes

$$(5.3) \quad \begin{aligned} \min_{\{n_i\}} \quad & \det\left(\sum_i n_i \mathbf{v}_i' \mathbf{v}_i + \tilde{R} + A\right) / \det\left(\sum_i n_i \mathbf{v}_i' \mathbf{v}_i + \tilde{R}\right) \\ \text{subject to} \quad & \sum_i n_i = n \\ & n_i \geq 0. \end{aligned}$$

An explicit solution can be found in the following case:

**Example 5.1** (Simple linear regression): Assume (4.2) with  $\Sigma_1 = \sigma_1^2 I$  and

$$(5.4) \quad E(Y_x | \theta) = \theta_0 + x\theta_1.$$

By theorem 5.1 the support points of the optimal design are -1 and 1. Let  $n_1$  and  $n_2$  be the number of observations at -1 and 1, and put  $c_{ii} = n + r_{ii}$ ,  $i = 0, 1$ ,  $b = c_{00}a_{11} + c_{11}a_{00} + (a_{00}a_{11} - a_{01}^2)$  and  $\Delta = b^2 - 4a_{01}^2 c_{00}c_{11}$ . It is easy to show that

**Theorem 5.2** *An optimal design for model (5.4) is obtained by taking  $n_1$  observations at -1 where  $n_1 = \text{median}\{0, (n - \beta^* + r_{01})/2, n\}$  and*

$$\beta^* = \begin{cases} (b - \sqrt{\Delta}) / (2a_{01}) & \text{if } a_{01} \neq 0 \\ 0 & \text{if } a_{01} = 0 \end{cases}$$

**Proof:** Let  $\beta = n + r_{01} - 2n_1$ ; (5.3) reduces to

$$\begin{aligned} & \min_{\beta} && (b - 2\beta a_{01}) / (c_{00}c_{11} - \beta^2) \\ & \text{subject to} && -n + r_{01} \leq \beta \leq n + r_{01} \end{aligned}$$

whose solution is  $\text{median}\{-n + r_{01}, \beta^*, n + r_{01}\}$ .  $\square$

Theorem 5.2 shows that for simple linear regression one point designs at either +1 or -1 are possible optimal designs. For instance, if  $a_{01} < 0$  and  $r_{01} \geq n$  then an optimal design puts  $n$  observations at -1, if  $a_{01} > 0$  and  $r_{01} \leq -n$  then an optimal design puts  $n$  observations at +1. If  $A$  is diagonal then  $n_1 = \text{median}\{0; (n + r_{01})/2; n\}$  which turns out to be the Bayes  $A$ -,  $D$ -, and  $E$ -optimal design (Pilz 1991, page 173). This happens for instance for estimation of  $\theta$  with  $\Sigma_4$  diagonal.

If we want to predict  $Z$  at  $x_0$ ,  $T = (1, x_0)$  and the solution is the Bayes  $c$ -optimal design which does not depend on  $\Sigma_4$  and  $\Sigma_2$ . Simple algebra shows that the design depends on  $\tilde{R}$  and on  $x_0$  as follows:

(i) if  $|x_0| \leq \sqrt{c_{11}/c_{00}}$  then  $n_1 = \text{median}\{0, (n + r_{01} - x_0 c_{00})/2, n\}$ ;

(ii) if  $|x_0| > \sqrt{c_{11}/c_{00}}$  then  $n_1 = \text{median}\{0, (n + r_{01} - c_{11}/x_0)/2, n\}$

Observe that if  $r_{00} = r_{11}$  then (i) and (ii) correspond respectively to interpolation and extrapolation (see Chaloner, 1984).

There does not seem to be a simple way of obtaining a general solution of (5.3), but if both the prior covariance matrix of the parameters and the matrix  $A$  have a high degree of symmetry then an explicit solution can be found as follows:

**Theorem 5.3** *If both  $A$  and  $\tilde{R}$  are invariant w.r.t. the group of all the permutations and sign changes which leaves the first row and column fixed then a design that solves (5.3) is the one with the same number of observations  $n/2^k$  at each point  $v_i$ ;  $i = 1, \dots, 2^k$ .*

**Proof:** Taking the group of all the permutations and sign changes of  $\{x_1, \dots, x_k\}$  as the group  $\mathcal{G}$  acting on the design space, conditions (4.24) and (4.25) are satisfied with  $Q_g$  a block diagonal matrix  $Q_g = \text{diag}(1, \Pi)$  and  $\Pi$  a  $\pm 1$ -permutation matrix. Thus Theorem 4.6 can be applied. An optimal design is the unique one invariant under all permutations and sign changes of the coordinates  $x_j$ 's.  $\square$

This ends Example 5.1.  $\square$

Note that the optimal design found in Theorem 5.3 is the classical  $D$ -optimal design for multiple linear regression. Two interesting cases covered by Theorem 5.3 are the following:

**Example 5.2** (Estimation of  $\theta$ :  $m = k + 1$ ,  $\Sigma_2 = O$ ,  $T = I_{k+1}$ ). If

- (i) the parameters  $\theta_0, \theta_1, \dots, \theta_k$  are a priori independent with  $\tilde{R} = \text{diag}(r_{00}, r_{11}, \dots, r_{11})$ , and

(ii)  $\Sigma_4 = \text{diag}(s_{00}, s_{11}, \dots, s_{11})$

then it is straightforward to show that the hypotheses of Theorem 5.3 hold with  $A = 2\sigma_1^2 \Sigma_4$  so that the design of Theorem 5.3 is optimal.  $\square$

**Example 5.3** Assume that the aim of the experiment is the prediction of an observation  $Z$  at the centre of the design region, i.e.  $T = (1, 0, 0, \dots, 0)$ . If (i) of Example 5.2 holds, then the same design of Theorem 5.3 is optimal.  $\square$

**Case 2: One-way analysis of variance:** Assume a normal linear model satisfying (4.2) with

$$(5.5) \quad E(Y_{ij}|\theta) = \theta_i, i = 0, \dots, k, j = 1, \dots, n_i$$

and  $\Sigma_1 = \sigma_1^2 I$  as in case 1. The design problem is finding the optimal number of observations  $n_i$  for each treatment  $\theta_i$ . Minimization of  $\Psi(M)$  becomes

$$(5.6) \quad \min_{\{n_i\}} \det(\text{diag}(n_0, \dots, n_k) + \tilde{R} + A) / \det(\text{diag}(n_0, \dots, n_k) + \tilde{R})$$

subject to  $\sum_i n_i = n$   
 $n_i \geq 0$ .

If  $\tilde{R}$  and  $A$  have a special structure, (5.6) can be easily solved. Note that the problems of estimation and prediction of the treatment effects are essentially equivalent: in both cases  $T = I$ .

**Example 5.4 Estimation/prediction of the effects of exchangeable treatments** If the treatments are exchangeable then the matrix  $\tilde{R}$  is permutation invariant. If  $\Sigma_4$  is permutation invariant too, then by Theorem 4.6 we can restrict the search of the optimal design to designs invariant w.r.t. permutations of all the treatments and hence an optimal design is the one with the same number of observations  $n/(k+1)$  at each treatment. The solution is the classical “universally” optimal design (Kiefer 1959, §4).  $\square$

**Example 5.5** Assume

$$(5.7) \quad \tilde{R} = \text{diag}(r_{00}, r_{11}, \dots, r_{11}).$$

(5.7) will hold, for instance, if the experiment involves  $k$  new test treatments  $\theta_j$ ,  $j = 1, \dots, k$  which are a priori exchangeable and uncorrelated and a control  $\theta_0$ , which is assumed to be uncorrelated with the treatments. Furthermore, assume

$$(5.8) \quad \Sigma_4 = \text{diag}(s_{00}, s_{11}, \dots, s_{11}), \text{ and } \Sigma_2 = \sigma_2^2 I, \text{ where } \sigma_2^2 \text{ can be } 0.$$

$\tilde{R}$  and  $A$  are invariant under permutations which leave the first row and column fixed. Thus by Theorem 4.6 there exists an optimal design which is invariant under the group of permutations of the test treatments, and hence an optimal design will have  $n_0$  observations at  $\theta_0$  and  $n_1$  observations at each  $\theta_j$ ,  $j = 1, \dots, k$ .

Let

$$A = \text{diag} \left( \frac{2s_{00}\sigma_1^2}{2s_{00}\sigma_2^2 + 1}, \frac{2s_{11}\sigma_1^2}{2s_{11}\sigma_2^2 + 1} I \right) = \text{diag}(a_{00}, a_{11}, \dots, a_{11}).$$

Define  $t = \text{trace}(\tilde{R})$ ,  $b = a_{11}[2k(n+t) + a_{00}(k+1)]$  and  $\Delta = b^2 + 4a_{11}(a_{00} - k^2 a_{11})(n+t)(n+t + a_{00})$ .

**Theorem 5.4** If  $T = I$  and (5.7) and (5.8) hold, an optimal design for estimating the treatment effects has  $n_1 = \text{median}\{0, \beta^* - r_{11}, n/k\}$  at each  $\theta_j$ ,  $j = 1, \dots, k$  where

$$\beta^* = \begin{cases} (-b + \sqrt{\Delta})/[2(a_{00} - k^2 a_{11})] & \text{if } a_{00} \neq k^2 a_{11} \\ (n+t)[(n+t)+a]/[2k(n+t)+a(k+1)] & \text{if } a_{00} = k^2 a_{11} = a \end{cases}$$

**Proof:** Let  $\beta = n_1 + r_{11}$ . Then (5.6) reduces to

$$(5.9) \quad \min_{\beta} \quad [1 + a_{00}/(n+t-k\beta)][1 + a_{11}/\beta]^k \\ \text{subject to} \quad r_{11} \leq \beta \leq n/k + r_{11}$$

whose solution is  $\text{median}\{r_{11}, \beta^*, n/k + r_{11}\}$   $\square$

Note that if  $s_{00} = s_{11}$ , and thus we use a uniformly weighted distance, then  $n_1 = \text{median}\{0; (n+r_{00}-r_{11})/(k+1); n/k\}$  which is the Bayes  $D$ -optimal design for the estimation of  $\theta$  (Verdinelli, 1992).  $\square$

**Example 5.6** Assume  $\Sigma_4 = \sigma_4 I$  and that the aim of the experiment is the estimation of the differences between  $\theta_0$ , the control, and the other treatments, so that  $T = (-1 : I_k)$ . Then the matrix  $A$  is invariant under permutations which leave the first row and column fixed.

Assume also that  $\tilde{R}$  is invariant under the same permutation group, e.g. assume  $\theta_0$  and  $\theta_j$  uncorrelated  $\forall j$  and  $\theta_j$  ( $j = 1, \dots, k$ ) exchangeable, so that  $\tilde{R} = \text{diag}(r_{00}, (r_{11} - \rho)I + \rho J)$ . Then a similar argument to that of example 5.5 implies that an optimal design is the one which has  $n_0$  and  $n_1$  observations at  $\theta_0$  and at each  $\theta_j$ , where  $n_0$  and  $n_1$  can be found by minimizing (5.6) subject to  $n_0 + kn_1 = n$ .

If  $t = \text{trace}(\tilde{R})$ ,  $h = 2\sigma_1^2\sigma_4^2$  and  $\beta = n_1 + r_{11} - \rho$ , then (5.6) becomes

$$(5.10) \quad \min_{\beta} \quad (1 + h/\beta)^{k-1} [1 + h/(\beta + k\rho) + hk/(n+t-k\rho-k\beta)] \\ \text{subject to} \quad r_{11} - \rho \leq \beta \leq n/k + r_{11} - \rho$$

If  $k = 1$ , the criterion reduces to Bayes  $c$ -optimality and the solution is  $n_0 = \text{median}\{0, [n - (r_{00} - r_{11})]/2, n\}$ . The number of observations to put at  $\theta_0$  and  $\theta_1$  depends on the difference between their prior precisions: the greater the precision, the smaller the number of observations.  $\square$

**Example 5.7** Assume all the treatments are a priori uncorrelated, i.e.

$$(5.11) \quad \tilde{R} = (r_{00}, r_{11}, \dots, r_{kk})$$

and let  $\Sigma_4 = \sigma_4^2 I$ . The proof of the following theorem is straightforward.

**Theorem 5.5** An optimal design is found by

$$(5.12) \quad \min_{\{n_j\}} \quad \prod_{j=1}^p (1 + h/(n_j + r_{jj})) \\ \text{subject to} \quad \sum_j n_j = n \\ n_j \geq 0,$$



where  $h = 2\sigma_1^2\sigma_4^2/(\sigma_4^2\sigma_2^2 + 1)$ . If

$$(5.13) \quad \max_j r_{jj} < \frac{n + \text{trace}(\tilde{R})}{k + 1}$$

the solution is  $n_j = (n + \text{trace}(\tilde{R}))/(k + 1) - r_{jj}$ .

Note that in this case the optimal design does not depend on  $\Sigma_4$ , as expected, since we are assuming that  $\Sigma_4 = \sigma_4^2 I$  and hence we use a uniformly weighted distance. If (5.13) does not hold, then the solution of (5.12) is on the boundary of the constraint region so that a numerical search is needed.  $\square$

## Appendix

We denote by  $W(m \times m) \geq 0$  a non-negative definite matrix, and by  $\tilde{M}(p \times p) > 0$  a positive definite one. Matrix concavity of a matrix function  $\phi$  is defined by  $\phi(\lambda A_1 + (1 - \lambda)A_2) \geq_L (\lambda\phi(A_1) + (1 - \lambda)\phi(A_2))$  for all  $A_1, A_2$  and  $0 \leq \lambda \leq 1$  where  $\geq_L$  is the Loewner ordering of symmetric matrices, namely  $A \geq_L B \iff A - B \geq 0$ . Similarly for matrix convexity. Let  $T$  be an  $m \times p$  matrix of rank  $m$ . Define

$$(A.1) \quad \tilde{\Psi}(\tilde{M}) = (W + T\tilde{M}^{-1}T')^{-1}.$$

### A Convexity and a Monotonicity Result

**Theorem A.1** *If  $W(m \times m) \geq 0$ , then  $\tilde{\Psi}$  defined by (A.1) over a convex set of  $p \times p$  positive definite matrices, is a concave matrix function of  $\tilde{M}$*

**Proof:** This result generalizes the one by Olkin with  $W = O$  (see Marshall and Olkin, 1979, p.469) and the proof is essentially the same. For given  $\tilde{M}_1, \tilde{M}_2$ , and  $0 \leq \lambda \leq 1$  define

$$\tilde{M}(\lambda) = \lambda\tilde{M}_1 + (1 - \lambda)\tilde{M}_2$$

and

$$Q(\lambda) = W + T\tilde{M}(\lambda)^{-1}T'.$$

It is enough to show that  $Q(\lambda)^{-1}$  is a concave matrix function of  $\lambda$  for any given  $M_1$  and  $M_2$ , or equivalently that the second derivative is non-positive definite. Now

$$d^2Q^{-1}/d\lambda^2 = 2Q^{-1}(dQ/d\lambda)Q^{-1}(dQ/d\lambda)Q^{-1} - Q^{-1}(d^2Q/d\lambda^2)Q^{-1}$$

so that

$$d^2Q^{-1}/d\lambda^2 \leq 0$$

iff

$$2(dQ/d\lambda)Q^{-1}(dQ/d\lambda) - d^2Q^{-1}/d\lambda^2 \leq 0.$$

Furthermore

$$dQ/d\lambda = -T\tilde{M}^{-1}(d\tilde{M}/d\lambda)\tilde{M}^{-1}T'$$

and

$$d^2Q^{-1}/d\lambda^2 = 2T\tilde{M}^{-1}(d\tilde{M}/d\lambda)\tilde{M}^{-1}(d\tilde{M}/d\lambda)\tilde{M}^{-1}T'$$

since  $d^2\tilde{M}^{-1}/d\lambda^2 = O$ . Thus we want to show that

$$T\tilde{M}^{-1}(d\tilde{M}/d\lambda)[\tilde{M}^{-1}T'Q^{-1}T\tilde{M}^{-1} - \tilde{M}^{-1}](d\tilde{M}/d\lambda)\tilde{M}^{-1}T' \leq 0.$$

In order to prove the last statement, we put  $X = \tilde{M}^{-1/2}T'$  and show that

$$\tilde{M}^{-1}T'Q^{-1}T\tilde{M}^{-1} - \tilde{M}^{-1} \leq 0,$$

i.e.  $XQ^{-1}X' \leq_L I_p$ , with  $I_p$  the  $p \times p$  identity matrix. This is equivalent to

$$Q^{-1/2}X'XQ^{-1/2} \leq_L I_m$$

which is true since  $Q - X'X = Q - T\tilde{M}^{-1}T' = W \geq 0$ , thus completing the proof.  $\square$

**Corollary A.1** *Let*

$$(A.2) \quad W(m \times m) \geq 0, M(p \times p) \geq 0, R(p \times p) > 0.$$

*Then  $\tilde{\Psi}_1(M) = (W + T(M + R)^{-1}T')^{-1}$  is matrix-concave in  $M$ .*

**Theorem A.2** *Under assumptions (A.2)*

(i)  $\Psi(M) = \det(W + T(M + R)^{-1}T')$  *is a convex function of  $M$ .*

(ii)  $\Psi(M)$  *is decreasing w.r.t. the Loewner ordering, i.e.  $A \geq_L B \Rightarrow \Psi(A) \leq \Psi(B)$ .*

**Proof:** (i) Since the real function  $\log \det(\cdot)$  is concave and increasing w.r.t. the Loewner ordering, by Corollary A.1  $-\log \det(W + T(M + R)^{-1}T') = \log \det(W + T(M + R)^{-1}T')^{-1}$  is a concave function of  $M$ , thus  $\det(W + T(M + R)^{-1}T')$  is log-convex and hence convex in  $M$ .

The proof of (ii) follows from well-known facts about  $\det(\cdot)$  and the matrix inverse, see for instance Marshall and Olkin (1979), Ch.16 Sect.E.  $\square$

**Theorem A.3**  $\Psi(M)$  *is differentiable.*

**Proof:** Given any two moment matrices  $M_1$  and  $M_2$  the directional derivative of  $\Psi$  at  $M_1$  in the direction of  $M_2$  exists, because of convexity, and can be shown to be:

$$(A.3) \quad \begin{aligned} F_\Psi[M_1, M_2] &= \lim_{\epsilon \rightarrow 0} [\Psi(M_1 + \epsilon(M_2 - M_1)) - \Psi(M_1)]/\epsilon \\ &= \text{trace}[T(M_1 + R)^{-1}(M_1 - M_2)(M_1 + R)^{-1} \\ &\quad T'(W + T(M_1 + R)^{-1}T')^{-1}] \det(W + T(M_1 + R)^{-1}T') \end{aligned}$$

and thus it is linear in  $M_2$ .  $\square$

**Corollary A.2** *If  $M = M(\mathbf{v})$  is a linear function of the vector  $\mathbf{v}$ , then  $\Psi(M(\mathbf{v}))$  is convex in  $\mathbf{v}$ .*

**Proof:** This statement is straightforward.  $\square$

## References

- Aitchison J. and Dunsmore I.R. (1975). *Statistical Prediction Analysis*. Cambridge University Press.
- Brooks, R.J. (1972). A decision theory approach to optimal regression designs. *Biometrika*, 59 563–571.
- Brooks, R.J. (1974). On the choice of an experiment for prediction in linear regression. *Biometrika* 61 303–311.
- Brooks, R.J. (1976). Optimal regression designs for prediction when prior knowledge is available. *Metrika* 23 221–230.
- Chaloner, K. (1984). Optimal Bayesian experimental designs for linear models. *Ann. Statist.* 12 283–300.
- Eaton, M.L. (1982). A method for evaluating improper prior distribution in *Statistical Decision Theory and Related Topics*. (S.S.Gupta and J. Berger, eds.) Academic Press, 1, 329–352.
- Eaton, M.L. (1986). Admissibility in Fair Bayes Prediction Problems. University of Minnesota Technical Report No. 482.
- Eaton, M.L. (1992). A statistical diptych: admissible inferences—recurrence of symmetric Markov chains. *Ann. Statist.* 20 1147–1179.
- Fedorov, V.V. (1972). *Theory of Optimal Experiments*. Academic Press N.Y.
- Giovagnoli A., Pukelsheim F. and Wynn H.P. (1987). Group invariant orderings and experimental designs. *J. Statist. Planning Inference* 17 159–171.
- Kiefer, J. (1959). Optimum Experimental Designs (with discussion). *J. Roy. Statist. Soc. B.* 21 272–319.
- Lindley, D.V. (1968). The choice of variables in multiple regression. *J. Roy. Statist. Soc. B* 30 31–36.
- Marshall A. and Olkin I. (1979). *Inequalities: Theory of Majorization and its applications*. Academic Press.
- Piccinato, L. (1980). La teoria delle decisioni statistiche. *B.U.M.I.* 17-A 376–405.
- Pilz, J. (1991). *Bayesian estimation and Experimental Design in Linear Regression Models*. Wiley, New York.
- Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.
- Savage, L.J. (1971). The elicitation of personal probabilities and expectations. *J. Amer. Statist. Ass'n.* 66 783–801.

Silvey, S.D. (1980). *Optimal Designs*. Chapman and Hall.

Verdinelli, I. (1992). Advances in Bayesian experimental design (with discussion). In *Bayesian Statistics 4*, (J.M. Bernardo et al. eds), Oxford University Press. 467–482.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B.* **35** 123–130.