Ann. Inst. Statist. Math.
Vol. 47, No. 2, 385–399 (1995)

DETERMINANT FORMULAS WITH APPLICATIONS TO DESIGNING WHEN THE OBSERVATIONS ARE CORRELATED*

WOLFGANG BISCHOFF

Institute of Mathematical Stochastics, Department of Mathematics, University of Karlsruhe, D-76128 Karlsruhe, Germany

(Received June 13, 1994; revised December 5, 1994)

Abstract. In the general linear model consider the designing problem for the Gauß-Markov estimator or for the least squares estimator when the observations are correlated. Determinant formulas are proved being useful for the D-criterion. They allow, for example, a (nearly) elementary proof and a generalization of recent results for an important linear model with multiple response. In the second part of the paper the determinant formulas are used for deriving lower bounds for the efficiency of a design. These bounds are applied in examples for tridiagonal covariance matrices. For these examples maximin designs are determined.

Key words and phrases: Determinant formula, general linear model, correlated observations, D-criterion, efficiency of designs, linear model with multiple response, lower bounds for the efficiency, tridiagonal matrices as covariance structure, maximin designs.

1. Introduction, notations and preliminary results

Consider a general linear model

$$Y = X\beta + Z$$

where X is a known real $(n \times m)$-matrix, $n \geq m$, $\beta \in \mathbb{R}^m$ is an unknown parameter vector and Z is an n-dimensional real random vector with

$$\mathrm{E}Z = \mathbf{0}_n = (0, \dots, 0)^\top \in \mathbb{R}^n, \qquad \operatorname{Cov} Z = C \quad \text{positive definite.}$$

In this paper we are interested in estimating β. In the first instance we assume that C is known. For β being estimable X must be of full rank, that is, $\operatorname{rank}(X) = m$. We consider two estimators mostly used for estimating β. Firstly the best linear unbiased estimator, the so-called Gauß-Markov estimator

$$(X^\top C^{-1} X)^{-1} X^\top C^{-1} : \mathbb{R}^n \to \mathbb{R}^m$$

---

* Parts of the paper are based on a part of the author's Habilitationsschrift Bischoff (1993a).

having covariance matrix $(X^\top C^{-1}X)^{-1}$, secondly the ordinary least squares estimator

$$(X^\top X)^{-1}X^\top : \mathbb{R}^n \to \mathbb{R}^m,$$

having covariance matrix $(X^\top X)^{-1}X^\top CX(X^\top X)^{-1}$. It is well-known that range$(CX) =$ range(X) is a necessary and sufficient condition such that the Gauß-Markov estimator and the least squares estimator coincide; see for example Zyskind (1967) or Kruskal (1968), see also Rao (1967). Note, we do not distinguish between a linear mapping from $\mathbb{R}^i$ to $\mathbb{R}^j$ and its unique matrix representation with respect to the standard bases of unit vectors.

If designing for the general linear model is of interest the model matrix X is determined by choosing an (exact) design $\tau \in \mathcal{E}^n$ where $\mathcal{E}$ is the experimental region. To emphasize the dependence on τ we write X_τ. Then the class $\mathcal{E}_n$ of feasible designs for estimating β is given by $\mathcal{E}_n = \{\tau \in \mathcal{E}^n : \text{rank}(X_\tau) = m\}$. So, for τ ranging over $\mathcal{E}_n$, we have a class of linear models which we denote by $LM(X_\tau, C : \tau \in \mathcal{E}_n)$.

Almost all optimality criteria in design theory depend on the covariance matrix of the Gauß-Markov estimator. An optimal design $\tau \in \mathcal{E}_n$ minimizes an appropriate functional of the corresponding covariance matrix or equivalently maximizes an appropriate functional of the inverse of the covariance matrix; see Pukelsheim ((1993), Sections 5 and 6). One of the more commonly used criteria for choosing a design $\tau \in \mathcal{E}_n$ is the famous D-optimality criterion. A design $\tau^* \in \mathcal{E}_n$ is called D-optimal (for $LM(X_\tau, C : \tau \in \mathcal{E}_n)$) if

$$\det(X_\tau^\top C^{-1}X_\tau) \le \det(X_{\tau*}^\top C^{-1}X_{\tau*})$$

for all $\tau \in \mathcal{E}_n$. The statistical aim of the paper is to give a general concept how one can tackle the problem of finding optimal or efficient designs with respect to the D-criterion when observations are correlated.

For special factorial linear models with correlated observations exact optimal designs are known; see for example, Kiefer and Wynn (1983, 1984), Budde (1984), Kunert and Martin (1987), and the references cited there.

Only little is known on optimal designs of regression models when the observations are correlated; see Bischoff (1992) and the references cited there. Recently, Bischoff (1992, 1993*b*) has stated conditions such that a D-optimal design for uncorrelated observations with common variance is also D-optimal for correlated observations.

Because it is difficult to determine optimal designs for a linear model with correlated observations, a hybrid approach is often chosen: namely to look for an optimal design not in the class of all possible designs $\mathcal{E}_n$ but only in the class of all designs which are optimal for the uncorrelated case with common variance; see Kiefer and Wynn (1981) and the literature cited there, see also Budde (1984). We explain the above approach in more detail because it may be used in two different ways. However, we consider these two approaches for the D-criterion only. To this end let the class of linear models $LM(X_\tau, C : \tau \in \mathcal{E}_n)$ be given.

Firstly an optimal design $\tau_0 \in \mathcal{E}_n$ in the above sense minimizes the determinant of the covariance matrix of the ordinary least squares estimator in the class of all designs of $\mathcal{E}_n$ being D-optimal for $LM(X_\tau, I_n : \tau \in \mathcal{E}_n)$. That means that $\tau_0 \in \mathcal{E}_n$ fulfills the conditions

$$\max_{\zeta \in \mathcal{E}_n} \det(X_\zeta^\top X_\zeta) = \det(X_{\tau_0}^\top X_{\tau_0})$$

and

$$\det(X_\tau^\top C X_\tau) \geq \det(X_{\tau_0}^\top C X_{\tau_0})$$

for every $\tau \in \mathcal{E}_n$ satisfying

$$\max_{\zeta \in \mathcal{E}_n} \det(X_\zeta^\top X_\zeta) = \det(X_\tau^\top X_\tau).$$

Kiefer and Wynn (1981) investigated and gave reasons for that approach.

Secondly the Gauß-Markov estimator may be used for estimating β. Then an optimal design $\tau_0 \in \mathcal{E}_n$ in the above sense minimizes the determinant of the covariance matrix of the Gauß-Markov estimator in the class of all designs of $\mathcal{E}_n$ being D-optimal for $LM(X_\tau, I_n : \tau \in \mathcal{E}_n)$. That means that $\tau_0 \in \mathcal{E}_n$ fulfills the conditions

$$\max_{\zeta \in \mathcal{E}_n} \det(X_\zeta^\top X_\zeta) = \det(X_{\tau_0}^\top X_{\tau_0})$$

and

$$\det(X_\tau^\top C^{-1} X_\tau) \leq \det(X_{\tau_0}^\top C^{-1} X_{\tau_0})$$

for every $\tau \in \mathcal{E}_n$ satisfying

$$\max_{\zeta \in \mathcal{E}_n} \det(X_\zeta^\top X_\zeta) = \det(X_\tau^\top X_\tau).$$

But on the other hand the question arises. How efficient is an (in the above sense) optimal design τ_0 for the given linear model with correlated observations among all designs $\tau \in \mathcal{E}_n$?

Therefore if the Gauß-Markov estimator is considered let us define the efficiency with respect to the D-criterion of an arbitrary design $\tau \in \mathcal{E}_n$ in the class of linear models $LM(X_\tau, C : \tau \in \mathcal{E}_n)$ by

$$\mathrm{eff}(\tau) = \mathrm{eff}(\tau, LM(X_\zeta, C : \zeta \in \mathcal{E}_n)) := \left[\frac{\det(X_\tau^\top C^{-1} X_\tau)}{\sup_{\zeta \in \mathcal{E}_n} \det(X_\zeta^\top C^{-1} X_\zeta)}\right]^{1/m};$$

see Pukelsheim ((1993), Sections 5.15 and 6.2).

If the structure of the covariance matrix is not known to be sufficiently regular for calculating the Gauß-Markov estimator, then it is common practice to use the ordinary least squares estimator. (See also the approach of Kiefer and Wynn described above.) Using the ordinary least squares estimator the efficiency with respect to the D-criterion of an arbitrary design $\tau \in \mathcal{E}_n$ in the class of linear models $LM(X_\tau, C : \tau \in \mathcal{E}_n)$ is given by

$$\mathrm{eff}_{\ell s}(\tau) = \mathrm{eff}_{\ell s}(\tau, LM(X_\zeta, C : \zeta \in \mathcal{E}_n)) = \left[\frac{\det(X_\tau^\top X_\tau)^2 \cdot \det(X_\tau^\top C X_\tau)^{-1}}{\sup_{\zeta \in \mathcal{E}_n} \det(X_\zeta^\top C^{-1} X_\zeta)}\right]^{1/m}.$$

It is worth mentioning that the Gauß-Markov Theorem states that given a fixed design τ the Gauß-Markov estimator is the only best linear unbiased estimator of β. Therefore we have $\mathrm{eff}(\tau) \geq \mathrm{eff}_{\ell s}(\tau)$ with equality if and only if the Gauß-Markov estimator and the ordinary least squares estimator coincide; for the function $\det(\cdot)$ is strictly isotonic on the set of all positive definite $(m \times m)$-matrices with respect to the Loewner ordering.

In Section 2 determinant formulas are developed. We use these formulas in Section 3 to give general formulas for the efficiency of an arbitrary design $\tau \in \mathcal{E}_n$, see Theorem 3.1. Then by these means we can generalize and provide yet easier evidences for statements that have been recently proven by Bischoff (1992, 1993*b*). Especially, Theorem 3.1 is applied to an important general linear model having multiple response. This allows a (nearly) elementary proof and a generalization of recent results given in Krafft and Schaefer (1992).

In Section 4 general lower bounds for $\mathrm{eff}(\tau)$ and $\mathrm{eff}_{\ell s}(\tau)$ are developed.

Finally, in Section 5 we show by examples how good designs may be found with the help of the developed results. There maximin designs are determined for tridiagonal covariance matrices. These are designs maximizing the minimal efficiency where the minimum is taken over all possible covariance matrices and the maximum is taken over all feasible designs.

We like to mention that lower bounds for $\mathrm{eff}(\tau)$ and $\mathrm{eff}_{\ell s}(\tau)$ are also investigated in Bischoff (1994) in case the models have an intercept term. Further symmetrical circulants are considered as special covariance structure there.

2. Determinant formulas

Let D be a nonnegative definite real $(n \times n)$-matrix, and let V be an m-dimensional subspace of $\mathbb{R}^n$ with $1 \leq m \leq n$. Next, we consider an arbitrary isometric mapping $\phi : V \to \mathbb{R}^m$ with respect to the standard scalar products. Then a unique nonnegative definite real $(m \times m)$-matrix $\tilde{D}$ with respect to the standard basis of unit vectors of $\mathbb{R}^m$ exists such that

$$\phi(v)^\top \tilde{D} \phi(w) = v^\top D w \quad \text{for all} \quad v, w \in V.$$

Note, the eigenvalues of $\tilde{D}$ only depend on D and on V but not on ϕ. Therefore we may denote the m eigenvalues of $\tilde{D}$ by $\lambda_1(D;V), \ldots, \lambda_m(D;V)$. In the sequel we write for shortness $\lambda_i(D;X)$ instead of $\lambda_i(D;\mathrm{range}(X))$ when X is a real $(n \times m)$-matrix. The above discussion suggests the notation

$$\det(D;V) := \det \tilde{D} = \Pi_{i=1}^m \lambda_i(D;V)$$

that we like to use in the following. Analogously as above we write $\det(D;X)$ instead of $\det(D;\mathrm{range}(X))$. Because $\det(D;V)$ is the decisive notion in the sequel we like to mention two further interpretation of $\det(D;V)$.

Firstly, let E be the possibly degenerated ellipsoid corresponding to D, that is

$$E = \{x \in \mathbb{R}^n \mid x^\top D x \leq 1\}.$$

Then $E \cap V$ is a possibly degenerated ellipsoid in V. Thus, if $E \cap V$ is not degenerated, then $\det(D; V)$ equals $(a_1 \cdots a_m)^{-2}$ where $a_1, \dots, a_m$ are the lengths of the semi-axes of $E \cap V$; in case $E \cap V$ is degenerated then $\det(D; V) = 0$. That means $\det(D; V)$ is the determinant of a nonnegative definite matrix describing $E \cap V$ with respect to the standard inner product in an analogous way as D describes E.

Secondly, let us consider a coordinate free approach. For that let $V^\perp$ be the orthogonal subspace of V with respect to the standard inner product in $\mathbb{R}^n$, that means

$$V^\perp = \{x \in \mathbb{R}^n \mid x^\top v = 0 \text{ for all } v \in V\}.$$

Then the matrix D may be understood as nonnegative definite linear operator from $V \oplus V^\perp$ to $V \oplus V^\perp$. Thus D can be uniquely partitioned in the two nonnegative definite linear operators $D_{11} : V \to V$, $D_{22} : V^\perp \to V^\perp$ and in the linear operator $D_{12} : V^\perp \to V$ satisfying for all $(v, w) \in V \oplus V^\perp$

$$D(v, w) = (D_{11}v + D_{12}w, D_{22}w + D_{12}'v) \in V \oplus V^\perp$$

where D_{12}' is the adjoint of D_{12}; see Eaton ((1983), Proposition 2.15). Obviously, we have $\det(D; V) = \det D_{11}$. The operator D_{11} was investigated by Halmos (1950); but his studies do not overlap this paper.

Taking into account the above discussion we obtain for a nonnegative definite $(n \times n)$-matrix D and an $(n \times m)$-matrix B with $B^\top B = I_m$:

$$\det(D; B) = \det(B^\top D B).$$

The above formula is used in the proof of the following crucial lemma.

LEMMA 2.1. *Let A be an arbitrary $(n \times m)$-matrix, and let D be a nonnegative definite $(n \times n)$-matrix. Then*

$$\det(A^\top D A) = \det(A^\top A) \cdot \det(D; A).$$

PROOF. Let B be an $(n \times m)$-matrix fulfilling $B^\top B = I_m$ and $\operatorname{range}(A) = \operatorname{range}(B)$. So a real $(m \times m)$-matrix L exists with $A = BL$. Thus the assertion follows easily. □

Remark 2.1. The above lemma may also be proved in a more technical way by a QR-decomposition of A.

Remark 2.2. In Section 1 the fact was shown that the Gauß-Markov Theorem implies the inequality $\operatorname{eff}(\tau) \geq \operatorname{eff}_{\ell s}(\tau)$. The equality holds true if and only if $\operatorname{range}(CX_\tau) = \operatorname{range}(X_\tau)$. Therefore Lemma 2.1 implies for an arbitrary subspace $V \neq \{\mathbf{0}_n\}$ of $\mathbb{R}^n$ that

$$\det(C^{-1}; V) \geq \det(C; V)^{-1}$$

with equality if and only if $C(V) = V$.

Next, have a look to the coordinate free approach described before Lemma 2.1. Let again $V = \text{range}(X)$, and let $C(v,w) = (C_{11}v + C_{12}w, C_{22}w + C'_{12}v) \in V \oplus V^{\perp}$, $v \in V$, $w \in V^{\perp}$. Then it may be easily verified that the inverse of C can be uniquely partitioned in the two nonnegative definite linear operators $[C_{11} - C_{12}C_{22}^{-1}C'_{12}]^{-1} : V \to V$, $[C_{22} - C'_{12}C_{11}^{-1}C_{12}]^{-1} : V^{\perp} \to V^{\perp}$ and in the linear operator $-C_{11}^{-1}C_{12}[C_{22} - C'_{12}C_{11}^{-1}C_{12}]^{-1} : V^{\perp} \to V$. Thus

$$\det(C^{-1}; V) = \det(C_{11} - C_{12}C_{22}^{-1}C'_{12})^{-1}. \tag{2.1}$$

Further a suitable $(n \times n)$-matrix Γ exists with $\Gamma^{\top}\Gamma = I_n$ and

$$\Gamma C \Gamma^{\top} = \begin{pmatrix} C_{11} & C_{12} \\ C_{12}^{\top} & C_{22} \end{pmatrix}$$

where C_{ij} has the same meaning as above; note we make no distinction between the coordinate free expression of an operator and a suitable matrix expression of it. By a well-known determinant formula for partitioned matrices we get

$$\det C = \det(C_{22}) \cdot \det(C_{11} - C_{12}C_{22}^{-1}C_{12}^{\top}).$$

Thus we obtain the following lemma by (2.1).

LEMMA 2.2. *With the above notation we have*

$$\det(C^{-1}; V) = \det C_{22} \cdot \det C^{-1} = \det(C; V^{\perp}) \cdot \det C^{-1}.$$

3. Efficiency of designs with respect to the D-criterion

As an immediate consequence of Section 2 we obtain the following theorem. There we use the abbreviation $\det(C; X_\tau^{\perp})$ for $\det(C; \text{range}(X_\tau)^{\perp})$.

THEOREM 3.1. *Let the class of linear models $LM(X_\tau, C : \tau \in \mathcal{E}_n)$ be given, and let $\tau \in \mathcal{E}_n$ be arbitrary. Then the following equality holds true*

$$\begin{aligned} \det(X_\tau^{\top} C^{-1} X_\tau) &= \det(X_\tau^{\top} X_\tau) \cdot \det(C^{-1}; X_\tau) \\ &= \det(X_\tau^{\top} X_\tau) \cdot \det(C; X_\tau^{\perp}) \cdot \det C^{-1}. \end{aligned}$$

Thus we get

$$\begin{aligned} \text{eff}(\tau) = \text{eff}(\tau, LM(X_\zeta, C : \zeta \in \mathcal{E}_n)) &= \left[\frac{\det(X_\tau^{\top} X_\tau) \cdot \det(C^{-1}; X_\tau)}{\sup_{\zeta \in \mathcal{E}_n} \det(X_\zeta^{\top} X_\zeta) \cdot \det(C^{-1}; X_\zeta)}\right]^{1/m} \\ &= \left[\frac{\det(X_\tau^{\top} X_\tau) \cdot \det(C; X_\tau^{\perp})}{\sup_{\zeta \in \mathcal{E}_n} \det(X_\zeta^{\top} X_\zeta) \cdot \det(C; X_\zeta^{\perp})}\right]^{1/m} \end{aligned}$$

and

$$
\begin{aligned}
\operatorname{eff}_{\ell s}(\tau) &= \operatorname{eff}_{\ell s}(\tau, LM(X_\zeta, C : \zeta \in \mathcal{E}_n)) \\
&= \left[\frac{\det(X_\tau^\top X_\tau)\cdot \det(C; X_\tau)^{-1}}{\sup_{\zeta\in\mathcal{E}_n} \det(X_\zeta^\top X_\zeta)\cdot \det(C^{-1}; X_\zeta)}\right]^{1/m} \\
&= \left[\frac{\det(X_\tau^\top X_\tau)\cdot \det(C^{-1}; X_\tau^\perp)^{-1}}{\sup_{\zeta\in\mathcal{E}_n} \det(X_\zeta^\top X_\zeta)\cdot \det(C; X_\zeta^\perp)}\right]^{1/m}.
\end{aligned}
$$

Remark 3.1. The above determinant formulas have some advantages. Firstly, we need not to calculate the inverse of the covariance matrix C. (Note, the covariance matrices considered in the literature depend on an unknown parameter. So a closed formula for the inverse of the covariance matrices does not necessarily exist.) Secondly, the above equation gives a relation between the designing problem of the uncorrelated homoscedastic case (usually considered) and the designing problem with correlated observations; see also Section 4.

Remark 3.2. The above result like the most of the following results may be generalized to covariance matrices depending on τ. But for shortness we mostly restrict ourself to covariance matrices not depending on τ.

3.1 *First applications*

Next, we use the above theorem for determining the efficiency of designs when the covariance matrix of Z is given by $\sigma^2 C + \rho \mathbf{1}_n \mathbf{1}_n^\top$ instead by C where $\mathbf{1}_n = (1, \dots, 1)^\top \in \mathbb{R}^n$. In particular, the class of completely symmetric covariance matrices is covered. For instance, such a model is used for body surface potential mappings (BSPM). There the electrical potential of the heart is measured at different points on the surface of a human body. The measurements are taken at the same point of time. BSPMs are described in detail in Bischoff *et al.* (1987, 1990).

COROLLARY 3.1. *Let C be an arbitrary positive definite $(n \times n)$-matrix, let a class of linear models $LM(X_\zeta, C : \zeta \in \mathcal{E}_n)$ be given, let $\tau \in \mathcal{E}_n$ be fixed, let the constants $\sigma^2 > 0$, $\rho > -\sigma^2 \mathbf{1}_n^\top C \mathbf{1}_n / n^2$ are known or unknown, and let the following condition be satisfied*

$$\mathbf{1}_n \in \operatorname{range}(X_\zeta) \quad \textit{for each} \quad \zeta \in \mathcal{E}_n.$$

a) *Then*

$$\operatorname{eff}(\tau, LM(X_\zeta, C : \zeta \in \mathcal{E}_n)) = \operatorname{eff}(\tau, LM(X_\zeta, \sigma^2 C + \rho \mathbf{1}_n \mathbf{1}_n^\top : \zeta \in \mathcal{E}_n)).$$

b) *If additionally $\mathbf{1}_n$ is an eigenvector of C, then*

$$\operatorname{eff}_{\ell s}(\tau, LM(X_\zeta, C : \zeta \in \mathcal{E}_n)) = \operatorname{eff}_{\ell s}(\tau, LM(X_\zeta, \sigma^2 C + \rho \mathbf{1}_n \mathbf{1}_n^\top : \zeta \in \mathcal{E}_n)).$$

PROOF. a) The assertion follows because $\det(\sigma^2 C + \rho \mathbf{1}_n \mathbf{1}_n^\top; X_\tau^\perp) = \det(\sigma^2 C; X_\tau^\perp)$.

b) Because of $\det((\sigma^2 C + \rho \mathbf{1}_n \mathbf{1}_n^\top)^{-1}; X_\tau^\perp) = \det(\sigma^{-2} C^{-1}; X_\tau^\perp)$ assertion b) holds true. □

Remark 3.3. Part a) of the above corollary may be obviously generalized in following way. Let C_τ be an arbitrary covariance matrix for each $\tau \in \mathcal{E}_n$ fulfilling $C_\tau u = Cu$ for each $u \in \text{range}(X_\tau)^\perp$. Then we have

$$\text{eff}(\tau; LM(X_\tau, C_\tau; \tau \in \mathcal{E}_n)) = \text{eff}(\tau; LM(X_\tau, C : \tau \in \mathcal{E}_n)).$$

In an analogous way part b) of the above corollary may be generalized, too.

Remark 3.4. The theory of approximate designs is the key idea for solving designing problems in regression models with uncorrelated observations and common variance. Note that in the uncorrelated case with common variance the connection between exact designs and approximate designs is given by the formula

$$\frac{1}{n} X_\tau^\top X_\tau = \int_{\mathcal{E}} f(t) f(t)^\top \nu(dt)$$

where $X_\tau = (f(t_1) \mid \cdots \mid f(t_n))^\top$, $\tau = (t_1, \ldots, t_n)^\top \in \mathcal{E}_n$, and ν is a suitable probability measure on $\mathcal{E}$. But an analogous relation does not exist for the correlated case in general. In practice on the other hand an approximate optimal design is only used as basis for choosing an efficient design; see, for instance, Pukelsheim and Rieder (1992). Therefore if we have

$$\text{eff}(\tau, LM(X_\zeta, I_n : \zeta \in \mathcal{E}_n)) = \text{eff}(\tau, LM(X_\zeta, C : \zeta \in \mathcal{E}_n))$$

for all $\tau \in \mathcal{E}_n$, or more generally if the equivalence

$$\det(X_\tau^\top X_\tau) \geq \det(X_\zeta^\top X_\zeta) \Longleftrightarrow \det(X_\tau^\top C^{-1} X_\tau) \geq \det(X_\zeta^\top C^{-1} X_\zeta)$$
$$(\Longleftrightarrow \det(X_\tau^\top X_\tau) \cdot \det(C; X_\tau^\perp) \geq \det(X_\zeta^\top X_\zeta) \cdot \det(C; X_\zeta^\perp))$$

holds true for all $\tau, \zeta \in \mathcal{E}_n$, then an approximate D-optimal design for $LM(X_\tau, I_n : \tau \in \mathcal{E}_n)$ may be considered as approximate D-optimal for $LM(X_\tau, C : \tau \in \mathcal{E}_n)$, too. Thus Corollary 3.1 may also be interpreted in such a way. These considerations are also important for the model considered next.

3.2 *Applications to a regression model with multiple response*

Next we apply Theorem 3.1 to a linear regression model with multiple response considered recently by Krafft and Schaefer (1992). They have shown that a D-optimal design for uncorrelated observations with common variance remains D-optimal for special covariance matrices. This result was generalized to several directions and the proof was simplified under less stringent assumptions by Bischoff (1993*b*). By the technique developed here the proof can be further simplified and the result can be generalized to least squares estimation. To this end the model is introduced briefly.

For simplification we use the same notation as Krafft and Schaefer (1992). Let $f_j^{(i)}$, $1 \le j \le r_i$, $1 \le i \le m$, be known real functions defined on an experimental region χ and $a_j^{(i)}$ be unknown parameters to be estimated. For points $x_1, x_2, \ldots, x_n \in \chi$ one observes random variables $Y_i(x_k)$ satisfying

$$\mathrm{E}Y_i(x_k) = \Sigma_{j=1}^{r_i} a_j^{(i)} \cdot f_j^{(i)}(x_k)$$

for $1 \le k \le n$, $1 \le i \le m$. To write this model in terms of matrices, we put for $1 \le i \le m$

$$Y_i = (Y_i(x_1), Y_i(x_2), \ldots, Y_i(x_n))^\top, \qquad a^{(i)} = (a_1^{(i)}, a_2^{(i)}, \ldots, a_{r_i}^{(i)})^\top,$$

$$F_i = \begin{pmatrix} f_1^{(i)}(x_1) & \cdots & f_{r_i}^{(i)}(x_1) \\ \vdots & & \vdots \\ f_1^{(i)}(x_n) & \cdots & f_{r_i}^{(i)}(x_n) \end{pmatrix}$$

and

$$\boldsymbol{Y} = (Y_1^\top, Y_2^\top, \ldots, Y_m^\top)^\top, \qquad \boldsymbol{a} = (a^{(1)^\top}, a^{(2)^\top}, \ldots, a^{(m)^\top})^\top,$$

$$X = \begin{pmatrix} F_1 & 0 & \cdots & 0 \\ 0 & F_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & F_m \end{pmatrix} = \mathrm{diag}(F_1, F_2, \ldots, F_m).$$

Then the above model can be written as $\mathrm{E}\,\boldsymbol{Y} = X\boldsymbol{a}$. Additionally, it is assumed that Cov $\boldsymbol{Y}$ exists and is positive definite. This model just described we shall refer to as regression model with multiple response. For more details on that model see Krafft and Schaefer (1992) and Bischoff (1993*b*).

A typical design matrix X_τ of the regression model with multiple response is given by $\mathrm{diag}(F_{1,\tau}, F_{2,\tau}, \ldots, F_{m,\tau})$ where $\tau \in \chi^n$. An exact design $\tau \in \chi^n$ is feasible for estimating $\boldsymbol{a}$ if and only if $F_{i,\tau}$ has full rank for every $i \in \{1, \ldots, m\}$. Let the set of such designs be denoted by χ_n. For further considerations we assume

$$\mathrm{range}(F_{1,\tau}) \supseteq \mathrm{range}(F_{2,\tau}) \supseteq \cdots \supseteq \mathrm{range}(F_{m,\tau}) \tag{3.1}$$

for each $\tau \in \chi_n$. For instance, condition (3.1) is fulfilled, if each response belongs to a polynomial regression.

In order to state the next result we need the notation of a Kronecker-product. Let $A = (a_{ij})_{1 \le i \le s, 1 \le j \le t}$ and B be arbitrary matrices then the notation $A \otimes B$ is used for the Kronecker-product of A with B, that is $A \otimes B = (a_{ij} \cdot B)_{1 \le i \le s, 1 \le j \le t}$. It is worth noting that with the above notation the set of all exact designs for which $\boldsymbol{a}$ is estimable is given by

$$\mathcal{E}_{m \cdot n} = \{\mathbf{1}_m \otimes \tau \mid \tau \in \chi_n\}.$$

Obviously $\mathcal{E}_{m \cdot n}$ may be indentified with χ_n. By using Theorem 3.1 we can generalize Theorem 1 of Krafft and Schaefer (1992) and Theorem 3.1 of Bischoff (1993*b*). Moreover, the proof is much simpler than the corresponding ones.

THEOREM 3.2. *Let the regression model with multiple response $LM(X_\tau, I_{n\cdot m} : \tau \in \chi_n)$ be given where $X_\tau = \operatorname{diag}(F_{1,\tau}, \dots, F_{m,\tau})$, $\tau \in \chi_n$, let condition (3.1) be satisfied, and let D be an arbitrary positive definite $(m \times m)$-matrix. Then $\det(D \otimes I_n; X_\tau)$ is constant, that means the determinant is independent of $\tau \in \chi_n$.*

Thus the following equalities hold true for each $\tau \in \chi_n$:

$$\operatorname{eff}(\tau; LM(X_\zeta, I_{nm} : \zeta \in \chi_n)) = \operatorname{eff}(\tau; LM(X_\zeta, D \otimes I_n : \zeta \in \chi_n))$$

and

$$\operatorname{eff}_{\ell s}(\tau; LM(X_\zeta, I_{nm} : \zeta \in \chi_n)) = c \cdot \operatorname{eff}_{\ell s}(\tau; LM(X_\zeta, D \otimes I_n : \zeta \in \chi_n))$$

where $c \geq 1$ is a constant independent of τ.

PROOF. Because of (3.1) we can choose $(n \times r_i)$-matrices $B_{i,\tau}$, $i = 1, \dots, m$, fulfilling for $m \geq i \geq j \geq 1$:

$$\operatorname{range}(B_{i,\tau}) = \operatorname{range}(F_{i,\tau}), \qquad B_{i,\tau}^\top B_{j,\tau} = (I_{r_i} \mid \mathbf{0}_{r_i, r_j - r_i})$$

where $\mathbf{0}_{p,q}$ is the $(p \times q)$-null-matrix. Then

$$\det(\operatorname{diag}(B_{1,\tau}, \dots, B_{m,\tau})^\top (D \otimes I_n) \operatorname{diag}(B_{1,\tau}, \dots, B_{m,\tau}))$$

is independent of $\tau \in \chi_n$ for each positive definite $(m \times m)$-matrix D. Thus the assertions follow by Theorem 3.1 and Remark 2.2. □

Remark 3.5. The covariance structure considered in Theorem 3.2 is quite natural. For if the i-th experiment is performed under the control variable x_i we observe the response $(Y_1(x_i), \dots, Y_m(x_i))$ where $Y_j(x_i)$ belongs to the j-th response. Thus it would be artifical to assume that $\operatorname{Cov}((Y_1(x_i), \dots, Y_m(x_i))^\top) = \sigma^2 I_m$ but it is quite natural that this random vector is correlated because the observations are taken from the same experiment. On the other hand it may be possible that the observations of different experiments are uncorrelated and that $\operatorname{Cov}((Y_1(x_i), \dots, Y_m(x_i))^\top) = D$ for $i = 1, \dots, n$. Therefore the models considered in Theorem 3.2 are the natural extensions of the uncorrelated homoscedastic univariate linear model to the linear model with multiple response. For further motivation and results we like to refer to the work of Krafft and Schaefer (1992).

4. Lower bounds for $\operatorname{eff}(\tau)$ and $\operatorname{eff}_{\ell s}(\tau)$

Let a class of linear models $LM(X_\zeta, C : \zeta \in \mathcal{E}_n)$ be given, let $m = \dim(\operatorname{range}(X_\zeta))$, and let $\lambda_1 \leq \cdots \leq \lambda_n$ be the ordered eigenvalues of C. Using a corollary of the Courant-Fischer-Theorem (cf., for instance, Marshall and Olkin (1979), Result A.1.c, p. 510) then Theorem 3.1 implies

$$\begin{aligned}
\operatorname{eff}_{\ell s}(\tau) &\geq \operatorname{eff}(\tau, LM(X_\zeta, I_n : \zeta \in \mathcal{E}_n)) \cdot \det(C; X_\tau)^{-1/m} \cdot \left[\sup_{\zeta \in \mathcal{E}_n} \det(C^{-1}; X_\zeta)\right]^{-1/m} \\
&\geq \operatorname{eff}(\tau, LM(X_\zeta, I_n : \zeta \in \mathcal{E}_n)) \cdot \det(C; X_\tau)^{-1/m} \cdot \Pi_{i=1}^m \lambda_i^{1/m},
\end{aligned}$$

and analogously

$$\text{eff}_{\ell s}(\tau) \geq \text{eff}(\tau, LM(X_\zeta, I_n : \zeta \in \mathcal{E}_n)) \cdot \det(C^{-1}; X_\tau^\perp)^{-1/m} \cdot \Pi_{i=1}^{n-m} \lambda_{m+i}^{-1/m}.$$

Note, in a similar way we get lower bounds for $\text{eff}(\tau)$.

Next, let us consider the principles mentioned in Section 1. First isolate the class of designs for which $\det(X_\tau^\top X_\tau)$ is maximized. Then within this class look for a design which minimizes the determinant of the covariance matrix of the Gauß-Markov estimator and the least squares estimator, respectively. By the above inequalities we recognize that the principles work well if a design τ exists being D-optimal for the uncorrelated homoscedastic case and for which $\det(C^{-1}; X_\tau)$ is about $\Pi_{i=1}^m \lambda_i^{-1}$; that means that the regression manifold corresponding to τ is "nearly" spanned by eigenvectors of the covariance matrix corresponding to the m smallest eigenvalues. But on the other hand if $\det(C^{-1}; X_\tau) \cdot \Pi_{i=1}^m \lambda_i$ is small for each τ being D-optimal for the uncorrelated homoscedastic case, then one should investigate designs ζ for which $\det(C^{-1}; X_\zeta)$ is greater than the corresponding expression for τ.

5. Tridiagonal matrices as covariance structure

In this section we show by examples how good designs may be found. To this end we assume that $C = C_\rho = \text{Cov}(Z)$ is of special structure, namely

$$\text{Cov}(Z_i, Z_j) = \begin{cases} \sigma^2 & \text{if } i = j \\ \sigma^2 \rho & \text{if } i = j+1 \text{ or } i = j-1 \\ 0 & \text{otherwise.} \end{cases}$$

The eigenvalues of $C_\rho = (\text{Cov}(Z_i, Z_j))$ are given by $\lambda_k(\rho) = \sigma^2 \cdot (1 + 2\rho \cdot \cos \frac{k\pi}{n+1})$, $k = 1, \ldots, n$. Without loss of generality we may assume $\sigma^2 = 1$. Thus C_ρ is a covariance matrix if and only if $|\rho| < (2 \cdot \cos \frac{\pi}{n+1})^{-1}$. We assume that ρ is unknown. In order to emphasize the dependence on ρ we write $\text{eff}(\tau; \rho)$ and $\text{eff}_{\ell s}(\tau; \rho)$ in the sequel. In general there does not exist a design which is highly efficient for all ρ. So the following maximin principle seems to be of interest:

$$\max_{\tau \in \mathcal{E}_n} \inf_{\rho \in R} \text{eff}(\tau; \rho)$$

and

$$\max_{\tau \in \mathcal{E}_n} \inf_{\rho \in R} \text{eff}_{\ell s}(\tau; \rho),$$

respectively, where $R \subseteq (-(2\cos \frac{\pi}{n+1})^{-1}, (2\cos \frac{\pi}{n+1})^{-1})$ is the set of possible values for ρ. For shortness we denote designs $\tau^* \in \mathcal{E}_n$ with $\inf_{\rho \in R} \text{eff}_{\ell s}(\tau^*; \rho) = \max_{\tau \in \mathcal{E}_n} \inf_{\rho \in R} \text{eff}_{\ell s}(\tau; \rho)$ maximin designs; if in the definition of maximin designs the maximum is taken over a subset $\tilde{\mathcal{E}}_n$ of $\mathcal{E}_n$ only, then we say maximin designs in $\tilde{\mathcal{E}}_n$. A subset of $\mathcal{E}_n$ being of interest is the set $\mathcal{E}_n(\lambda_i)$ of all designs τ whose regression manifold range(X_τ) contains an eigenvector corresponding to λ_i. In an analogous way $\mathcal{E}_n(\lambda_i, \lambda_j)$ is defined.

We like to mention in passing that, obviously, information on ρ may be used in order to get good designs with respect to the Bayes principle instead of the above maximin principle.

THEOREM 5.1. *Let an arbitrary class $LM(X_\tau, C_\rho : \tau \in \mathcal{E}_n)$, $\rho \in R$, of linear models be given where the covariance matrix C_ρ has the structure given above.*

a) *Let $R = (a, (2\cos\frac{\pi}{n+1})^{-1})$ with $a > -(2\cos\frac{\pi}{n+1})^{-1}$, and suppose that $\mathcal{E}_n(\lambda_n) \neq \emptyset$. Then $\tau \in \mathcal{E}_n$ is a maximin design if and only if τ is a maximin design in $\mathcal{E}_n(\lambda_n)$.*

b) *Let $R = (-(2\cos\frac{\pi}{n+1})^{-1}, a)$ with $a < (2\cos\frac{\pi}{n+1})^{-1}$, and suppose that $\mathcal{E}_n(\lambda_1) \neq \emptyset$. Then $\tau \in \mathcal{E}_n$ is a maximin design if and only if τ is a maximin design in $\mathcal{E}_n(\lambda_1)$.*

c) *Let $R = (-(2\cos\frac{\pi}{n+1})^{-1}, (2\cos\frac{\pi}{n+1})^{-1})$, and suppose that $\mathcal{E}_n(\lambda_1, \lambda_n) \neq \emptyset$. Then $\tau \in \mathcal{E}_n$ is a maximin design if and only if τ is a maximin design in $\mathcal{E}_n(\lambda_1, \lambda_n)$.*

PROOF. a) For shortness we write $\lambda_i(\rho, \tau)$ for the eigenvalue $\lambda_i(C_\rho; X_\tau)$ introduced in Section 2, $i = 1, \ldots, m$, $\tau \in \mathcal{E}_n$. Further let these eigenvalues $\lambda_1(\rho; \tau) \leq \cdots \leq \lambda_m(\rho; \tau)$ be ordered. We consider a design $\tau \in \mathcal{E}_n$ whose regression manifold range(X_τ) contains no eigenvector corresponding to λ_n and a design $\zeta \in \mathcal{E}_n(\lambda_n)$. Then we have for each $\rho > 0$

$$\lambda_n(\rho) < \lambda_1(\rho; \tau) \leq \cdots \leq \lambda_m(\rho; \tau) \leq \lambda_1(\rho)$$

and

$$\lambda_n(\rho) = \lambda_1(\rho; \zeta) \leq \cdots \leq \lambda_m(\rho; \zeta) \leq \lambda_1(\rho).$$

Note, there exists a constant $c(\tau) > 0$ with $\lambda_1(\rho; \tau) \geq c(\tau)$ for all $\rho > 0$. Thus we have

$$\det(C_\rho; X_\tau) = \Pi_{i=1}^m \lambda_i(\rho; \tau) \geq c(\tau)^m \quad \text{for all} \quad \rho > 0,$$
$$\det(C_\rho; X_\zeta) = \Pi_{i=1}^m \lambda_i(\rho; \zeta) \longrightarrow 0 \quad \text{when} \quad \rho \to \left(2\cos\frac{\pi}{n+1}\right)^{-1}$$

whence we get by Theorem 3.1

$$\mathrm{eff}_{\ell s}(\tau; \rho) \leq \left[\frac{\det(X_\tau^\top X_\tau) \cdot \det(C_\rho; X_\tau)^{-1}}{\det(X_\zeta^\top X_\zeta) \cdot \det(C_\rho^{-1}; X_\zeta)}\right]^{1/m} \longrightarrow 0$$
$$\text{when} \quad \rho \to \left(2\cos\frac{\pi}{n+1}\right)^{-1}.$$

On the other hand for ζ a constant $c > 0$ exists with $\inf_{\rho \in R} \mathrm{eff}_{\ell s}(\zeta; \rho) \geq c$ because $\inf_{\rho \in R}(\rho + (2\cos\frac{\pi}{n+1})^{-1}) > 0$. So assertion a) is proved.

In the same way assertions b) and c) may be shown. □

Example. Let us consider a regression model with regression function $f : \mathcal{E} \longrightarrow \mathcal{E}$, $t \mapsto t$ where $\mathcal{E} \subseteq \mathbb{R}$ is the experimental region. Given the design $\tau =$

$(t_1,\dots,t_n)^\top \in \mathcal{E}_n$, $n \geq 2$, we have a linear model $Y = \tau\beta + Z$. We consider the two cases $\mathcal{E} = [-1,1]$ (see a) below) and $\mathcal{E} = [0,1]$ (see b) below) because the results of one case cannot be derived from the other case; for the above model has no intercept term. We assume that ρ is unknown.

a) $\mathcal{E} = [-1,1]$. Let us consider two special designs.

1) Let $\tau_1 = (1,-1,1,-1,\dots)^\top \in \mathcal{E}_n$. Because ρ is unknown we cannot calculate the Gauß-Markov estimator corresponding to τ_1. Therefore we evaluate a lower bound for $\mathrm{eff}_{\ell s}(\tau_1;\rho)$ by the inequality given in Section 4:

$$\mathrm{eff}_{\ell s}(\tau_1;\rho) \geq \left(1 - 2|\rho|\cdot\cos\frac{\pi}{n+1}\right)\cdot\det(C_\rho; X_{\tau_1}) = \left(\frac{1-2|\rho|\cos\dfrac{\pi}{n+1}}{1-\left(2-\dfrac{2}{n}\right)\rho}\right).$$

Note, given $\rho \geq 0$ then τ_1 is the best design for the correlated case in the class of all designs being D-optimal for the homoscedastic case.

2) Let

$$\tau_2 = \begin{cases} \left(\sin\dfrac{\pi}{n+1}, -\sin\dfrac{2\pi}{n+1},\dots,(-1)^{n-1}\sin\dfrac{n\pi}{n+1}\right)^\top \\ \quad\text{if } n \text{ is odd} \\ \left(\sin\dfrac{n\pi}{2(n+1)}\right)^{-1}\left(\sin\dfrac{\pi}{n+1}, -\sin\dfrac{2\pi}{n+1},\dots,(-1)^{n-1}\sin\dfrac{n\pi}{n+1}\right)^\top \\ \quad\text{if } n \text{ is even.} \end{cases}$$

Then τ_2 is an eigenvector corresponding to λ_n. Thus we obtain by the considerations of Section 4

$$\begin{aligned}\mathrm{eff}(\tau_2;\rho) &= \mathrm{eff}_{\ell s}(\tau_2;\rho) \\ &\geq \begin{cases} \dfrac{n+1}{2n}\,\dfrac{1-2|\rho|\cos\dfrac{\pi}{n+1}}{1-2\rho\cos\dfrac{\pi}{n+1}} & \text{if } n \text{ is odd} \\[2ex] \dfrac{n+1}{2n}\cdot\left(\sin\dfrac{n\pi}{2(n+1)}\right)^{-2}\cdot\dfrac{1-2|\rho|\cos\dfrac{\pi}{n+1}}{1-2\rho\cos\dfrac{\pi}{n+1}} & \text{if } n \text{ is even.}\end{cases}\end{aligned}$$

Firstly, we assume that $\rho \geq 0$. Note, $\mathrm{eff}(\tau_2,\rho)$ is greater than $\frac{1}{2}$ independent of ρ. So τ_1 is a highly efficient design if ρ is not too great but τ_1 is getting worse if ρ goes to $(2\cos\frac{\pi}{n+1})^{-1}$. In that case τ_2 is much better. Indeed for fixed $n \geq 2$ the design τ_2 is better than τ_1 if $\rho \in \left(\frac{n-1}{2n-\frac{2}{n}+4n\cos\frac{\pi}{n+1}}, \frac{1}{2\cos\frac{\pi}{n+1}}\right)$. Thus if one cannot exclude values of ρ near $(2\cos\frac{\pi}{n+1})^{-1}$ one prefers the design τ_2 to τ_1 from the above maximin principle. Indeed Theorem 5.1 implies that τ_2 is a maximin design for $R = (a, (2\cos\frac{\pi}{n+1})^{-1})$ with $a > -(2\cos\frac{\pi}{n+1})^{-1}$.

If $\rho < 0$, both designs τ_1 and τ_2 are bad. In that case the designs $\tau_3 = \mathbf{1}_n$ and a suitable eigenvector τ_4, say, corresponding to λ_1 take over the role of τ_1 and τ_2, respectively.

But if nothing is known about ρ that is $\rho \in (-(2\cos\frac{\pi}{n+1})^{-1}, (2\cos\frac{\pi}{n+1})^{-1})$, then each of the above designs is bad for certain values of ρ. Then from the above maximin point of view a design that is a suitable eigenvector corresponding to $\lambda_{n/2}$ if n is even and $\lambda_{(n+1)/2}$ if n is odd, respectively, is better than the above designs.

b) $\mathcal{E} = [0,1]$. Then $\tau_3 = \mathbf{1}_n$ is the only D-optimal design for the homoscedastic case; hence it is the optimal design for the correlated case for each of the principles described in Section 1. But it is really bad for most $\rho > 0$ and τ_2 is better than τ_3 in case ρ is not too small.

If $\rho < 0$, τ_3 is a highly efficient design because the only eigenvectors corresponding to $\mathcal{E}_n$ are eigenvectors corresponding to the eigenvalue λ_1. But such designs are very bad for $\rho < 0$.

Acknowledgements

The author would like to thank an anonymous referee for careful reading of the manuscript and some useful comments which improved the presentation.

References

Bischoff, W. (1992). On exact D-optimal designs for regression models with correlated observations, *Ann. Inst. Statist. Math.*, **44**, 229–238.

Bischoff, W. (1993*a*). On the efficiency of designs with respect to the D-criterion when the observations are correlated, Habilitationsschrift, University of Karlsruhe, Germany.

Bischoff, W. (1993*b*). On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response, *J. Statist. Plann. Inference*, **37**, 69–80.

Bischoff, W. (1994). Lower bounds for the efficiency of designs with respect to the D-criterion when the observations are correlated (submitted).

Bischoff, W., Cremers, H. and Fieger, W. (1987). Optimal regression models in electrocardiography, *Innovation et Technologie en Biologie et Medcine*, **8**, 24–34.

Bischoff, W., Cremers, H. and Fieger, W. (1990). Optimal designs in electrocardiography with non-constant variances, *Z. Angew. Math. Mech.*, **70**, 731–733.

Budde, M. (1984). Optimale Zweifachblockpläne bei seriell korrelierten Fehlern, *Metrika*, **31**, 203–213.

Eaton, M. L. (1983). *Multivariate Statistics (A Vector Space Approach)*, Wiley, New York.

Halmos, P. (1950). Normal dilations and extensions of operators, *Summa Brasiliensis Mathematicae*, **2**, 125–134.

Kiefer, J. and Wynn, H. P. (1981). Optimum balanced block and latin square designs for correlated observations, *Ann. Statist.*, **9**, 737–757.

Kiefer, J. and Wynn, H. P. (1983). Autocorrelation-robust design of experiments, *Scientific Inference and Data Analysis and Robustness*, Academic Press, New York.

Kiefer, J. and Wynn, H. P. (1984). Optimum and minimax exact treatment designs for one-dimensional autoregressive error processes, *Ann. Statist.*, **12**, 414–450.

Krafft, O. and Schaefer, M. (1992). D-optimal designs for a multivariate regression model, *J. Multivariate Anal.*, **42**, 130–140.

Kruskal, W. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach, *Ann. Math. Statist.*, **39**, 70–75.

Kunert, J. and Martin, R. J. (1987). On the optimality of finite Williams II(a) designs, *Ann. Statist.*, **15**, 1604–1628.

Marshall, A. W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York.

Pukelsheim, F. (1993). *Optimal Design of Experiments*, Wiley, New York.

Pukelsheim, F. and Rieder, S. (1992). Efficient rounding of approximate designs, *Biometrika*, **79**, 763–770.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its applications to measurement of signals, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1 (eds. L. M. LeCam and J. Neyman), 355–372, University of California Press, Berkeley.

Zyskind, G. (1967). On canonical forms non-negative covariance matrices and best and simple least squares linear estimators in linear models, *Ann. Math. Statist.*, **38**, 1092–1109.