

## COHERENT DISPERSION CRITERIA FOR OPTIMAL EXPERIMENTAL DESIGN

BY A. PHILIP DAWID AND PAOLA SEBASTIANI<sup>1</sup>

*University College London and The Open University*

We characterize those coherent design criteria which depend only on the dispersion matrix (assumed proper and nonsingular) of the “state of nature,” which may be a parameter-vector or a set of future observables, and describe the associated decision problems. Connections are established with the classical approach to optimal design theory for the normal linear model, based on concave functions of the information matrix. Implications of the theory for more general models are also considered.

**1. Introduction.** There are two distinct statistical approaches to the problem of optimal experimental design. *Criterion-based design* [Kiefer (1959)] focuses on finding designs to optimize some appealing *optimality criterion*, usually defined as a suitable function of the Fisher information matrix of the design. In the context of linear models, the Fisher information matrix is proportional to the inverse of the dispersion matrix of the least squares estimators of the parameters. Bayesian criterion-based methods have also been developed, in which the information matrix is replaced by the inverse of the posterior dispersion matrix of the parameters [Pilz (1991)]. Particular attention is often given to design criteria which are concave nondecreasing functions of the information matrix. Nondecreasingness is fundamental, since we wish to maximize “information”; concavity is less fundamental, but is usually imposed in order that the criterion be amenable to convex optimization methods [Kiefer (1974)].

An alternative approach, *decision-based design*, builds on the general principles of Bayesian decision theory [see, e.g., Raiffa and Schlaifer (1961)]. For this, we need to set up a decision problem requiring us to choose an act after seeing the data of our experiment, and then delivering a loss which is specified as a function of the act chosen and the unknown state of nature. Given a full Bayesian specification of the problem, the general principle of acting so as to minimize the overall expected loss can then be applied to determine the optimal experiment (as well as the optimal act to take after seeing any data). In such decision problems, the “state of nature” is usually identified with the parameter-vector. In this case, we may speak of an *estima-*

---

Received May 1996; revised October 1998.

<sup>1</sup>Supported in part by the European Science Foundation and Nuffield Foundation Grant SCI/180/95/493/G.

AMS 1991 subject classifications. Primary 62K05; secondary 62C10.

Key words and phrases. Bayesian decision theory, coherence, concavity, dispersion standard, optimal design, optimality criterion, proper scoring rule, uncertainty function.

*tive decision problem.* However, it is more relevant to real world problems to consider instead the unknown value(s) of some future observation(s) on the process as determining the loss. In this case we have a *predictive decision problem.*

Decision-theoretic justifications of some classical and Bayesian optimality criteria are well known [Chaloner and Verdinelli (1995)]. However, there is currently no general theory which allows us to associate directly a criterion-based design problem with a decision problem. The purpose of this paper is to formalize the connections between the two approaches to experimental design, under the constraint that attention is confined to properties of the (estimative or predictive) dispersion matrix. (The possibility of designs having singular information matrix is allowed and causes no problems so long as we deal with Bayesian design based on a proper prior distribution; otherwise, it may be necessary to restrict attention to designs under which the quantities of interest are estimable.) In particular, we aim to identify when a dispersion-based optimality criterion is (*dispersion-*)*coherent*, in that optimizing it is equivalent to solving some Bayesian decision problem. Conversely, we shall give conditions for such a Bayesian decision problem to be soluble by means of criterion-based convex optimization methods. Of particular importance are criteria which can be simultaneously justified from both points of view and which we here characterize. In particular, while there are decision-theoretic justifications for the familiar optimality criteria of *D*-optimality and *A*-optimality, we show that (under the constraints we impose) there is none such for *E*-optimality.

Section 2 reviews the criterion-based approach, and its Bayesian analogue, as well as extensions to problems of prediction. In Section 3 we describe the general decision-based approach, leading to the choice of design being based on a “standard” constructed from a function describing the “uncertainty” in the relevant posterior or predictive distribution. As a first step in relating this to the criterion-based approach, with its emphasis on the information matrix, in Section 4 we characterize those coherent uncertainty functions which depend only on the dispersion matrix. Section 5 considers the corresponding “coherent dispersion standards” in the context of the normal linear model and their relationship with optimality criteria. In particular, we show that suitable criterion functions, whether applied to the classical information matrix, to its Bayesian counterpart, or to the analogous matrix for prediction problems, lead to design strategies which, simultaneously, are amenable to convex optimization, and correspond to the solution of a Bayesian decision problem. Finally, in Section 6 we give an asymptotic extension of the theory developed here to design criteria in more general models, and offer some concluding comments in Section 7.

**2. Criterion-based optimal experimental design.** In this section we consider classical and Bayesian criterion-based approaches to optimal design for the linear model.

2.1. *Classical optimal design.* We shall assume a linear model for the data  $Y$  given the vector of parameters  $\theta$  in  $\mathcal{R}^q$ , and the design matrix  $X$ ,

$$(1) \quad E(Y | \theta) = X\theta \quad \text{and} \quad \text{Var}(Y | \theta) = \sigma^2 I_n,$$

where  $\text{Var}$  denotes dispersion matrix. A typical row of  $X$  will be a vector of the form  $f(x) = (f_1(x), \dots, f_q(x))$ ,  $x \in \mathcal{X} \subset \mathcal{R}^k$ , where the  $f_i$  are known functions. The *information matrix* of the design is  $M_0 := X^T X / \sigma^2$ . When  $M_0$  is nonsingular, the dispersion matrix of the least squares estimator  $\hat{\theta}$  of  $\theta$  is  $V_0 = M_0^{-1}$ , the *classical estimative dispersion*.

A design  $\xi$  is a finite collection of design points  $x_i$  with multiplicities  $n_i$ . Then  $X, M_0, V_0$  are functions of  $\xi$ . We wish to choose  $\xi$  from some set  $\Xi$  (typically having fixed  $\sum_i n_i$ ) to maximize  $\Phi(\xi)$ , where  $\Phi: \Xi \rightarrow \mathcal{R}$  is a suitable *optimality criterion*. Let  $\mathcal{S}_q$  denote the set of  $q \times q$  non-negative definite symmetric matrices and  $\mathcal{S}_q^+$  its subset of positive definite matrices. The classical approach to optimal design [Kiefer (1974)] involves optimality criteria which depend on the design  $\xi$  only through its information matrix  $M_0$ , so that  $\Phi \equiv \Psi_0(M_0)$  (i.e.,  $\Phi(\xi) \equiv \Psi_0\{M_0(\xi)\}$ ), where  $\Psi_0: \mathcal{D} \rightarrow \mathcal{R}$  may be termed the *criterion function*. We assume that the domain  $\mathcal{D}$  of  $\Psi_0$  satisfies  $\mathcal{S}_q^+ \subseteq \mathcal{D} \subseteq \mathcal{S}_q$ , so that  $\Psi_0$  is defined for all nonsingular matrices, at least; see Section 2.1.1 below. The problem is then to maximize  $\Phi$  over some set  $\Xi$  of designs having information matrix in  $\mathcal{D}$ . Since smaller  $V_0$ , and thus larger  $M_0$ , is more desirable, attention is usually restricted to criterion functions that are nondecreasing in the Loewner ordering: if  $M_0 - M_1 \in \mathcal{S}_q$ , then  $\Psi_0(M_0) - \Psi_0(M_1) \geq 0$ . Simplification of the computational optimization problem can often be achieved by limiting attention to concave criterion functions [Kiefer (1974)].

DEFINITION 1. An optimality criterion  $\Phi$  will be called *regular* if  $\Phi$  can be expressed as  $\Psi_0(M_0)$ , where  $\Psi_0$  is concave and nondecreasing.

LEMMA 1. If  $\Phi$  is regular, then so is  $\Phi^* = g(\Phi)$  for any concave nondecreasing real function  $g(\cdot)$ .

PROOF. The composition of concave nondecreasing functions is concave nondecreasing.  $\square$

An important one-parameter class of optimality criteria is  $\tilde{\Phi}_s$ ,  $s \in \mathcal{R}$ , given by  $\tilde{\Phi}_s = \Pi_s(M_0)$ , where  $\Pi_s$  is the *matrix power mean* function [Chaloner and Verdinelli (1995); Kiefer (1974)] defined, for  $C \in \mathcal{S}_q^+$ , by

$$(2) \quad \Pi_s(C) = \begin{cases} \left\{ \frac{1}{q} \text{tr}(C^s) \right\}^{1/s}, & s \neq 0, \\ \det(C)^{1/q}, & s = 0. \end{cases}$$

We note for future reference the identity

$$(3) \quad \Pi_s(C^{-1}) \equiv \{\Pi_{-s}(C)\}^{-1}.$$

We further introduce criteria  $\tilde{\Phi}_l = \Pi_l(M_0)$ ,  $\tilde{\Phi}_{0,\alpha} = \Pi_{0,\alpha}(M_0)$  ( $\alpha \neq 0$ ), where the functions  $\Pi_l$ ,  $\Pi_{0,\alpha}$  on  $\mathcal{S}_q^+$  are defined by

$$(4) \quad \Pi_l(C) \equiv \log \det(C),$$

$$(5) \quad \Pi_{0,\alpha}(C) \equiv \text{sign}(\alpha)(\det C)^\alpha.$$

Some popular optimality criteria may be expressed in terms of those introduced above:

1. *A-optimality* (minimize  $\text{tr } V_0$ ):  $\tilde{\Phi}_{-1}$ ;
2. *D-optimality* (maximize  $\det M_0$ ):  $\tilde{\Phi}_0$ ,  $\tilde{\Phi}_l$  or  $\tilde{\Phi}_{0,\alpha}$ ;
3. *E-optimality* (maximize the minimum eigenvalue of  $M_0$ ):  $\tilde{\Phi}_{-\infty}$ .

Regularity of the optimality criterion  $\tilde{\Phi}_s$  for  $s \leq 1$  (and thus, in particular, of *A*-, *D*- and *E*-optimality) is a consequence of the following theorem [Pukelsheim, (1993), page 151].

**THEOREM 1.** *For  $q > 1$  the function  $\Pi_s$ , defined on  $\mathcal{S}_q^+$  by (2), is strictly concave if  $s < 1$ , linear for  $s = 1$  and strictly convex for  $s > 1$ . Furthermore,  $\Pi_s$  is nondecreasing, nonnegative and homogeneous of degree 1.*

Regularity of  $\tilde{\Phi}_l$ , and of  $\tilde{\Phi}_{0,\alpha}$  for  $\alpha \leq 1/q$ , follows from Lemma 1.

**2.1.1. Singular designs.** In some cases (e.g., if  $\Psi_0 \equiv \Pi_s$ ,  $s \geq 0$ ) it is possible to extend the domain  $\mathcal{D}$  of the criterion function  $\Psi_0$  beyond  $\mathcal{S}_q^+$  and so allow an optimal design which might be singular. In particular, suppose attention focuses on some linear function  $\beta = A\theta$ , where  $A$  is  $k \times q$  of rank  $k$ ; then there is no reason to avoid a singular design, so long as it makes  $\beta$  estimable (which will hold if the null space of  $M_0$  is contained in that of  $A$ ). In this case the least squares estimator  $\hat{\beta}$  is well defined and has nonsingular dispersion matrix  $V'_0 = AM_0^-A^T$ , where  $M_0^-$  is any  $g$ -inverse of  $M_0$ . Any optimality criterion of the form  $\Psi'_0(M'_0)$ , where  $\Psi'_0$  is a criterion function defined on  $\mathcal{S}_k^+$  (at least), and  $M'_0 = V_0'^{-1}$ , will then allow inclusion of such singular designs. Possible optimality criteria are  $\tilde{\Phi}'_s \equiv \Pi_s(M'_0)$ , for any  $s$ . For  $s = -1$  this yields *L-optimality*, or *c-optimality* when  $A = c^T$ ,  $c \in \mathcal{R}^q$ ; likewise, *D<sub>A</sub>-* and *D<sub>S</sub>-optimality* [Silvey (1980)] are special cases for  $s = 0$ .

**2.2. Bayesian optimal design.** The approach described in Section 2.1 can be given an “improper Bayesian” interpretation if we further assume a normal sampling distribution for  $Y: Y | \theta \sim N(X\theta, \sigma^2 I_n)$  ( $\sigma^2$  known); and regard  $\theta$  as the value of a random quantity  $\Theta$  with an improper uniform prior distribution. Then, when  $M_0$  is nonsingular,  $V_0 = M_0^{-1}$  can be reinterpreted as the posterior dispersion of  $\Theta$ .

For a simple proper Bayesian approach, we consider a normal prior distribution for  $\Theta$ . We can suppose that this has nonsingular dispersion, since otherwise  $\Theta$  is essentially confined to a subspace, which we can then take as our (redefined) parameter-space. Thus we take  $\Theta \sim N(\theta_0, R^{-1})$  with  $\theta_0$  and

$R \in \mathcal{S}_q^+$  known. Then the posterior distribution of  $\Theta$  given  $Y = y$  is normal, with dispersion

$$(6) \quad V_e := (R + M_0)^{-1}$$

and expectation  $V_e(R\theta_0 + X^T y/\sigma^2)$ . In particular, the posterior distribution is always proper, and its dispersion  $V_0$  nonsingular, even when the information  $M_0$  is singular. We term  $V_e$  the (*posterior*) *estimative dispersion* and  $M_e = V_e^{-1} = R + M_0$  the (*posterior*) *estimative information*. Note that  $V_e$  and  $M_e$  are functions of the design  $\xi$ , through  $M_0$ .

The proper Bayesian criterion-based approach [Pilz (1991)] parallels the classical one: given a criterion function  $\Psi_e: \mathcal{S}_q^+ \rightarrow \mathcal{R}$ , a design  $\xi^*$  is sought to maximize  $\Phi \equiv \Phi(\xi) = \Psi_e(M_e)$ . We can then reexpress  $\Phi$  as  $\Psi_0(M_0)$ , where  $\Psi_0: \mathcal{S}_q \rightarrow \mathcal{R}$  is defined by  $\Psi_0(C) \equiv \Psi_e(R + C)$ . Note that the domain of  $\Psi_0$  is the whole of  $\mathcal{S}_q$ , so that in the proper Bayesian approach there is no problem in admitting singular  $M_0$ , which we shall do in this context throughout this paper. If  $\Psi_e$  is concave nondecreasing, so too is  $\Psi_0$ , and thus  $\Phi$  will be regular [Kiefer (1974)]. Common Bayesian optimality criteria are again specializations, or monotone transformations, of the matrix power mean, now applied to  $M_e$ ; for example, Bayes  $D$ -optimality corresponds to  $s = 0$  and Bayes  $A$ -optimality to  $s = -1$  [Chaloner and Verdinelli (1995); Pilz (1991)].

**2.3. Predictive design.** We now consider the formulation of the design problem when the aim of the experiment is prediction of a future observation  $Z \in \mathcal{R}^m$  and give a sufficient condition for a predictive optimality criterion to be regular. We suppose that, given  $\theta$ ,  $Z$  is independent of the experiment  $\xi$  and its data  $Y$ , with  $E(Z|\theta) = T\theta$  and  $\text{Var}(Z|\theta) = \Sigma_Z$ . So long as we use a design which makes  $T\theta$  estimable, the least squares predictor of  $Z$  based on  $Y$  is  $T\hat{\theta}$ , where  $\hat{\theta}$  denotes any solution to the normal equations; and the *classical predictive dispersion*,  $\text{Var}(Z - T\hat{\theta})$ , is  $V_c = \Sigma_Z + TV_0T^T$ , where  $V_0$  now denotes any generalized inverse of  $M_0$ . This is also the Bayesian predictive dispersion of  $Z$  given  $Y$  when the prior for  $\Theta$  is improper uniform. A design to optimize prediction could be found by minimizing some nondecreasing function of  $V_e$  [Bates, Buck, Riccomagno and Wynn (1996); Sacks, Welch, Mitchell and Wynn (1989)], or equivalently by maximizing a nondecreasing function of the *classical predictive information*,  $M_c = V_c^{-1}$ .

In a proper Bayesian approach, we again assume  $\Theta \sim N(\theta_0, R^{-1})$ ,  $Y | \Theta = \theta \sim N(X\theta, \sigma^2 I_n)$  with  $\theta_0$ ,  $R$  and  $\sigma^2$  known, and  $Z|\Theta = \theta \sim N(T\theta, \Sigma_Z)$ . Then the *Bayesian predictive dispersion*,  $\text{Var}(Z | Y)$ , is  $V_p = \Sigma_Z + TV_eT^T$ , with associated *Bayesian predictive information*  $M_p = V_p^{-1}$ . Again,  $V_p$  and  $M_p$  are functions of the design  $\xi$  through  $M_0$ .

Consider now an optimality criterion of the form  $\Phi \equiv \Psi_p(M_p)$ , where  $\Psi_p: \mathcal{S}_m^+ \rightarrow \mathcal{R}$  is a nondecreasing function. Then  $\Phi$  is also expressible as  $\Psi_e(M_e)$ , with  $\Psi_e = \Psi_p \circ \gamma_{\Sigma_Z, T}$  where, for given  $B$  and nonnegative definite  $A$ ,  $\gamma_{A, B}$  is defined by

$$(7) \quad \gamma_{A, B}(C) \equiv (A + BC^{-1}B^T)^{-1}.$$

This in turn allows reexpression of  $\Phi$  as  $\Psi_0(M_0)$ , where  $\Psi_0(C) \equiv \Psi_e(R + C)$  (again  $\Psi_0$  is defined on the whole of  $\mathcal{S}_q$ ). Although the optimality criterion  $\Phi$  is initially specified using the predictive criterion function  $\Psi_p$ , its regularity depends on the concavity of the function  $\Psi_0$  relating  $\Phi$  to  $M_0$ . The following result establishes a sufficient condition for this.

**THEOREM 2.** *If  $\Psi_p$  is concave and nondecreasing, then so too are  $\Psi_e$  and  $\Psi_0$ , and hence  $\Phi \equiv \Psi_p(M_p)$  is regular.*

The proof follows from concavity and nondecreasingness of the function  $\gamma_{A,B}$  [Eaton, Giovagnoli and Sebastiani (1994)], and the fact that the composition of concave nondecreasing functions is concave nondecreasing.

A corresponding property holds for the limiting improper uniform case  $R = 0$ ; thus any concave nondecreasing function of the classical predictive information  $M_c$  will yield a regular optimality criterion (over designs making  $T\theta$  estimable).

**3. Decision-based optimal experimental design.** The above classical and Bayesian approaches to choice of experiment for the linear model give no guidance for choosing the optimality criterion. Bayesian decision theory addresses this issue. The formulation of the choice of experiment as a decision theory problem has a long tradition [Savage (1972); Lindley (1956, 1972)]. Bernardo and Smith (1994), and the recent review in Chaloner and Verdinelli (1995), also consider predictive decision problems. In this section we describe the approach in very general terms, following Dawid (1994). We return to the normal linear model in Section 5.

Consider a “terminal decision problem” with loss  $L(z, a)$  depending on the value  $z$  of a quantity  $Z \in \mathcal{Z}$  and on the action  $a \in \mathcal{A}$  taken. Let  $P \in \mathcal{P}$ , a convex family of proper distributions for  $Z$ , and write  $L(P, a)$  for the expected loss  $E_{Z \sim P}\{L(Z, a)\}$ . Suppose that, for each  $P \in \mathcal{P}$ ,  $\inf_a L(P, a)$  is finite and is achieved at a unique value,  $a(P)$  say. Then  $a(P)$  is the Bayes act when  $Z$  is assigned the distribution  $P$ , and  $H(P) := L\{P, a(P)\}$  is the associated Bayes loss, or *uncertainty*. The function  $H: \mathcal{P} \rightarrow \mathcal{R}$  is the *coherent uncertainty function* associated with the decision problem. An important property [DeGroot (1970), page 125] is that  $H$  is a concave function on  $\mathcal{P}$ .

Consider now the formulation of the choice of experiment as a decision problem [Lindley (1972)]. Let  $\Xi$  denote the class of possible experiments. If  $\xi \in \Xi$  is chosen, then observation will be made of a quantity  $Y \equiv Y_\xi$ , with distribution  $P_{\theta, \xi}^Y$  over a sample space  $\mathcal{Z}_\xi$  when  $\Theta = \theta$ . After  $Y_\xi = y$  is observed, an action  $a \in \mathcal{A}$  is chosen; then the quantity of interest,  $Z$ , is observed, and the loss suffered if  $Z = z$  is  $L(z, a)$ . We assume that  $Z$  has distribution  $P_\theta^Z$ , conditional on the value  $\Theta = \theta$ , independently of  $\xi$ ,  $y$  and  $a$ . Note particularly that this formulation includes the case of estimation, on taking  $Z \equiv \Theta$ , in which case  $P_\theta^Z$  is  $\delta_\theta$ , the point mass at  $\theta$  [Eaton, Giovagnoli and Sebastiani (1996)]. We further assign to  $\Theta$  a prior distribution, independent of  $\xi$ . We wish to choose  $\xi$ , and then, after observing  $Y_\xi$ , to take an optimal action  $a$ , so as to

minimize the overall expected loss. Let  $P_{\xi, y}^Z$  denote the predictive distribution for  $Z$  after conducting  $\xi$  and observing  $Y_\xi = y$ . The principle of minimizing expected loss then entails that the optimal choice for  $\xi$  is that which minimizes the *expected Bayes risk*

$$(8) \quad U(\xi) = E\{H(P_{\xi, Y_\xi}^Z)\},$$

where  $H$  is the coherent uncertainty function corresponding to the loss function  $L$ , and the expectation is over the marginal distribution of  $Y_\xi$  in  $\xi$ . We shall call a function  $U: \Xi \rightarrow \mathcal{R}$  of the form (8), for some coherent uncertainty function  $H$ , a *(design) standard*. The minimization of such a standard thus provides the general coherent prescription for choice of experiment. Decision-theoretic justifications for certain Bayesian optimality criteria, such as  $A$ - or  $D$ -optimality, are well known; see Chaloner and Verdinelli (1995) for a recent overview. However, there is as yet no general theory which allows us to associate directly an optimality criterion and a design standard. A major purpose of this paper is to fill this gap.

**4. Coherent uncertainty functions and dispersion.** We wish to relate the two approaches to optimal design, based on optimality criteria and on uncertainty functions, respectively. With some exceptions (e.g., nonlinear estimation problems), popular optimality criteria depend only on the information matrix of the design, or, equivalently, on the associated classical estimative dispersion. Hence, for purposes of comparison, in this paper we shall focus attention on uncertainty functions which depend on a distribution (for the parameter or for a future observable) only through its dispersion. That is to say, if two different distributions have the same dispersion matrix, we shall want the uncertainty function to assign them the same value. This is a “non-parametric” condition, in that, for example, we are not confining attention to normal distributions, but requiring the uncertainty function to be defined for essentially all possible distributions. Thus we are here excluding, for example, the use of the entropy uncertainty function [Good (1952); Lindley (1956); Stone (1959)], since, although for normal distributions it depends only on the dispersion, this property fails as soon as we move outside the normal family. Note also that our restriction to dependence on the dispersion excludes many interesting decision problems, for example, those in which one wishes to optimize an outcome [Verdinelli and Kadane (1992)].

Clearly the dispersion-based uncertainty condition will put constraints on the form of the uncertainty function. In this section we shall identify necessary and sufficient conditions for it to have this property, and, when it does, display a loss function which gives rise to it. To avoid misunderstandings, note that we are *not* requiring that the underlying *loss function*—when this is expressible in terms of an assessed distribution—should only depend on the dispersion of that distribution, and indeed we shall see from the proof of Theorem 4 below that this is not the case.

Let  $\mu_P$  and  $V_P$  denote, respectively, the mean vector and dispersion matrix of a distribution  $P$  over  $\mathcal{R}^m$ . We shall restrict attention to  $\mathcal{P}_m = \{P: V_P \text{ exists and is in } \mathcal{S}_m^+\}$ . Then  $\mathcal{P}_m$  is convex.

**DEFINITION 2.** A function  $K: \mathcal{S}_m^+ \rightarrow \mathcal{R}$  is called a *coherent dispersion function*, or *dispersion-coherent*, if  $H: \mathcal{P}_m \rightarrow \mathcal{R}$ , defined by  $H(P) \equiv K(V_P)$ , is a coherent uncertainty function (hence concave) on  $\mathcal{P}_m$ . In this case we may also term  $H$  a *dispersion-coherent uncertainty function*.

**THEOREM 3.** *If  $K$  is a coherent dispersion function, then  $K$  is concave and nondecreasing on  $\mathcal{S}_m^+$ .*

**PROOF.** Suppose that  $H(P) \equiv K(V_P)$  defines a dispersion-coherent uncertainty function. Then  $H$  must be a concave function on  $\mathcal{P}_m$ . Given  $V, W \in \mathcal{S}_m^+$ , we can find  $P, Q \in \mathcal{P}_m$ , with  $V_P = V, V_Q = W$  and  $\mu_P = \mu_Q = \mu$ , say. Then defining  $R = \alpha P + (1 - \alpha)Q$  for  $\alpha \in [0, 1]$ , we have  $R \in \mathcal{P}_m$ , and  $V_R = \alpha V + (1 - \alpha)W$ . Hence  $K\{\alpha V + (1 - \alpha)W\} = H(R) \geq \alpha H(P) + (1 - \alpha)H(Q) = \alpha K(V) + (1 - \alpha)K(W)$ . Thus  $K$  is concave.

Now let  $V \in \mathcal{S}_m^+, W \in \mathcal{S}_m$ , and take  $P$  and  $Q$  such that  $V_P = V, V_Q = W$ . Let  $R$  be the convolution  $P * Q$ . Then  $R \in \mathcal{P}_m$ , and  $V_R = V + W$ . We can express  $R = E(P_Y)$ , where  $P_y = P * \delta_y$ ,  $\delta_y$  being the point mass at  $y$ , and  $Y \sim Q$ . So  $K(V + W) = H(R) = H\{E(P_Y)\} \geq E\{H(P_Y)\} = K(V)$ . Hence  $K$  is nondecreasing.  $\square$

**COROLLARY 1.** *If  $K$  is a coherent dispersion function, then, for any fixed  $V, K(\lambda V)$  is a concave nondecreasing function of  $\lambda$ .*

We shall now show the converse of Theorem 3 constructively. Let  $K: \mathcal{S}_m^+ \rightarrow \mathcal{R}$  be concave and nondecreasing. Regarding  $\mathcal{S}_m^+$  as an open subset of  $\mathcal{R}^{m(m+1)/2}$ , it follows from concavity [Rockafellar (1970), page 217] that  $K$  has a ‘‘supergradient’’  $\Gamma_W$  at each  $W \in \mathcal{S}_m^+$ . That is,  $\Gamma_W$  is a symmetric matrix such that

$$(9) \quad K(V) \leq K(W) + \text{tr}\{\Gamma_W(V - W)\} \quad \text{all } V \in \mathcal{S}_m^+.$$

Since  $K$  is nondecreasing,  $\Gamma_W \in \mathcal{S}_m$ . If  $K$  is differentiable at  $W$ , then  $\Gamma_W$  is just the derivative  $dK(W)/dW$  of  $K$  at  $W$ , that is, the matrix with diagonal elements  $\partial K(W)/\partial w_{ii}$  and off-diagonal elements  $(1/2)\partial K(W)/\partial w_{ij}$ . Otherwise,  $\Gamma_W$  may not be unique, but we suppose that a specific choice is made at each  $W \in \mathcal{S}_m^+$ .

**THEOREM 4.** *If  $K$  is concave and nondecreasing in  $\mathcal{S}_m^+$ , then  $K$  is a coherent dispersion function.*

**PROOF.** Consider the decision problem with observation space  $\mathcal{R}^m$ , action space  $\mathcal{A} = \mathcal{P}_m$ , and loss function given by

$$(10) \quad L(z, Q) = (z - \mu_Q)^T \Gamma_Q (z - \mu_Q) + K(V_Q) - \text{tr}(V_Q \Gamma_Q),$$



where we have now written  $\Gamma_Q$  for  $\Gamma_{V_Q}$ . Then  $L(P, Q)$  exists for  $P \in \mathcal{P}_m$ , and

$$(11) \quad L(P, Q) = K(V_Q) + (\mu_P - \mu_Q)^T \Gamma_Q (\mu_P - \mu_Q) + \text{tr}\{(V_P - V_Q)\Gamma_Q\}.$$

In particular,  $L(P, P) = K(V_P)$ , and it is easy to show using  $\Gamma_W \in \mathcal{S}_m$  and (9) that  $L(P, Q) \geq L(P, P)$ . Hence the optimal act when  $Z \sim P$  is just  $P$  itself [i.e., (10) defines a *proper scoring rule*], and the associated Bayes loss  $H(P)$  is  $K(V_P)$ .  $\square$

**COROLLARY 2.** *For  $m > 1$ , the matrix power mean function  $\Pi_s$  is a coherent dispersion function if and only if  $s \leq 1$ .*

The proof is from Theorems 1, 3 and 4.

We now consider the effect of applying a concave nondecreasing transformation to a coherent dispersion function  $K$ . Define  $k_0 = \inf\{K(V) : V \in \mathcal{S}_m^+\}$ . The range of  $K$  is contained in  $\mathcal{X} = (k_0, +\infty)$ . Let  $g: \mathcal{X} \rightarrow \mathcal{R}$  be concave and nondecreasing, and define  $K^*(V) \equiv g\{K(V)\}$ . Theorems 3 and 4, together with the proof of Lemma 1, imply the following lemma.

**LEMMA 2.** *If  $K$  is a coherent dispersion function, then so is  $K^* = g(K)$  for any concave nondecreasing real function  $g(\cdot)$ .*

We can relate the proper scoring rules, say  $L$  and  $L^*$ , associated with  $H$  and  $H^*$  as in (10), as follows. Since  $g$  is finite, concave and nondecreasing, at each  $y \in \mathcal{X}$  there exists a nonnegative supergradient for  $g$ , that is, a scalar  $\psi_y \geq 0$  such that

$$(12) \quad g(x) \leq g(y) + \psi_y(x - y) \quad \text{all } x \in \mathcal{X}.$$

Now from (9), if  $V, W \in \mathcal{S}_m^+$ , then  $K(W) + \text{tr}\{\Gamma_W(V - W)\} \in \mathcal{X}$ , and  $K^*(V) = g\{K(V)\} \leq g[K(W) + \text{tr}\{\Gamma_W(V - W)\}]$ , and thus, by (12),  $K^*(V) \leq K^*(W) + \psi_{K(W)} \text{tr}\{\Gamma_W(V - W)\}$ . It follows that  $\Gamma_W^* = \psi_{K(W)} \Gamma_W$  is a supergradient of  $K^*$  at  $W$ , and hence, from (10), that we can take  $L^*(z, Q) \equiv K^*(V_Q) + \psi_{K(V_Q)}\{L(z, Q) - K(V_Q)\}$ .

**THEOREM 5.** *Let  $K_0$  be a homogeneous coherent dispersion function, that is,  $K_0(\lambda V) \equiv \lambda K_0(V)$ , ( $\lambda > 0$ ). Then  $K(V) \equiv g\{K_0(V)\}$  defines a coherent dispersion function  $K$  if and only if  $g: \mathcal{R} \rightarrow \mathcal{R}$  is concave and nondecreasing.*

**PROOF.** “If” follows from Lemma 2. Now suppose  $K$  is dispersion-coherent and fix  $V_0$  with  $K_0(V_0) = 1$ . Then  $g(\lambda) \equiv g\{\lambda K_0(V_0)\} \equiv g\{K_0(\lambda V_0)\} \equiv K(\lambda V_0)$ . Now apply Corollary 1.  $\square$

From Theorem 1, Theorem 5 applies when  $K(\cdot)$  is the matrix power mean function  $\Pi_s(\cdot)$  ( $s \leq 1$ ). In particular, a dispersion function depending only on  $\det V$  is dispersion-coherent if and only if it is of the form  $g(\det V^{1/m})$  with  $g$  concave and nondecreasing. Thus  $\log \det V$  is a coherent dispersion function (and may be so used whether or not the distribution is normal).

TABLE 1  
Some coherent dispersion functions and their associated loss functions

$H(P) \equiv K(V_P)$	$L(z, Q)$
$\Pi_s(V_P)$ ( $s \neq 0, s \leq 1$ )	$\left(\frac{1}{m}\right)^{1/s} \{\text{tr}(V_Q^s)\}^{(1/s)-1} (z - \mu_Q)^T V_Q^{s-1} (z - \mu_Q)$
$\Pi_0(V_P)$	$\frac{1}{m} \det(V_Q)^{1/m} (z - \mu_Q)^T V_Q^{-1} (z - \mu_Q)$
$\Pi_l(V_P)$	$\log \det V_Q + (z - \mu_Q)^T V_Q^{-1} (z - \mu_Q)$
$\Pi_{0,\alpha}(V)$ ( $\alpha \leq 1/m$ )	$\text{sign}(\alpha) (\det V_Q)^\alpha \left\{ \alpha (z - \mu_Q)^T V_Q^{-1} (z - \mu_Q) + 1 - m\alpha \right\}$

Some examples of uncertainties based on coherent dispersion functions, with their associated loss functions, are given in Table 1 (where, in the definitions of  $\Pi_s$  etc. we take  $q = m$ ). All functions are differentiable in  $\mathcal{S}_m^+$ , so that we can take the associated decision problem to have action space  $\mathcal{P}_m$  and loss function given by (10), with  $\Gamma_W = dK(V)/dW$ . In computing the loss function associated with  $H(P) \equiv \Pi_s(V_P)$ , we have used the matrix derivative  $d \text{tr}(V^s)/dV = sV^{s-1}$  [Sebastiani (1996b)]. Note further that if  $K$  is a coherent dispersion function on  $\mathcal{S}_M^+$ , then so is  $K_A$  on  $\mathcal{S}_m^+$ , defined by  $K_A(V) \equiv cK(AVA^T)$ , for fixed positive constant  $c$  and  $m \times m$  matrix  $A$ . This arises from an associated decision problem with loss function involving  $AZ$  in place of  $Z$ . In this way we obtain further coherent dispersion functions from Table 1. In particular,  $\Pi_{1,A}(V) \equiv (1/m) \text{tr}(VC)$ , where  $C = A^T A$ , defines the coherent dispersion function corresponding to the proper scoring rule  $S(z, Q) \equiv (1/m)(z - \mu_Q)^T C(z - \mu_Q)$ , or equivalently to the simple quadratic loss function  $L(z, a) = (1/m)(z - a)^T C(z - a)$  with  $a \in \mathcal{A} = \mathcal{R}^m$ ; see Chaloner (1984) for applications to normal linear regression models.

**5. Coherent dispersion standards for the normal linear model.** In this section, we apply the above considerations to the choice of experiment in the specific case of a normal regression model with conjugate prior distribution. As in Section 2.2, we shall assume that  $Y | \theta \sim N(X\theta, \sigma^2 I_n)$ ,  $\theta \in \mathcal{R}^p$ ,  $X$  a  $n \times q$  matrix determined by the experiment  $\xi$  and  $\Theta \sim N(\theta_0, R^{-1})$  with  $\theta_0$  and  $R$  known. Without loss of generality we suppose  $\sigma^2$  known; Verdinelli (1996) has shown that when the variance is unknown the design problem is essentially independent of  $\sigma^2$ . Then  $\Theta | Y$  has a normal distribution, with dispersion matrix  $V_e = M_e^{-1} = (R + M_0)^{-1}$ .

5.1. *Coherent dispersion standards.* Let  $K_e: \mathcal{S}_q^+ \rightarrow \mathcal{R}$  be concave and non-decreasing. Thus, with  $P \in \mathcal{P}_q$  a distribution for  $\Theta$ ,  $H(P) \equiv K_e(V_P)$  defines a coherent uncertainty function corresponding to an estimative decision problem. If we use this uncertainty function in the Bayesian normal linear model, the coherent solution (8) reduces to minimizing  $U \equiv U(\xi) = K_e(V_e)$ , where  $V_e$  is given by (6). Note that, since  $V_e$  is independent of the data  $Y_\xi$  observed

in  $\xi$ , the expectation in (8) is not required. We shall call  $U \equiv K_e(V_e)$ , with  $K_e$  concave and nondecreasing (and  $V_e$  regarded as a function of the experiment  $\xi$ ), an *estimative coherent dispersion standard*.

In particular, when  $M_0$  is nonsingular, the classical estimative variance  $V_0 = M_0^{-1}$  is the formal posterior variance of  $\Theta|Y$  for the limiting case  $R \rightarrow 0$ , when the prior for  $\Theta$  is the (improper) uniform distribution. Thus if  $K_0: \mathcal{S}_q^+ \rightarrow \mathcal{R}$  is concave and nondecreasing, then we can regard  $U \equiv K_0(V_0)$  as the relevant quantity to minimize, to solve some formal Bayesian estimative decision problem, based on an improper uniform prior (and restricting attention to experiments yielding a proper posterior). We may call  $U \equiv K_0(V_0)$  a *formally coherent dispersion standard*. (Specific choices for  $K_0$  may admit extension of its domain so that  $U$  remains defined for some or all singular  $M_0$ . For example, if, as in Section 2.1.1, attention focuses on  $\beta = A\theta$ , we might define  $U = K'_0(V'_0)$ , with  $K'_0: \mathcal{S}_k^+ \rightarrow \mathcal{R}$  concave and nondecreasing, and  $K_0$  reexpressing  $U$  as a function of  $V_0$ .)

Suppose now that  $U \equiv K_e(V_e)$  is a (proper) estimative coherent dispersion standard, and define  $K_0: \mathcal{S}_q^+ \rightarrow \mathcal{R}$  by  $K_0 = K_e \circ \gamma_{R, I_q}$ , with  $\gamma_{A, B}$  given by (7). Since  $\gamma_{R, I_q}(V_0) = V_e$ , we have  $U \equiv K_0(V_0)$ . Thus  $K_0$  is concave and nondecreasing, so that the estimative coherent dispersion standard  $U$  is also a formally coherent dispersion standard. (Note, moreover, that  $U$  remains well defined for all singular  $M_0$ .)

For a predictive decision problem, we would consider standards of the form  $U \equiv K_p(V_p)$  with  $K_p$  concave and nondecreasing. But then also  $U \equiv K_e(V_e)$ , with  $K_e: \mathcal{S}_q^+ \rightarrow \mathcal{R}$  defined by  $K_e(C) \equiv K_p(\Sigma_Z + TCT^T)$ . When  $K_p$  is concave and nondecreasing, so is  $K_e$ . Hence any predictive coherent dispersion standard is also an estimative coherent dispersion standard. This is a special case of a well-known result [Aitchison and Dunsmore (1975), page 64]; a predictive decision problem can always be reformulated as an estimative decision problem. Finally, as shown above, we can further reexpress  $U \equiv K_e(V_e)$  as a formally coherent dispersion standard  $K_0(V_0)$ . Thus we may speak of a standard  $U$ , whether initially defined as a concave nondecreasing function of  $V_0$ , of  $V_e$  or of  $V_p$ , as, simply, a *(coherent) dispersion standard*.

**5.2. Regular-coherent optimality criteria and dispersion standards.** We now consider the relationship between an optimality criterion  $\Phi$  and a dispersion standard  $U$ . Suppose there exists a decreasing function  $g: \mathcal{R} \rightarrow \mathcal{R}$  such that  $\Phi \equiv g(U)$ , or equivalently  $U \equiv h(\Phi)$ , with  $h$  the inverse function of  $g$ . Then maximizing  $\Phi$  is equivalent to minimizing  $U$ . In particular, if we have, simultaneously, both coherence of  $U$  and regularity of  $\Phi$ , then the problem is both based on a genuine decision problem, and amenable to solution by convex optimization methods. In this case we shall call the design problem *regular-coherent*,  $\Phi$  a *regular-coherent optimality criterion*, and  $U$  a *regular-coherent dispersion standard*.

For instance, if we only consider designs with a nonsingular information matrix, regular-coherence is obtained when  $\Phi$  and  $U$  are expressible as  $\Phi \equiv$

$\Psi_0(M_0)$ ,  $U \equiv K_0(V_0)$ , respectively, with both  $\Psi_0$  and  $K_0$  concave and non-decreasing in  $\mathcal{S}_q^+$ , and  $\Phi \equiv g(U)$  or  $U \equiv h(\Phi)$ , with the mutually inverse functions  $g$  and  $h$  decreasing. Given only  $K_0$ , or only  $\Psi_0$ , regular-coherence requires the existence of the other, satisfying the above conditions. This is expressible as either (and then both) of the functions  $\Psi_0$  and  $K_0$  being *associable*, as given by the following definition.

DEFINITION 3. A function  $F: \mathcal{S}_q^+ \rightarrow \mathcal{R}$  is *associable* if:

- (i)  $F$  is concave and nondecreasing.
- (ii) There exist a decreasing function  $h: \mathcal{R} \rightarrow \mathcal{R}$  and a concave nondecreasing function  $G: \mathcal{S}_q^+ \rightarrow \mathcal{R}$  such that

$$(13) \quad F(C) \equiv h\{G(C^{-1})\}.$$

In this case  $G$  is also associable, and we call  $F$  and  $G$  *associated functions*.

LEMMA 3. If  $F$  is associable, and  $F^*(C) \equiv k\{F(C)\}$  with  $k$  concave and nondecreasing, then  $F^*$  is associable.

The proof follows easily from Lemma 2.

THEOREM 6. If  $q > 1$ , the power mean function  $\Pi_s: \mathcal{S}_q^+ \rightarrow \mathcal{R}$ , given by (2), is associable if and only if  $-1 \leq s \leq 1$ .

PROOF. From Theorem 1, (i) of Definition 3 holds for  $F \equiv \Pi_s$  if and only if  $s \leq 1$ . We shall show that (ii) of Definition 3 holds if and only if  $s \geq -1$ . Suppose first that  $s \geq -1$  and define  $G(C) \equiv \Pi_{-s}(C)$ . Then  $G$  is concave and nondecreasing. Then, using (3),  $\Pi_s(C) \equiv \{\Pi_{-s}(C^{-1})\}^{-1} \equiv \{G(C^{-1})\}^{-1}$ , so that (ii) holds with  $h(x) \equiv x^{-1}$ . Suppose now  $s < -1$  and  $\Pi_s(C) = h\{G(C^{-1})\}$  for some decreasing function  $h$ . Then  $\Pi_s(C^{-1}) = h\{G(C)\}$ . From (3),  $\Pi_s(C^{-1}) \equiv \{\Pi_{-s}(C)\}^{-1}$ . So  $G(C) \equiv h^{-1}\{\{\Pi_{-s}(C)\}^{-1}\}$  is an increasing function of  $\Pi_{-s}(C)$ , which, by Theorem 1, is strictly convex. So  $G$  cannot be concave, and (ii) fails.  $\square$

Thus, for  $|s| \leq 1$ , we have regular-coherence for the design strategies based on the optimality criterion  $\tilde{\Phi}_s \equiv \Pi_s(M_0)$ , as well as on the (formally coherent) dispersion standard  $\Pi_s(V_0)$  and likewise for  $\tilde{\Phi}_l \equiv \Pi_l(M_0)$  or its associated dispersion standard  $\Pi_l(V_0)$ . In particular, we obtain regular-coherence of  $A$ -optimality ( $\tilde{\Phi}_{-1}$ ) and  $D$ -optimality ( $\tilde{\Phi}_0$  or  $\tilde{\Phi}_l$ ). However,  $E$ -optimality ( $\tilde{\Phi}_{-\infty}$ ), while regular, is not associated with a coherent dispersion standard, and thus cannot arise from solving any decision problem with a dispersion-based uncertainty function. This goes some way towards confirming a conjecture of Chaloner and Verdinelli (1995). Note, however, that we have *not* shown that  $E$ -optimality, as a normal-theory design criterion, could not arise as the solution of any decision problem. This stronger conclusion would follow if a general uncertainty function which, when confined to the normal distributions,

depended only on the dispersion, would then have to be concave and nondecreasing in that normal dispersion. Although we are not aware of any counter examples, we have not been able to show this.

5.2.1. *Estimation and prediction.* Suppose we start with an estimative Bayesian optimality criterion expressed as  $\Phi \equiv \Psi_e(M_e)$ , where the function  $\Psi_e$  is associable. Then, as shown in Section 2.2,  $\Phi$  is regular. Let  $K_e$  be an associated function of  $\Psi_e$ . Then maximizing  $\Phi$  is equivalent to minimizing the coherent dispersion standard  $U := K_e(V_e)$ . Hence this proper Bayes estimation problem is regular-coherent. Conversely, we obtain a regular-coherent problem if we start from an estimative dispersion standard  $U \equiv K_e(V_e)$ , with  $K_e$  associable. Similarly, applying Section 2.3, we could start with a predictive optimality criterion  $\Phi \equiv \Psi_p(M_p)$ , which will be regular and be associated with a coherent predictive dispersion standard  $U := K_p(V_p)$ , or, alternatively, start with such a  $U$  and construct  $\Phi$ . Again, we have a regular-coherent problem. We note particularly that, for these proper Bayes estimation and prediction problems, there is no need to require  $M_0$  to be nonsingular.

We thus see that we can produce a regular-coherent problem by using an associable function, no matter whether we apply it to  $M_0$ , to  $V_0$ , to  $M_e$ , to  $V_e$ , to  $M_p$  or to  $V_p$ . In all cases, the problem will correspond to a genuine Bayesian decision problem and will be amenable to convex optimization techniques.

**6. More general models.** Although we have so far mostly limited consideration to the normal linear model, the approach described extends, in principle, to situations involving more general assumptions on the model constituents  $Y|x, \theta$  and  $Z|\theta$ , and on the prior density  $\pi(\theta)$ . In this section, we conduct an asymptotic analysis of such more general problems. It is only assumed that the model constituents are suitably well behaved, and that  $\pi(\cdot)$  is continuous and everywhere positive. We continue to suppose that distinct observations are independent, so that a design may be identified with its design measure  $\xi$  [Chaloner and Verdinelli (1995)].

6.1. *Optimality criteria.* One might introduce an *optimality criterion*  $\Phi \equiv \Phi(\xi)$ , and seek the allowable design  $\xi$  which maximizes it. Concavity of  $\Phi(\xi)$  on the space  $\Xi$  of design measures  $\xi$  will assist optimization by allowing application of the general equivalence theorem [Whittle (1973)]. Commonly used optimality criteria have the general form

$$(14) \quad \Phi \equiv \Phi(\xi) \equiv \int \Psi_0\{I_\xi(\theta)\}\pi(\theta) d\theta,$$

where  $I_\xi(\theta)$  is the expected Fisher information matrix at  $\theta$  in experiment  $\xi$ , and  $\Psi_0: \mathcal{P}_m \rightarrow \mathcal{R}$  is an appropriate *criterion function*: see Chaloner and Verdinelli (1995) for a detailed discussion. Typically the function  $\Psi_0$  is concave and nondecreasing, and then the optimality criterion  $\Phi$  defined by (14) will be concave on  $\Xi$ , as desired. We can further transform such a  $\Phi$  by a concave, nondecreasing real function, to obtain an equivalent concave optimality criterion.

6.2. *Decision-based approach.* In full generality, minimizing a design standard  $U(\xi)$  as given by (8) supplies the coherent solution to optimal choice of experiment. However, the computational problems involved in applying this coherent design strategy can be daunting. Here we investigate solutions based on asymptotic approximations. Since a predictive decision problem can always be reformulated as one of estimation, we shall confine attention to the posterior distribution of  $\Theta$ . We shall again restrict ourselves to decision problems for which the uncertainty in a distribution  $P$  for  $\Theta$  depends only on its dispersion matrix  $V_P$ ; that is,  $H(P) \equiv K_0(V_P)$ , with  $K_0$  a coherent (so concave and nondecreasing) dispersion function. Assuming that the usual asymptotic normal approximation to the posterior distribution applies,  $U(\xi)$  can then be approximated as

$$(15) \quad U(\xi) \doteq \int K_0\{I_\xi(\theta)^{-1}\}\pi(\theta) d\theta$$

[Chaloner and Verdinelli (1995)]. Thus, we may call an optimality criterion  $\Phi$  *dispersion-coherent* if it can be expressed as a decreasing function of a quantity  $U$ , of the form of the right-hand side of (15), for some coherent dispersion function  $K_0$ . Then maximizing  $\Phi$  will deliver, asymptotically at least, the solution to a decision problem. An example of a dispersion-coherent optimality criterion is

$$(16) \quad g(E[f\{\det I_\xi(\Theta)^{-1/q}\}]),$$

with  $f$  concave and nondecreasing and  $g$  decreasing, the expectation being over the prior distribution of  $\Theta$ . It follows from Section 4 that any dispersion-coherent optimality criterion depending only on the determinant of the expected information matrix must be of the form (16), thus answering the question raised by Atkinson and Donev [(1992), pages 213 and 214] as to the appropriate way to generalize the criterion of  $D$ -optimality to nonlinear nonnormal problems. They considered criteria equivalent to:

1.  $E[\log \det\{I_\xi(\Theta)\}]$ ;
2.  $-\log E[\det\{I_\xi(\Theta)^{-1}\}]$ ;
3.  $-\log \det E\{I_\xi(\Theta)^{-1}\}$ ;
4.  $\log E\{\det I_\xi(\Theta)\}$  [Zacks (1977); Pronzato and Walter (1985)];
5.  $\log \det[E\{I_\xi(\Theta)\}]$ .

Of these, only 1 and 4 are of the form (16), with, respectively,  $f(x) \equiv q \log x$ ,  $g(x) \equiv -x$  and  $f(x) \equiv -x^{-q}$ ,  $g(x) \equiv \log(-x)$ . These findings help to address the concerns of Firth (1996) as to the “naturalness” of these several variant forms.

Other dispersion-coherent optimality criteria are  $g(E[\Pi_s\{I_\xi(\Theta)^{-1}\}])$ , for  $s \leq 1$  and  $g$  decreasing, or, equivalently (on putting  $t = -s$ )  $g\{E([\Pi_t\{I_\xi(\Theta)\}]^{-1})\}$ , for  $t \geq -1$ ; and, again with  $g$  decreasing,

$$g(E[\Pi_{0,\alpha}\{I_\xi(\Theta)^{-1}\}]) = g[E\{\text{sign}(\alpha) \det I_\xi(\Theta)^{-\alpha}\}],$$

for  $\alpha \leq 1/q$ , being of the form (16) with  $f(x) \equiv \text{sign}(\alpha)x^{\alpha q}$ . So, for example, the optimality criterion  $E\{\det I_{\xi}(\Theta)^{1/2}\}$ , which has been proposed as a criterion invariant to reparametrization when using a uniform prior [Firth and Hinde (1997b)], is dispersion-coherent.

**6.3. Regular-coherence.** As in the normal linear model, we wish to relate the two approaches: the criterion-based approach, which seeks to maximize (14), where  $\Psi_0$  is concave and nondecreasing and the (approximate) decision-theoretic approach, which seeks to minimize (15), with  $K_0$  concave and nondecreasing. In particular, we are interested in criteria which are *regular-coherent*, being simultaneously interpretable in both these ways.

We now describe a sufficient condition on the dispersion function  $K_0$  under which the criterion  $\Phi(\xi) := -U(\xi)$ , as given by (15), will be regular-coherent for any model in the general regression class considered, and any prior distribution. This result extends a remark of Chaloner and Verdinelli (1995).

**DEFINITION 4.** A dispersion function  $K_0$  will be termed *strongly regular* if the function  $M \mapsto -K_0(M^{-1})$  is concave and nondecreasing.

Note that an associable function  $F$ , if used as a coherent dispersion function, will be strongly regular whenever the function  $h$  appearing in (ii) of Definition 3 can be taken as convex—in which case we may always take  $h(x) \equiv -x$ .

The importance of strong regularity is captured by the following result, which follows easily from the definitions.

**THEOREM 7.** *Suppose that  $K_0$  is a strongly regular and coherent dispersion function. Then the optimality criterion*

$$(17) \quad \Phi(\xi) := -U(\xi) = - \int K_0\{I_{\xi}(\theta)^{-1}\}\pi(\theta) d\theta$$

*is regular-coherent.*

**LEMMA 4.** *The power mean function  $\Pi_s$  is strongly regular for  $s \geq -1$ .*

**PROOF.** From (3), we can write  $-\Pi_s(M^{-1})$  as  $g\{\Pi_{-s}(M)\}$ , where  $g(x) \equiv -x^{-1}$  is a concave nondecreasing function of  $x$ , and  $\Pi_{-s}(M)$  a concave nondecreasing function of  $M$ .  $\square$

Similar arguments show the strong regularity of  $\Pi_s^*$ , defined by  $\Pi_s^*(V) \equiv -\Pi_s(V)^{-1} \equiv -\Pi_{-s}(V^{-1})$  for  $s \geq -1$ , of  $\Pi_{-1,A}$ , of  $\Pi_{0,\alpha}$  for  $\alpha \geq -1/q$  [leading to concavity properties for the corresponding  $\Phi$ , given by (17), related to those of Firth and Hinde (1997a)] and of  $\Pi_l(V) \equiv \log \det(V)$ .

It follows from Corollary 2 and Lemma 4 that the optimality criterion

$$-E([\Pi_s\{I_{\xi}(\Theta)\}]^{-1})$$

is regular-coherent for  $-1 \leq s \leq 1$ . Likewise regular-coherent are the optimality criteria  $E([\Pi_s\{I_{\xi}(\Theta)\}])$  obtained from the strongly regular and coherent

dispersion function  $K(V) \equiv \Pi_{-s}^*(V)$  ( $-1 \leq s \leq 1$ ). Taking  $K(V) \equiv \Pi_{0,\alpha}(V) \equiv \text{sign}(\alpha) \det V^\alpha$ , the optimality criterion  $-E[\Pi_{0,\alpha}\{I_\xi(\Theta)^{-1}\}] = E[\Pi_{0,-\alpha}\{I_\xi(\Theta)\}]$  will be regular-coherent for  $-1/q \leq \alpha \leq 1/q$ . In particular,  $-E\{\det(I_\xi(\Theta)^{-1})\}$  is regular but not dispersion-coherent, having the form (16) with  $f(x) \equiv x^q$ , which is convex for  $q > 1$ ; while  $E\{\det I_\xi(\Theta)\}$  is dispersion-coherent, but not regular.

**7. Conclusion.** In this paper we have attempted to tie together the two main approaches to optimal experimental design, criterion-based and decision-based. We have confined attention to measures of performance determined by an appropriate dispersion matrix, and, under this restriction, have developed conditions under which optimal design for a normal linear model may be simultaneously justified from either point of view. For more general problems, we have extended the approach to criteria based on asymptotic approximation of the posterior distribution. In fact, even for small sample problems, the use of design criteria derived from coherent dispersion functions can be shown to be valuable for simplifying computationally intensive solution methods [Sebastiani (1996a)].

Throughout this paper we have assumed that the experimenter knows the functional form of the model for all uncertain quantities involved in the experiment. In many realistic situations this will not be the case, and the problem will then need to incorporate some aspect of model choice, or model mixture, prior to estimation or prediction. The decision-based formulation can, in principle, be extended to incorporate such model uncertainty; however, the complexity of the ensuing computational problem will demand the development of new methods of solution, perhaps based on stochastic optimization. Such methods may also prove fruitful for handling nonconcave criteria and for the nonasymptotic analysis of nonlinear nonnormal problems.

**Acknowledgments.** The authors are grateful to the Editors and referees for helping to improve the initial version of this paper.

## REFERENCES

- AITCHISON, J. and DUNSMORE, I. R. (1975). *Statistical Prediction Analysis*. Cambridge Univ. Press.
- ATKINSON, A. C. (1996). The usefulness of optimum experimental designs (with discussion). *J. Roy. Statist. Soc. Ser. C* **58** 59–76; 95–111.
- ATKINSON, A. C. and DONEV, A. N. (1992). *Optimum Experimental Design*. Clarendon Press, Oxford.
- BATES, R. A., BUCK, R. J., RICCOMAGNO, E. and WYNN, H. P. (1996). Experimental design and observation for large systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **58** 77–111.
- BERNARDO, J. M. and SMITH, A. F. M. (1994). *Bayesian Theory*. Wiley, New York.
- CHALONER, K. (1984). Optimal Bayesian experimental designs for linear models. *Ann. Statist.* **12**, 283–300.
- CHALONER, K. and VERDINELLI, I. (1995). Bayesian experimental design: a review. *Statist. Sci.* **10** 273–304.
- DAWID, A. P. (1994). Proper measures of discrepancy, uncertainty and dependence, with applications to predictive experimental design. Research Report 139, Dept. Statistical Science, Univ. College London.



- DEGROOT, M. H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.
- EATON, M. L., GIOVAGNOLI, A. and SEBASTIANI, P. (1994). A predictive approach to the Bayesian design problem with application to normal regression models. Technical Report 598, School of Statistics, Univ. Minnesota.
- EATON, M. L., GIOVAGNOLI, A. and SEBASTIANI, P. (1996). A predictive approach to the Bayesian design problem with application to normal regression models. *Biometrika* **83** 111–125.
- FIRTH, D. (1996). Discussion of the papers by Atkinson (1996) and by Bates et al. (1996). *J. Roy. Statist. Soc. Ser. B* **58** 100.
- FIRTH, D. and HINDE, J. P. (1997a). On Bayesian  $D$ -optimum design criteria and the equivalence theorem in nonlinear models. *J. Roy. Statist. Soc. Ser. B* **59** 793–797.
- FIRTH, D. and HINDE, J. P. (1997b). Parameter-neutral optimum design for non-linear models. *J. Roy. Statist. Soc. Ser. B* **59** 799–811.
- GOOD, I. J. (1952). Rational decisions. *J. Roy. Statist. Soc. Ser. B* **14** 107–114.
- KIEFER, J. C. (1959). Optimum experimental designs (with discussion). *J. Roy. Statist. Soc. Ser. B* **21** 272–319.
- KIEFER, J. C. (1974). General equivalence theorem for optimum designs: approximate theory. *Ann. Statist.* **2** 849–879.
- LINDLEY, D. V. (1956). On a measure of information provided by an experiment. *Ann. Math. Statist.* **27** 986–1005.
- LINDLEY, D. V. (1972). *Bayesian Statistics A Review*. SIAM, Philadelphia.
- PILZ, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models*. Wiley, New York.
- PRONZATO, L. and WALTER, E. (1985). Robust experimental design via stochastic approximations. *Math. Biosci.* **75** 103–120.
- PUKELSHEIM, F. (1993). *Optimal Design of Experiments*. Wiley, New York.
- RAIFFA, H. A. and SCHLAIFER, R. S. (1961). *Applied Statistical Decision Theory*. MIT Press.
- ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton Univ. Press.
- SACKS, J., WELCH, W., MITCHELL, T. and WYNN, H. P. (1989). Design and analysis of computer experiments (with discussion). *Statist. Sci.* **4** 409–423.
- SAVAGE, L. J. (1972). *The Foundations of Statistics*, 2nd ed. Dover, New York.
- SEBASTIANI, P. (1996a). Discussion of the papers by Atkinson (1996) and by Bates et al. (1996). *J. Roy. Statist. Soc. Ser. B* **58** 96–97.
- SEBASTIANI, P. (1996b). On the derivatives of matrix powers. *SIAM J. Matrix Anal. Appl.* **17** 640–648.
- SILVEY, S. D. (1980). *Optimal Design*. Chapman and Hall, London.
- STONE, M. (1959). Application of a measure of information to the design and comparison of regression experiments. *Ann. Math. Statist.* **30** 55–70.
- VERDINELLI, I. (1996). Bayesian design for the normal linear model with unknown error variance. Technical Report 647, Carnegie Mellon Univ.
- VERDINELLI, I. and KADANE, J. B. (1992). Bayesian design for maximizing information and outcome. *J. Amer. Statist. Assoc.* **87** 510–515.
- WHITTLE, P. (1973). Some general points on the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35** 123–130.
- ZACKS, S. (1977). Problems and approaches in design of experiments for estimation and testing in nonlinear models. In *Multivariate Analysis IV* (P. R. Krishnaiah, ed.). North-Holland, Amsterdam.

DEPARTMENT OF STATISTICAL SCIENCE  
 UNIVERSITY COLLEGE LONDON  
 LONDON, WC1E 6BT UNITED KINGDOM  
 E-MAIL: dawid@stats.ucl.ac.uk

STATISTICS DEPARTMENT  
 THE OPEN UNIVERSITY  
 MILTON KEYNES, MK7 6AA  
 UNITED KINGDOM  
 E-MAIL: p.sebastiani@open.ac.uk