Statistica Sinica **5**(1995), 499-514

DESIGNING FOR MINIMALLY DEPENDENT OBSERVATIONS

Ben Torsney and Abdulhadi M. Alahmadi

University of Glasgow and King Abdulaziz University

Abstract. The problem of constructing designs to minimize the squared covariance or correlation between the estimates of two linear combinations of the parameters of a linear regression model is first considered. When the minimum is non-zero the covariance criterion can be equivalent to the c-optimal criterion. When the minimum is zero it often can be attained by a class of designs. It is then of interest to optimize a standard criterion over the class. Some analytic and algorithmic results are reported.

Key words and phrases: c-optimality, D_A-optimality, directional derivatives, linear optimality criterion, optimal design, optimization, vertex directional derivatives.

1. Introduction

In this contribution we develop what we believe to be two new items in the realm of optimal linear regression design theory. It is convenient to review this area.

In a linear design problem a general linear model is assumed i.e. we (minimally) assume independence between observations on a response variable y, constant variance σ^2 and the conditional expectation $E(y|\boldsymbol{v}) = \boldsymbol{v}^t\boldsymbol{\theta}$, $\boldsymbol{v} \in \boldsymbol{V}$. Here $\boldsymbol{\theta} \in R^k$ is a vector of unknown parameters and $\boldsymbol{V} = \{\boldsymbol{v} \in R^k : \boldsymbol{v} = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}))^t, \boldsymbol{x} \in \boldsymbol{\chi}\}$ with $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_k(\boldsymbol{x}))^t$ a vector of known regression functions and $\boldsymbol{\chi}$ a practical closed compact design space. Thus $\boldsymbol{V}$ is the image under $\boldsymbol{f}$ of $\boldsymbol{\chi}$ $-$ an induced design space.

In this context an approximate design is characterized by a probability measure p defined on $\boldsymbol{\chi}$ and hence on $\boldsymbol{V}$. In reality these spaces must be discretized. Suppose that $\boldsymbol{V} = \{v_1, v_2, \ldots, v_J\}$. Then p is characterized by a set of weights $p_1, p_2, \ldots, p_J$ satisfying $p_j \geq 0$, $j = 1, 2, \ldots, J$, and $\sum_{j=1}^J p_j = 1$, weight p_j being assigned to v_j. We wish to choose $p = (p_1, p_2, \ldots, p_J)$ optimally. If $\hat{\boldsymbol{\theta}}$ is the least squares estimator of $\boldsymbol{\theta}$ derived from observations obtained under the design p, then $\mathrm{Cov}(\hat{\boldsymbol{\theta}}) \propto M^-(p)$, where $M(p)$ is the $k \times k$ design matrix:

$$M(p) = \sum_{j=1}^J p_j v_j v_j^t = VPV^t,$$

V the $k \times j$ matrix$[v_1, v_2, \ldots, v_J]$ and $P = \mathrm{diag}(p_1, p_2, \ldots, p_J)$.

A good design will make the matrix $M(p)$ "large" in some sense.

Standard choices are designs which maximize $\phi(p) = \psi\{M(p)\}$ where $\psi(M)$ $= \mathrm{logdet}(M)$ (D-optimality; appropriate if there is interest in all the parameters in $\boldsymbol{\theta}$) or $\psi(M) = -\boldsymbol{c}^t M^- \boldsymbol{c}$ for a given vector $\boldsymbol{c}$ (c-optimality; appropriate if there is interest only in $\boldsymbol{c}'\boldsymbol{\theta}$) or $\psi(M) = -\mathrm{Trace}(AM^-A^t)$ and $\psi(M) = -\mathrm{logdet}(AM^-A^t)$ for a given $s \times k$ matrix A, $s < k$ (Linear optimality and D_A-optimality respectively; appropriate if there is interest in inference only for $A\boldsymbol{\theta}$).

We add to this list by first considering the problem of designing to ensure that two linear combinations of the components of $\boldsymbol{\theta}$ are estimated as independently of each other as possible. Interest in this problem was prompted by a practical problem arising in Chemistry which will be discussed later and was first reported by Torsney (1981). We formalize this problem in the next section. In Section 3 we explore an interesting problem, namely that of maximizing a standard design criterion subject to zero covariance.

2. Designing To Minimize Covariances and Correlations

2.1. New criteria

For given vectors $\boldsymbol{a}, \boldsymbol{b} \in R^k$ we consider designing to minimize numerical covariances or correlations between the estimates of $\boldsymbol{a}^t\boldsymbol{\theta}$ and $\boldsymbol{b}^t\boldsymbol{\theta}$. This leads to the following two problems:

Problem 1. *Maximize* $\phi_c(p) = -[\boldsymbol{a}^t M^-(p)\boldsymbol{b}]^2$ *subject to the constraint* $\sum_{i=1}^J p_i = 1$, $p_i \geq 0,\ i = 1, \ldots, J$, *where* $\boldsymbol{a}, \boldsymbol{b} \in R^k$.

Problem 2. *Maximize* $\phi_\varrho(p) = -[\boldsymbol{a}^t M^-(p)\boldsymbol{b}]^2/([\boldsymbol{a}^t M^-(p)\boldsymbol{a}][\boldsymbol{b}^t M^-(p)\boldsymbol{b}])$ *subject to the constraint* $\sum_{i=1}^J p_i = 1$, $p_i \geq 0$, $i = 1, \ldots J$.

We call these two criteria the covariance criterion and the correlation criterion respectively. These seem to be new criteria except if $\boldsymbol{a} \propto \boldsymbol{b}$, when the covariance criterion is equivalent to a c-optimal criterion and the correlation criterion is constant. The covariance and c-optimal criteria are likely to have similar properties.

2.2. Simplifying the covariance criterion

For simplicity first consider the case $J = k$. Also assume that the matrix $M(p)$ is non-singular . Then the covariance criterion will be of the form:

$$\phi_c(p) = -[\boldsymbol{a}^t M^{-1}(p)\boldsymbol{b}]^2 = -[\boldsymbol{a}^t (VPV^t)^{-1}\boldsymbol{b}]^2,$$

where V as defined in Section 1 is now $k \times k$. Hence we obtain the simplification

$$\phi_c(p) = -[\boldsymbol{c}^t P^{-1} \boldsymbol{d}]^2 = -\Big[\sum_{i=1}^{k} c_i d_i / p_i\Big]^2, \tag{1}$$

where $\boldsymbol{c} = V^{-1}\boldsymbol{a}$ and $\boldsymbol{d} = V^{-1}\boldsymbol{b}$.

Clearly from (1) we can write the covariance criterion as the square of a linear combination of the reciprocals of the weights, the coefficients being the products $c_i d_i$, $i = 1, 2, \ldots, k$. A crucial feature is that these can be positive or negative. We distinguish two cases:
Case a: when all $c_i d_i$, $i = 1, 2, \ldots, k$, have the same sign;
Case b: when the $c_i d_i$, $i = 1, 2, \ldots, k$, have differing signs.

Case a will be considered in Section 2.2.1 and Case b in Section 3.

2.2.1. Explicit solution

Suppose $c_i d_i > 0$, $i = 1, 2, \ldots, k$, i.e. Case a. Then $\phi_c(p)$ is proportional to the square of a positive linear combination of the reciprocals of the weights and so is equivalent to a c-optimal criterion. An explicit solution for the optimal weights is available. It is

$$p_i^* = \sqrt{|c_i d_i|} \Big/ \sum_{j=1}^{k} \sqrt{|c_j d_j|}, \quad i = 1, 2, \ldots, k, \tag{2}$$

which yields a maximum value for (1) of

$$\phi_c^*(p^*) = \Big\{\sum_{i=1}^{k} \sqrt{|c_i d_i|}\Big\}^4. \tag{3}$$

A similar explicit result would be possible if the support points consists of s linearly independent points, $s < k$, (in which case $M(p)$ is singular) as happens for the c-optimal criterion. See Pukelsheim and Torsney (1991).

Note that when all the $c_i d_i < 0$, $i = 1, 2, \ldots, k$, the optimal weights p^* and the maximum value for $\phi_c(p)$ are again given by (2) and (3) respectively.

As an example of this case, consider the quadratic regression model $E(y) = \theta_1 + \theta_2 x + \theta_3 x^2$, $1 \le x \le 2$, suppose we wish to estimate the unknown parameter θ_1 as independently of θ_3 as possible. Thus we want to solve Problem 1 under the choice of $\boldsymbol{a} = (1, 0, 0)^t$ and $\boldsymbol{b} = (0, 0, 1)^t$. Suppose that the support points are $\{1, x_0, 2\}$, $1 < x_0 < 2$. Then

$$(c_1 d_1, c_2 d_2, c_3 d_3)^t = S^2[2x_0(2 - x_0)^2, 2, x_0(x_0 - 1)^2]^t, \tag{4}$$

where $S = 1/\mathrm{Det}(V)$ and $\mathrm{Det}(V) = 3x_0 - x_0^2 - 2$, for the appropriate choice of V of Section 1.

Clearly from (4) all the $c_i d_i$ are greater than zero. Substituting from (4) in (2) we find the optimal weights and their support points to be as recorded in the following table:

Support points	1	x_0	2
Optimal weights	$(2-x_0)\sqrt{2x_0}/w$	$\sqrt{2}/w$	$(x_0-1)\sqrt{x_0}/w$

where $w = (2-x_0)\sqrt{2x_0} + \sqrt{2} + (x_0-1)\sqrt{x_0}$.

With a view to finding the best three point design subject to the endpoints being support points, we must choose x_0 to maximize the criterion $\phi_c^*(p^*)$ in (3) with respect to x_0. The optimal value is $x_0^* = 1.5$.

Similarly, for the same model we have an example for which the $c_i d_i$ are less than zero by taking $\boldsymbol{a} = (0,1,0)^t$ and $\boldsymbol{b} = (0,0,1)^t$. This means we are interested in making the numerical covariance between θ_2 and θ_3 as small as possible. Corresponding results are:

Support points	1	x_0	2
Optimal weights	$(2-x_0)\sqrt{2+x_0}/w$	$\sqrt{3}/w$	$(x_0-1)\sqrt{(x_0+1)}/w$

where $w = (2-x_0)\sqrt{2+x_0} + \sqrt{3} + (x_0-1)\sqrt{x_0+1}$. The optimal choice of the support point x_0 is again $x_0 = 1.5$.

2.3. Algorithm

Apart from the above case there are no other general explicit solutions either for the covariance criterion (particularly if $J > k$) or for the correlation criterion (whatever the value of J). We need an algorithm to determine optimal weights in general, as of course is the case with standard design criteria.

Suppose we wish to maximize a differentiable criterion $\phi(p)$ subject to $p_i \geq 0$, $\sum_{j=1}^J p_j = 1$. One class of iterations for this problem are:

$$p_i^{(r+1)} = \frac{p_i^{(r)} f(d_i^{(r)}, \delta)}{\sum_{j=1}^k p_j^{(r)} f(d_j^{(r)}, \delta)}, \quad i = 1, 2, \ldots, J, \tag{5}$$

where $d_i^{(r)} = \frac{\partial \phi}{\partial p_i}\Big|_{p=p^{(r)}}$ and $f(d,\delta)$ is a function which satisfies the following conditions:

(a) $f(d,\delta) > 0$;

(b) $f(d,\delta)$ is strictly increasing in the scalar variable d for some set of δ-values, say $\delta > 0$;

(c) $f(d, 0) = \text{constant} \neq 0$;

(d) the variable δ is a free parameter.

Under these conditions iterations (5) guarantee that $F_\phi(p^{(r)}, p^{(r+1)}) \geq 0$, where $F_\phi(p^{(r)}, p^{(r+1)})$ is the directional derivative of the criterion $\phi(p)$ at $p^{(r)}$ in the direction of $p^{(r+1)}$. This follows since $F_\phi(p, q) = \sum_{i=1}^J (q_i - p_i) d_i$ ($d_i = \partial\phi/\partial p_i$, $q_i \geq 0$, $\sum_{i=1}^J q_i = 1$) under the differentiability assumption on $\phi(p)$. Hence $F_\phi(p^{(r)}, p^{(r+1)}) = \text{Cov}(D, f(D, \delta))/E\{f(D, \delta)\}$, where D is a random variable satisfying $P(D = d_i^{(r)}) = p_i^{(r)}$. Since $f(d, \delta)$ is postive and increasing in d we have guarantees that the two terms in this ratio are nonnegative.

Further, first order conditions for a local maximum p^* are

$$F_i^* \begin{cases} = 0, & p_i^* > 0, \\ \leq 0, & p_i^* = 0, \end{cases}$$

where $F_i = F_\phi(p, e_i) = d_i - \sum_{j=1}^J p_j d_j$, $d_i = \partial\phi/\partial p_i$, $e_i = i$th unit vector. Hence, derivatives corresponding to nonzero p_i^* must share a common value. Moreover if $p^{(r)} = p^*$ then $p^{(r+1)} = p^{(r)}$ so that $F_\phi(p^{(r)}, p^{(r+1)}) = 0$ and p^* is a fixed point of the iteration.

This type of algorithm was first proposed by Torsney (1977), taking $f(d, \delta) = d^\delta$, $\delta > 0$. Subsequent empirical studies include Silvey et al. (1978), which is a study of the choice of δ when $f(d, \delta) = d^\delta$, and Torsney (1988), which mainly considers $f(d, \delta) = e^{\delta d}$ in a variety of applications, including estimation and image processing problems. Torsney and Alahmadi (1992) continue these investigations exploring other choices of $f(d, \delta)$ for which an approximate optimal δ has been found.

If $\phi(p)$ is the covariance criterion it can have negative derivatives while the correlation criterion always has negative and positive derivatives since, being a homogeneous function of degree zero, $\sum_{i=1}^J p_i(\partial\phi_\varrho/\partial p_i) = 0$. So we need a choice of $f(d, \delta)$ which is defined for negative d. Indeed this was the main reason why Torsney (1988) considered choices of $f(d, \delta)$ such as $e^{\delta d}$. In its original conception algorithm (5) was evolved for standard optimal design criteria which have positive derivatives with the choice of $f(d, \delta) = d^\delta$; in particular $\delta = 1$ for D-optimality and $\delta = 1/2$ for c-optimality and the linear criterion yields monotonic iterations. See Titterington (1976) and Torsney (1983).

For the covariance and correlation criteria we have explored the use of $f(d, \delta) = (1 + sd)^{s\delta}$ where $s = \text{sign}(d)$.

Now consider the Chemical example referred to in Section 1. This concerns the relationship between the Viscosity y and the Concentration x of a chemical solution. The model is $E(y|\boldsymbol{v}) = \boldsymbol{\theta}^t \boldsymbol{v}$; where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^t$, $\boldsymbol{v} = \boldsymbol{f}(x) =$

$(x, x^{1/2}, x^2)^t$ with x a scalar restricted to $(0.0, 0.2]$. It was desired to estimate θ_3 as independently of θ_1 and θ_2 as possible. To encompass the desire to minimize two numerical covariances or correlations, three choices of $\boldsymbol{a}$ were considered, namely $\boldsymbol{a} = (1,0,0)^t$, $\boldsymbol{a} = (0,1,0)^t$ and $\boldsymbol{a} = (-1,1,0)^t$, while $\boldsymbol{b} = (0,0,1)^t$ always.

Consider Problem 1 first. Taking the design interval to be $[0.02, 0.20]$ we believe, from running the algorithm, that the optimal support points are the same for each of our choices of $\boldsymbol{a}$ and $\boldsymbol{b}$, namely $\{.02, .12, .20\}$. Given this, the conditions of Section 2.2.1 are satisfied so that Equation (2) identifies the optimal weights which are listed in Table 1.

Table 1. Optimal weights for different choices of $\boldsymbol{a}$ and $\boldsymbol{b}$ in the case of the covariance criterion for the model $E(y) = \theta_1 x + \theta_2 x^{1/2} + \theta_3 x^2$, $x \in (0.0, 0.2]$.

	Support points			
$\boldsymbol{a}$	.02	.12	.20	$\boldsymbol{a}^t M^{-1}(p)\boldsymbol{b}$
$\boldsymbol{a} = (1,0,0)^t$	.423	.405	.172	−38565.6
$\boldsymbol{a} = (0,1,0)^t$	.509	.347	.144	6909.34
$\boldsymbol{a} = (-1,1,0)^t$	.437	.396	.167	45649.5
	Corresponding optimal weights			

Table 2. The number of iterations needed to achieve $\max F_i \leq 10^{-n}$, $n = 1, 2, 3, 4$ under the choice of $\boldsymbol{a} = (1,0,0)^t$ and $\boldsymbol{b} = (0,0,1)^t$ in the case of the covariance criterion for the above model.

J	δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
19	.95	1/19	637	747	821	889
10	.9	1/10	131	145	158	170
3	.5	1/3	2	2	2	2

Table 3. The number of iterations needed to achieve $\max F_i \leq 10^{-n}$, $n = 1, 2, 3, 4$ under the choice of $\boldsymbol{a} = (0,1,0)^t$ and $\boldsymbol{b} = (0,0,1)^t$ in the case of the covariance criterion for the above model.

J	δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
19	.95	1/19	313	353	393	433
10	.9	1/10	88	98	110	120
3	.5	1/3	2	2	2	2

Table 4. The number of iterations needed to achieve $\max F_i \leq 10^{-n}$, $n = 1,2,3,4$ under the choice of $\boldsymbol{a} = (-1,1,0)^t$ and $\boldsymbol{b} = (0,0,1)^t$ in the case of the covariance criterion for the above model.

J	δ	$p^{(0)}$	$n=1$	$n=2$	$n=3$	$n=4$
19	.95	1/19	603	668	734	809
10	.9	1/10	128	141	154	168
3	.5	1/3	2	2	2	2

In Tables 2 to 4 we record for different choices of $\boldsymbol{a}$ and $\boldsymbol{b}$ the value of δ which attained max $F_i \leq 10^{-4}$ in the smallest number of iterations when $p^{(0)}$ assigns equal weight to J points in $[0.02, 0.2]$ for $J = 19, 10, 3$. Also recorded are the number of iterations needed to achieve max $F_i \leq 10^{-n}$, $n = 1,2,3,4$.

For $J = 19, 10$ these points were equally spaced and therefore included the set $\{.02, .12, .20\}$ which were the three points chosen for $J = 3$. We note that for $J = 19, 10$ algorithm (5) with $f(d,\delta) = (1+sd)^{s\delta}$, $s = \text{sign}(d)$, always converged to the optimal design for these three design points under the above choices of $\boldsymbol{a}$ and $\boldsymbol{b}$.

Clearly the number of iterations needed to achieve max $F_i \leq 10^{-n}$, $n = 1,2,3,4$, depends on the number of design points in the initial design and on the value of δ. For instance, if the initial design consists only of the three support points of the optimal design and if $\delta = .5$ then the optimal design may be obtained in two steps although, of course, we do have an explicit formula in this case. In contrast, when the initial design consists of J points for $J = 10, 19$ a higher value of δ, namely, $\delta = .95, .9$ respectively, attains max $F_i \leq 10^{-4}$ in the smallest number of iterations. In terms of the number of iterations the convergence is slow especially when $J \geq 10$, but it can be improved by setting weights to zero when $p_j < \varepsilon_1$ and $F_j < -\varepsilon_2$ for some small $\varepsilon_1, \varepsilon_2$ or just when p_j goes below a small fixed value ε.

Similar results have been obtained by Fellman (1989) for the c-optimal criterion when $f(d,\delta) = d^\delta$. (Note that $\varphi(p) = -c^T M^-(p)c$, the c-optimal criterion, has the positive derivatives $d_j = (c^T M^-(p)v_j)^2$ if $p_j > 0$, so that d_j^δ is defined). In particular as noted by Torsney (1983) $f(d,\delta) = d^{1/2}$ attains the optimum in one step for the c-optimal criterion when the support points form a linearly independent set of vectors. Clearly $f(d,\delta) = (1+sd)^{s\delta}$, $s = \text{sign}(d)$ has similar effects.

We note finally that it would be unwise to take δ to be large since $p^{(r+1)} \to e_m$ as $\delta \to \infty$ where e_m is the mth unit vector, assuming that $d_m = \partial\phi_c/\partial p_m|_{p=p^{(r)}}$ is a unique maximum derivative at $p^{(r)}$.

Table 5. The Optimal support points and corresponding optimal weights in the case of the correlation criterion for the above model.

	$\boldsymbol{a} = (1,0,0)^t$			$\boldsymbol{a} = (0,1,0)^t$			$\boldsymbol{a} = (-1,1,0)^t$		
supp(p^*)	.02	.12	.2	.02	.14	.2	.02	.12	.12
p^*	.0133	.9845	.0022	.0158	.9810	.00032	.0132	.9849	.0019
$\phi_\rho(p^*)$	$-.8155$			$-.5537$			$-.7755$		

Now consider the example of Problem 2 arising from the above model and the same choices of $\boldsymbol{a}$ and $\boldsymbol{b}$. The optimal support points and corresponding optimal weights were determined by the same algorithm. These results are listed in Table 5. From this table algorithm (5) again converges to three design points, namely $\text{supp}(p^*) = \{.02, .14, .2\}$ under the choice of $\boldsymbol{a} = (0,1,0)^t$ while $\text{supp}(p^*) = \{.02, .12, .2\}$ under the other two choices of $\boldsymbol{a}$. These designs have the unusual feature of one large weight corresponding to the middle support point. Also the convergence is slow in terms of the number of iterations. This would seem to be due to a combination of small weights and zero homogeneity of the correlation criterion, the latter implying zero partial derivatives at the optimum corresponding to the positive weights. This is because the partial derivatives and the directional derivatives are equal under zero homogeneity. In general $F_j = (\partial\phi_\varrho/\partial p_j) - \sum_{i=1}^J p_i(\partial\phi_\varrho/\partial p_i)$, $j = 1, 2, \ldots, J$; but since $\phi_\varrho(p)$ is homogeneous of degree zero in the weights $\boldsymbol{p}$ then $\sum_{i=1}^J p_i(\partial\phi_\varrho/\partial p_i) = 0 \times \phi_Q(p) = 0$. (Note a function $\phi(p)$ is homogeneous of degree h if, for scalar c, $\phi(cp) = c^h\phi(p)$ and then $\sum_{i=1}^J p_i(\partial\phi/\partial p_i) = h\phi(p)$). Thus, in this example when the algorithm approaches the optimum the derivatives and some weights are small. Accordingly, proceeding from $p^{(r)}$ to $p^{(r+1)}$ will only slightly change criterion and weights values. In order to improve convergence in such cases, we might use $f(d,\delta) = (1 + s\alpha d)^{s\delta}$ instead of $f(d,\delta) = (1 + sd)^{s\delta}$, for some appropriate α. We report finally that since we wished to minimize two squared covariances or two squared correlations in this example we also considered minimizing convex combinations of these i.e.

$$\psi_c(p) = -\left\{\alpha\left(\boldsymbol{a}_1^t M^{-1}\boldsymbol{b}\right)^2 + (1-\alpha)\left(\boldsymbol{a}_2^t M^{-1}\boldsymbol{b}\right)^2\right\}, \text{ for } 0 \le \alpha \le 1, \text{ and } \boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{b} \in R^k$$

or

$$\psi_\varrho(p) = -\left\{\frac{\alpha\left(\boldsymbol{a}_1^t M^{-1}\boldsymbol{b}\right)^2}{\left(\boldsymbol{a}_1^t M^{-1}\boldsymbol{a}_1\right)\left(\boldsymbol{b}^t M^{-1}\boldsymbol{b}\right)} + \frac{(1-\alpha)\left(\boldsymbol{a}_2^t M^{-1}\boldsymbol{b}\right)^2}{\left(\boldsymbol{a}_2^t M^{-1}\boldsymbol{a}_2\right)\left(\boldsymbol{b}^t M^{-1}\boldsymbol{b}\right)}\right\}.$$

Starting from an initial uniform design on the discretization $\{x : x = .02, .04, \ldots, .2\}$ ($p_j^{(0)} = 1/10$), we determined, using the above algorithm, the optimal design under the choices of $\boldsymbol{a}_1 = (1,0,0)^t$, $\boldsymbol{a}_2 = (0,1,0)^t$ and $\boldsymbol{b} = (0,0,1)^t$ for each of the cases $\alpha = 0, .1, \ldots, .9, 1$.

For the covariance criterion the support points are always $\{.02, .12, .20\}$ and there is little change in the weights. For the correlation criterion the initial support gradually changes from $\{.02, .14, .20\}$ (the support for $\boldsymbol{a}_2$) to $\{.02, .12, .20\}$ (the support for $\boldsymbol{a}_1$) as α travels from 0 to 1. For α roughly in the range .2 to .4 there is the suggestion of a four point support namely $\{.02, .12, .14, .2\}$. However since the pair .12 and .14 are two neighbouring points in our discretization of the interval $[.02, .2]$ the pair probably represent a cluster 'replacing' a single support point which must lie in $[.12, .14]$. Under the recommendations of Atwood (1976) a good approximation to this point is the convex combination of .12 and .14 based on their optimal weights. Further it should be assigned the total of these weights. For example in the case of $\alpha = 0.3$, the support points .12 and .14 are assigned the weights .30 and .39 respectively. Instead, a total weight of .69 should be assigned to the value $x = (.30 \times .12 + .39 \times .14)/.69 \cong .13$.

3. Optimal Selection of Designs Satisfying Zero Covariance

It is possible, of course, that the optimal value of the covariance criterion is zero. This, in fact, is true in case b of Section 2.2 which also serves to illustrate that when zero is a possible value for the criterion, then typically there are many designs which attain this optimum. It then seems natural to choose one of these optimally. This leads us to consider the following problem:

Problem 3. *Maximize a standard design criterion, $\phi(p)$, subject to a zero covariance condition and $\sum_{i=1}^{J} p_i = 1$, $p_i \geq 0$, $i = 1, \ldots, J$.*

For simplicity we restrict attention to the case $J = k$ and hence to Case b of Section 2.2. A subsequent publication will be devoted to the case $J > k$.

We consider three different cases.

3.1. Case 1

A simple case is when one of the $c_i d_i$ is greater than zero, say $c_1 d_1$, one less than zero, say $c_k d_k$, and all the others are equal to zero. Equating (1) to zero, then implies

$$\frac{c_1 d_1}{p_1} = -\frac{c_k d_k}{p_k} \Rightarrow p_k = \alpha\, p_1, \tag{6}$$

where $\alpha = -c_k d_k / c_1 d_1$.

Thus there are many designs guaranteeing zero covariance. We consider an optimal choice of $p_1, p_2, \ldots, p_k$ by maximizing some of the D_A-, c- or the linear optimality criteria which we denote by ϕ_1, ϕ_2 or ϕ_3 respectively. Consider first ϕ_3. Our aim in this case is to maximize ϕ_3 subject to zero covariance i.e. subject to $p_k = \alpha\, p_1$. The criterion is of the form

$$\phi_3(p) = -\mathrm{Trace}\left[AM^{-1}(p)A^t\right] = \sum_{i=1}^{k} \eta_i^2/p_i, \tag{7}$$

where the ${\eta_i}^2$ are the diagonal elements of the matrix $V^{-1}A^tA(V^t)^{-1}$. Then, by substituting from (6) in (7) we find

$$\phi_3(p) = -\left[\sum_{i=1}^{k-1} \frac{\eta_i^2}{p_i} + \frac{\eta_k^2}{\alpha\, p_1}\right] = -\left[\sum_{i=2}^{k-1} \frac{\eta_i^2}{p_i} + \frac{\eta_1^2 + (\eta_k^2/\alpha)}{p_1}\right].$$

But since $\sum_{i=1}^{k} p_i = 1$, $p_i \geq 0$ then $\alpha\, p_1 + \sum_{i=1}^{k-1} p_i = 1 \Rightarrow (1+\alpha)p_1 + \sum_{i=2}^{k-1} p_i = 1$. If we let $q_1^+ = (1+\alpha)p_1$ and $q_j^+ = p_j$, $j = 2, \ldots, k-1$, then $\sum_{i=1}^{k-1} q_i^+ = 1$ and

$$\phi_3(q^+) = -\left[\frac{(1+\alpha)\,(\eta_1^2 + \eta_k^2/\alpha)}{q_1^+} + \sum_{i=2}^{k-1} \frac{\eta_i^2}{q_i^+}\right] = -\sum_{i=1}^{k-1} \frac{{\eta_i^+}^2}{q_i^+}, \tag{8}$$

where ${\eta_1^+}^2 = (1+\alpha)[\eta_1^2 + (\eta_k^2/\alpha)]$ and ${\eta_j^+}^2 = \eta_j^2$, $j = 2, \ldots, k-1$. Hence (8) leads to the explicit solution for the optimal weights

$$q_j^{+^*} = \left|\eta_j^+\right| / \sum_{i=1}^{k-1} \left|\eta_i^+\right|, \quad j = 1, 2, \ldots, k-1, \tag{9}$$

and the maximum value for $\phi_3(p)$ subject to zero covariance is

$$\phi_3^*(q^{+^*}) = -\left[\sum_{i=1}^{k-1} \left|\eta_i^*\right|\right]^2. \tag{10}$$

A particular case of this criterion is c-optimality (ϕ_2) when $A^t = \boldsymbol{c}^+$ for some vector $\boldsymbol{c}^+$. Then $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_k)^t = V^{-1}\boldsymbol{c}^+$.

3.2. Equivalent criteria

We have considered general A or $\boldsymbol{c}^+$ so far. It seems natural that these would be influenced by $\boldsymbol{a}$ and $\boldsymbol{b}$. Simplest choices would be $A^t = [\boldsymbol{a} \vdots \boldsymbol{b}]$ and $\boldsymbol{c}^+ = \boldsymbol{a} + \boldsymbol{b}$. In fact these lead to equivalent criteria for any choice of $\boldsymbol{a}$, $\boldsymbol{b}$ since

$$\begin{aligned} {\boldsymbol{c}^+}^t M^{-1}(p)\boldsymbol{c}^+ &\propto \mathrm{Var}\left[\left(\boldsymbol{a}^t + \boldsymbol{b}^t\right)\hat{\theta}\right] \\ &= \mathrm{Var}(\boldsymbol{a}^t\hat{\theta}) + \mathrm{Var}(\boldsymbol{b}^t\hat{\theta}) \quad \text{ if } \quad \mathrm{Cov}(\boldsymbol{a}^t\hat{\theta}, \boldsymbol{b}^t\hat{\theta}) = 0 \\ &\propto \mathrm{Trace}\left[A^tM^{-1}(p)A\right]. \end{aligned}$$

Note this result is true also when the number of support points exceeds the number of parameters. We focus attention on these choices later.

3.3. Case 2

We now consider the case when some of the $c_i d_i$ are greater than zero and all the others are less than zero. Without loss of generality suppose that $c_i d_i > 0$ for $i = 1, \ldots, n_1$; $c_{n_1+j} d_{n_1+j} < 0$ for $j = 1, \ldots, n_2$, where $n_1 + n_2 = k$. Let $e_i = c_i d_i$ and $q_i = p_i$ if $i = 1, \ldots, n_1$; $f_j = -c_{n_1+j} d_{n_1+j}$ and $w_j = p_{n_1+j}$ if $j = 1, \ldots, n_2$. We can then rewrite (1) as follows:

$$\phi_c(q, w) = -\left[\sum_{s=1}^{n_1} \frac{e_s}{q_s} - \sum_{t=1}^{n_2} \frac{f_t}{w_t}\right]^2, \tag{11}$$

where $q_i, w_j > 0$ and $\sum_{i=1}^{n_1} q_i + \sum_{j=1}^{n_2} w_j = 1$. Clearly (11) can be zero if $\sum_{s=1}^{n_1} e_s/q_s = \sum_{t=1}^{n_2} f_t/w_t$. In fact there are many designs which will guarantee this and hence $\phi_c(q, w) = 0$. We seek to choose $\boldsymbol{q}$, $\boldsymbol{w}$ optimally by maximizing a standard design criterion with a view to good estimation of $(\boldsymbol{a}^t\boldsymbol{\theta}, \boldsymbol{b}^t\boldsymbol{\theta})$. The complete class of such designs are given by the following transformation:

$$(\boldsymbol{q}, \boldsymbol{w}) \leftarrow (\boldsymbol{g}, \boldsymbol{h}) : \begin{array}{ll} q_i = \dfrac{g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s}}{z}, & i = 1, \ldots, n_1, \\ w_j = \dfrac{h_j \sum_{t=1}^{n_2} \frac{f_t}{h_t}}{z}, & j = 1, \ldots, n_2, \end{array} \tag{12}$$

where $g_i, h_j > 0$ and

$$z = \Big[\Big(\sum_{i=1}^{n_1} g_i\Big) \times \Big(\sum_{s=1}^{n_1} \frac{e_s}{g_s}\Big)\Big] + \Big[\Big(\sum_{j=1}^{n_2} h_i\Big) \times \Big(\sum_{t=1}^{n_2} \frac{f_t}{h_t}\Big)\Big].$$

3.4. Properties of the transformation

1. This transformation satisfies zero covariance.
2. The transformation satisfies the constraint $\sum_{i=1}^{n_1} q_i + \sum_{j=1}^{n_2} w_j = 1$.
3. The transformation $(\boldsymbol{q}, \boldsymbol{w}) \rightarrow (\boldsymbol{g}, \boldsymbol{h})$ is unique up to a constant multiple because the q_i and w_j and $\phi[\boldsymbol{q}(\boldsymbol{g}, \boldsymbol{h}), \boldsymbol{w}(\boldsymbol{g}, \boldsymbol{h})]$ are all homogeneous of degree zero in both $\boldsymbol{g}$ and $\boldsymbol{h}$. For this reason we invoke the constraints $\sum_{i=1}^{n_1} g_i = \sum_{j=1}^{n_2} h_j = 1$.

3.5. Case 3

A final case which attains zero covariance is when some of the $c_i d_i$ are greater than zero, some less than zero and the others are equal to zero. This case can be transformed, in a manner similar to Case 2, to a problem involving three sets of weights, two of them defined similarly to $\boldsymbol{g}$ and $\boldsymbol{h}$ above.

Without loss of generality suppose that $c_i d_i > 0$ for $i = 1, \ldots, n_1$; $c_{n+j} d_{n+j} < 0$ for $j = 1, \ldots, n_2$ and $c_{n_1+n_2+m} d_{n_1+n_2+m} = 0$ for $m = 1, \ldots, n_3$, $\sum_{i=1}^{3} n_i = k$. Then the covariance criterion in this case will be of the form

$$\phi_c(q, w, r) = -\left[\sum_{i=1}^{n_1} \frac{e_i}{q_i} - \sum_{j=1}^{n_2} \frac{f_j}{w_j} + \sum_{m=1}^{n_3} \frac{o_m}{r_m}\right]^2, \tag{13}$$

where $e_i = c_i d_i$ and $q_i = p_i$ for $i = 1, \ldots, n_1$; $f_j = -c_{n_1+j} d_{n_1+j}$ and $w_j = p_{n_1+j}$ for $j = 1, \ldots, n_2$ and $o_m = 0$ and $r_m = p_{n_1+n_2+m}$ for $m = 1, \ldots, n_3$, $\sum_{i=1}^{3} n_i = k$. As in Case 2 the complete class of designs which guarantee zero covariance is given by the transformation:

$$(\boldsymbol{q}, \boldsymbol{w}, \boldsymbol{r}) \leftarrow (\boldsymbol{g}, \boldsymbol{h}, \boldsymbol{u}) : \begin{array}{l} q_i = u_{n_3+1} \dfrac{g_i \sum_{s=1}^{n_1} \frac{e_s}{g_s}}{z}, \quad i = 1, \ldots, n_1, \\ w_j = u_{n_3+1} \dfrac{h_j \sum_{t=1}^{n_2} \frac{f_t}{h_t}}{z}, \quad j = 1, \ldots, n_2, \\ r_m = u_m, \quad m = 1, \ldots, n_3, \end{array} \tag{14}$$

where $u_{n_3+1} = \sum_{i=1}^{n_1} q_i + \sum_{j=1}^{n_2} w_j$, and z is as above.

This transformation has similar properties to the transformation mentioned in Case 2. In particular the criterion $\phi[\boldsymbol{q}(\boldsymbol{g}, \boldsymbol{h}, u_{n_3+1}), \boldsymbol{w}(\boldsymbol{g}, \boldsymbol{h}, u_{n_3+1}), \boldsymbol{r}(\boldsymbol{u})]$ is again homogeneous of degree zero in $\boldsymbol{g}$ and $\boldsymbol{h}$. In addition it will also be homogenous in $\boldsymbol{u}$ if $\phi(\cdot)$ is homogenous in the original weights $\boldsymbol{p}$ as is usually the case with design criteria. We therefore invoke the constraints $\sum_{i=1}^{n_1} g_i = \sum_{j=1}^{n_2} h_i = \sum_{m=1}^{n_3+1} u_m = 1$. Under this array of transformations Problem 3 changes to: Either maximize $\psi(\boldsymbol{g}, \boldsymbol{h}) = \phi[\boldsymbol{q}(\boldsymbol{g}, \boldsymbol{h}), \boldsymbol{w}(\boldsymbol{g}, \boldsymbol{h})]$ subject to

$$\sum_{i=1}^{n_1} g_i = \sum_{j=1}^{n_2} h_j = 1 \quad \text{and} \quad g_i,\ h_j > 0$$

or maximize $\psi(\boldsymbol{g}, \boldsymbol{h}, \boldsymbol{u}) = \phi[\boldsymbol{q}(\boldsymbol{g}, \boldsymbol{h}, u_{n_3+1}), \boldsymbol{w}(\boldsymbol{g}, \boldsymbol{h}, u_{n_3+1}), \boldsymbol{r}(\boldsymbol{u})]$ subject to

$$\sum_{i=1}^{n_1} g_i = \sum_{j=1}^{n_2} h_i = \sum_{m=1}^{n_3+1} u_m = 1 \quad \text{and} \quad g_i, h_j, u_m > 0.$$

Here $\phi(\cdot)$ is one of the criteria ϕ_1, ϕ_2 or ϕ_3. We denote the $\psi(\cdot)$ derived from these under the above process by ψ_1, ψ_2 or ψ_3 respectively.

3.6. Proposed algorithm

In Section 2.3 we described and subsequently used algorithm (5) for maximizing a criterion with respect to one set of weights $p_1, \ldots, p_J$. In Problem 3 we have two sets of weights g_i, $i = 1, 2, \ldots, n_1$, and h_j, $j = 1, 2, \ldots, n_2$, and possibly a third set u_m, $m = 1, 2, \ldots, n_3$. A natural extension of this algorithm is:

$$\left.\begin{array}{l} \left.\begin{array}{ll} g_i^{(r+1)} = \dfrac{g_i^{(r)} f_1(d_{1i}^{(r)}, \delta)}{\sum_{s=1}^{n_1} g_s^{(r)} f_1(d_{1s}^{(r)}, \delta)}, & i = 1, \ldots, n_1, \\ h_j^{(r+1)} = \dfrac{h_j^{(r)} f_2(d_{2j}^{(r)}, \delta)}{\sum_{t=1}^{n_2} h_t^{(r)} f_2(d_{2t}^{(r)}, \delta)}, & j = 1, \ldots, n_2, \end{array}\right\} \text{(16a)} \\ \text{and if necessary} \\ u_m^{(r+1)} = \dfrac{u_m^{(r)} f_3(d_{3m}^{(r)}, \delta)}{\sum_{v=1}^{n_3+1} u_v^{(r)} f_3(d_{3v}^{(r)}, \delta)}, \quad m = 1, \ldots, n_3 + 1, \end{array}\right\} \quad \text{(16b)}$$

where $f_i(d, \delta)$, $i = 1, 2, 3$, are choices of $f(d, \delta)$ satisfying the conditions listed in Section 3.2, and $d_{1i}^{(r)}$, $d_{2j}^{(r)}$ and $d_{3m}^{(r)}$ are the first partial derivatives with respect to the variables g_i, h_j and u_m respectively, at current iterate values, of the criterion (ψ_1, ψ_2, ψ_3) to be maximized.

Our choices of $f_i(d, \delta)$, must respect the fact that these criteria like the covariance criterion, can have negative derivatives. We again consider the case $f_i(d, \delta) = (1 + sd)^{s\delta}$, $s = \text{sign}(d)$.

3.7. Examples and discussion

We now explore two examples.

Example 1. An example of Case 1 and Case 3 can be derived from the first example considered by Silvey et al. (1978) and also by Wu (1978). The design space for this example is: $V = \{(1, -1, -1)^t, (1, -1, 1)^t (1, 1, -1)^t, (1, 2, 2)^t\}$.

Consider $\boldsymbol{a} = (1, 0, 0)^t$ and $\boldsymbol{b} = (0, 0, 1)^t$ and for the linear criterion (ϕ_3) $A^t = [\boldsymbol{a} \vdots \boldsymbol{b}]$. Then a subset of the design space which satisfies Case 1 and hence Case 3 is

$$V = \{(1, -1, 1)^t, (1, 1, -1)^t, (1, 2, 2)^t\}.$$

Values which emerge are: $\boldsymbol{cd} = (0.0625, -0.1875, 0.0)^t$, $\alpha = 0.1875/0.0625 = 3$ and $\boldsymbol{\eta}^2 = (0.2656, 0.3906, 0.0625)^t$. Substituting these results into (9) and (10) we obtain a maximum value for the linear optimality criterion of -2.2750 while the optimal weights are $\boldsymbol{q}^{+^*} = (.8343, .1657)^t$, $\boldsymbol{p}^* = (.208, .626, .166)^t$. Of course we had explicit formulae for calculating $\boldsymbol{q}^{+^*}$ and $\boldsymbol{p}^*$. But this example is an illustration of Case 3 although with $n_1 = n_2 = n_3 = 1$ and hence $g_1 = h_1 = 1$. Starting from $u_1^{(0)} = u_2^{(0)} = 1/2$ algorithm (16b) converged to the optimal solution attaining $\max F_j \leq 10^{-5}$ in six iterations.

Example 2. Finally, we consider an example for both Case 2 and Case 3. Consider the cubic regression model $E(y) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$, $x \in [-\beta, \beta]$,

$\beta = 1, 2, \ldots, 5$ and choose $\boldsymbol{a} = (1, 0, 0, 0)^t$, $\boldsymbol{b} = (0, 0, 1, 0)^t$ and $A^t = [\boldsymbol{a} \vdots \boldsymbol{b}]$.

Table 6. This table shows the optimal support points and corresponding optimal weights for the criteria ψ_2 and ψ_3 in the case of the cubic regression model (Example 2) when the design space is the interval $[-\beta, \beta]$ where β takes the values 1, 2, 3, 4 and 5.

	$\beta = 1$	$\beta = 2$	$\beta = 3$	$\beta = 4$	$\beta = 5$
The maximum value for ψ_2 and ψ_3	-9.45062	-2.29000	-1.53441	-1.29294	-1.17614
Weights g^* and h^*	0.16440	0.21991	0.24656	0.25422	0.25034
	0.14852	0.07006	0.04216	0.02631	0.01529
	0.85148	0.92994	0.95784	0.97369	0.98471
	0.83560	0.78009	0.75344	0.74578	0.74966
Weights g^* and w^*	0.03166	0.02300	0.01422	0.00900	0.00600
	0.11992	0.06273	0.03973	0.02538	0.01492
	0.68751	0.83269	0.90261	0.93923	0.96110
	0.16091	0.08158	0.04344	0.02639	0.01798
Corresponding design points	-1	-2	-3	-4	-5
	$-.97$	-1.93	-2.74	-3.65	-4.64
	-0.01	-0.01	-0.02	-0.02	-0.01
	1	2	3	4	5

Table 7. This table shows the number of iterations needed to achieve $\max F_j \leq 10^{-n}$; $n = 1, 2, 3, 4, 5$ and $j = 1, 2, 3, 4$ in the intervals $[-\beta, -\beta]$ for the criterion ψ_2 where β_3 where β takes the values $1, 2, 3, 4$ and 5 when $\boldsymbol{a} = (1, 0, 0, 0)^t$, $\boldsymbol{b} = (0, 0, 1, 0)^t$, in the case of the cubic regression model (Example 2) when the design space is the interval $[-\beta, \beta]$, $\beta = 1, 2, \ldots, 5$.

δ	$\beta \backslash n$	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
$\delta = 0.072$	$\beta = 1$	3	4	4	6	6
$\delta = 0.33$	$\beta = 2$	5	6	8	11	13
$\delta = 0.6$	$\beta = 3$	4	8	12	18	24
$\delta = 0.7$	$\beta = 4$	3	10	20	29	39
$\delta = 0.8$	$\beta = 5$	3	13	25	37	49

In Table 6 we report the optimal four point designs for the criteria ψ_2 and ψ_3 on $[-\beta, \beta]$ subject to the endpoints being support points for $\beta = 1, 2, \ldots, 5$. In

Table 7 we report the number of iterations needed to achieve $\max F_i \leq 10^{-n}$ for $n = 1, 2, 3, 4, 5$ for those values of δ which attain $\max F_i \leq 10^{-5}$ in the smallest number of iterations when finding the optimal weights of these particular support points.

The two middle points were found by a search through the supports $\{-\beta, x_1, x_2, \beta\}$ where $x_1 < x_2$ and $x_1, x_2 < 0$ or $x_1, x_2 > 0$, algorithms (16a) or (16b) being used for each pair x_1 and x_2. (Note that we exclude the case when x_1 and x_2 have opposite signs because then the $c_i d_i$ are either all negative or all positive, so that no design satisfies zero covariance). In particular algorithm(16b) needed to be used when $x_1 + x_2 = \pm\beta$. To see this consider the four point support $\{-1, \alpha, (1-\alpha), 1\}$, $0 < \alpha < 1$. Then if $\boldsymbol{a} = (1, 0, 0, 0)^t$, $\boldsymbol{b} = (0, 0, 1, 0)^t$ we find

$$\begin{aligned} \boldsymbol{d} = V^{-1}\boldsymbol{a} = S[&\alpha^2(\alpha-1)^2(2\alpha-1),\, 2\alpha(1-\alpha)(\alpha-2),\\ &-2\alpha(\alpha^2-1),\, -\alpha(1-\alpha)(2\alpha^3-3\alpha^2-3\alpha+2)]^t \end{aligned}$$

and $\boldsymbol{c} = V^{-1}\boldsymbol{b} = S[2\alpha(1-\alpha)(2\alpha-1),\ 2\alpha(2-\alpha)(1-\alpha),\ \ 2\alpha(1-\alpha^2), 0]^t$, $S = (1/\mathrm{Det}(V))$ and hence $(c_1d_1, c_2d_2, c_3d_3, c_4d_4)^t = S^2[-2\alpha^3(\alpha-1)^3(2\alpha-1)^2$, $-4\alpha^2(1-\alpha)^2(\alpha-2)^2$, $4\alpha^2(1-\alpha^2), 0]^t$. So in this case we have one positive, two negative and one zero $c_i d_i$ thus satisfying Case 3.

Again some optimal weights are small. This is because the zero covariance constraint limits some weights to ranges of small values, a phenomenon which has been noted in other examples.

4. Conclusion

We have introduced two new design criteria, the covariance criterion and the correlation criterion, and have explored the use of a specialised algorithm indexed by a free parameter δ for constructing designs which optimize these criteria. We have also considered the case of optimizing standard design criteria subject to zero covariance when the number of support points equals the number of parameters, transforming this problem to one of optimizing a criterion with respect to two or three sets of weights and using a natural extension of the above algorithm to solve this derived problem.

The examples considered are arguably small scale and the algorithm did not always enjoy a reasonable rate of convergence even for good choices of the free parameter δ. There is need for further work on improving the algorithm and on considering larger scale examples. We have also still to report on the case when the number of support points exceeds the number of parameters when optimising a standard design criteria subject to zero covariance.

Finally further investigations will include extensions of the above problems to include several covariances or correlations.

Acknowledgement

We thank a referee for a thorough assessment of the paper. This work was done while the second author was completing a Ph.D. at the University of Glasgow.

References

Atwood, C. L. (1976). Convergent design sequences, for sufficiently regular optimality criteria. Ann. Statist. **4**, 1124-1138.

Fellman, J. (1989). An empirical study of a class of iterative searches for optimal designs. J. Statist. Plann. Inference **21**, 85-92.

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. Ann. Statist. **19**, 1614-1625.

Silvey, S. D., Titterington, D. M. and Torsney, B. (1978). An algorithm for optimal designs on a finite design space. Comm. Statist. **14**, 1379-1389.

Titterington, D. M. (1976). Algorithms for computing D-optimal designs on a finite design space. Proceedings, Conference on Information Sciences and Systems, Dept. Elect. Eng., 213-216, Jones Hopkins Univ., Baltimore.

Torsney, B. (1977). Contribution to discussion of "Maximum likelihood from incomplete data via the EM algorithm" by Dempster et al. J. Roy. Statist. Ser.B **39**, 26-27.

Torsney, B. (1981). Algorithms for a constrained optimaization problem with applications in statistics and optimum design. Ph.D. Thesis, University of Glasgow.

Torsney, B. (1983). A moment inequality and monotonicity of an algorithm. Proc. Internat. Symp. on Semi-Infinite Programming and Appl. (Edited by K. O. Kortanek and A. V. Fiacco) at Univ. of Texas, Austin. Lecture Notes in Econom. and Math. Sys. **215**, 249-260.

Torsney, B. (1988). Computing optimizing distributions with applications in design, estimation and image processing. Optimal Design and Analysis of Experiments (Edited by Y. Dodge, V. V. Fedorov and H. P. Wynn), 361-370, Elsevier Science Publishers B. V., North Holland.

Torsney, B. and Alahmadi, A. M. (1992). Further development of algorithms for constructing optimizing distributions. Model Oriented Data Analysis. Proc. 2nd IIASA Workshop in St. Kyrik, Bulgaria (Edited by V. V. Fedorov, W. G. Müller and I. N. Vuchkov), 121-129, Physica-Verlag.

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

Wu, C. F. (1978). Some iterative procedures for generating nonsingular optimal designs. Comm. in Statist. **14**, 1399-1412.

Department of Statistics, University of Glasgow, 15 University Gdns., Glasgow G12 8QW, U.K.

Department of Mathematics, King Abdulaziz University, Faculty of Education, Madinah Munawwarah, P. O. Box 344, Saudi Arabia.

(Received October 1993; accepted March 1995)