Optimal weights for experimental designs on linearly independent support points

Friedrich Pukelsheim, Ben Torsney

Angaben zur Veröffentlichung / Publication details:

Pukelsheim, Friedrich, and Ben Torsney. 1991. "Optimal weights for experimental designs on linearly independent support points." *The Annals of Statistics* 19 (3): 1614–25. https://doi.org/10.1214/aos/1176348265.

The Annals of Statistics
1991, Vol. 19, No. 3, 1614–1625

OPTIMAL WEIGHTS FOR EXPERIMENTAL DESIGNS ON LINEARLY INDEPENDENT SUPPORT POINTS

By Friedrich Pukelsheim and Ben Torsney

Universität Augsburg and University of Glasgow

An explicit formula is derived to compute the A-optimal design weights on linearly independent regression vectors, for the mean parameters in a linear model with homoscedastic variances. The formula emerges as a special case of a general result which holds for a wide class of optimality criteria. There are close links to iterative algorithms for computing optimal weights.

1. Introduction. In experimental design the main tool for the checking of optimality of a candidate design are equivalence theorems [see Kiefer (1974) and Pukelsheim and Titterington (1983)]. Or else we may use algorithms that find optimal designs by machine computation. However, Torsney (1981) and Kitsos, Titterington and Torsney (1988), Section 6.1, recently established an explicit formula for c-optimal designs provided the regression vectors that support the design are linearly independent. The notion of c-optimality indicates that interest concentrates on the scalar subsystem $c'\theta$ of the full mean parameter vector $\theta \in \mathbb{R}^k$. A special case is reported in Karlin and Studden (1966), page 362, and a closely related formula appears in Studden (1977), page 413.

In the present article we investigate s-dimensional subsystems $K'\theta$ and arbitrary optimality criteria ϕ. In general, the ϕ-optimal weights for $K'\theta$ satisfy a nonlinear fix point equation. For A-optimality an explicit formula emerges, generalizing the c-optimality formula mentioned above. For D-optimality we obtain a new proof of Atwood's (1973) result that the optimal weights are bounded from above by $1/s$. Otherwise our result points to an algorithm that is a particular member of a family introduced by Silvey, Titterington and Torsney (1978).

Much of the previous work on the subject uses *differential* calculus. In the present article we derive our results from an equivalence theorem which is based on the *subdifferential* calculus of convex analysis [see Pukelsheim (1980), Theorem 5]. As a consequence we obtain a complete coverage of the "boundary problems" that arise with the optimal design problem. In particular, three items deserve special mentioning.

(i) Our optimality criteria ϕ are information functions; that is, they are nonnegative, homogeneous and concave. They need not be differentiable.

Received March 1990; revised September 1990.
AMS 1980 *subject classification*. 62K05.
Key words and phrases. General equivalence theorem, information functions, matrix means, algorithms for optimal designs, A-optimality, D-optimality, c-optimality.

(ii) Our optimization problem runs over all designs on the given regression vectors. We do not need to restrict attention to those designs under which the parameter system $K'\theta$ is identifiable (i.e., estimable, testable).

(iii) The optimal support may be a strict subset of the given regression vectors; that is, some of the weights of the optimal design may vanish. We do not require the weights to be positive at the point of optimality. This is of relevance if the number of given regression vectors exceeds the dimensionality of the parameter system of interest.

The set of competing designs considered in the present article is the set Ξ of *all* designs on the *finite* regression range $\mathscr{X} = \{x_1, \dots, x_l\} \subseteq \mathbb{R}^k$. Therefore we omit reference to the set of competing designs and simply speak of ϕ-optimal designs for $K'\theta$, where $K'\theta$ designates the parameter system of interest. Our crucial assumption is that the regression vectors $x_1, \dots, x_l$ are linearly independent.

2. The model. We consider the classical linear model where for each *regression vector* $x_i \in \mathbb{R}^k$ we make real observations Y_{ij}, for $j = 1, \dots, n_i$, whose expectation depends linearly on x_i and on an unknown mean parameter $\theta \in \mathbb{R}^k$,

$$E[Y_{ij}] = x_i'\theta. \tag{1}$$

The observations Y_{ij} are taken to be uncorrelated and of equal variance. In the present article we assume that the regression vectors $x_1, \dots, x_l$, say, are *linearly independent*.

If the full mean parameter system $\theta \in \mathbb{R}^k$ is of interest, there have to be at least $l = k$ regression vectors for θ to be estimable. More generally, we consider parameter subsystems $K'\theta$ with a *coefficient matrix* $K \in \mathbb{R}^{k \times s}$ that is of full column rank s. Then the number of regression vectors lies between s and k,

$$s \le l \le k. \tag{2}$$

The lower bound s is needed so that $K'\theta$ is identifiable, and the upper bound k is enforced by the assumption of linear independence. Optimal designs may be supported by as much as $s(s+1)/2 + s(k-s)$ regression vectors [see Fellman (1974), Theorem 4.1.4, or Pukelsheim (1980), Theorem 6]; for $s > 1$ this number exceeds k. Hence our restriction (2) means that we are investigating designs with a *small support size*, only. For practical purposes such designs are preferable anyway.

The experimental design underlying model (1) is determined by the replication numbers $n_1, \dots, n_l$. More generally, we take an experimental design to be given by the *weights* w_i,

$$\min_{i=1}^{l} w_i \ge 0, \qquad \sum_{i=1}^{l} w_i = 1.$$

The weight w_i is the proportion of observations to be made when the regression vector is x_i. For model (1) we have $w_i = n_i/\sum_{j=1}^{l} n_j$.

There exist many optimality criteria which serve to discriminate between competing designs [cf. Kiefer (1974)]. Almost all of them depend on the *moment matrix*

$$M(w) = \sum_{i=1}^{l} w_i x_i x_i' = X'\Delta_w X.$$

Here the matrix $X \in \mathbb{R}^{l\times k}$ assembles the regression vectors x_i according to $X' = (x_1, \dots, x_l)$. Since the l regression vectors are assumed to be linearly independent, the matrix X has full row rank l. The matrix Δ_w is defined to be the diagonal $l \times l$ matrix with the weight vector $w = (w_1, \dots, w_l)'$ on its diagonal.

Given an optimality criterion ϕ, a closed-form expression for the ϕ-optimal weights would be extremely useful. Our results imply such an explicit formula, for the particular case of the average-variance criterion. When applied to the full parameter system θ, the A-criterion demands minimization of

$$\frac{1}{k}\operatorname{trace} M(w)^{-1}, \tag{3}$$

the average of the variances of the least squares estimates $\hat{\theta}_1, \dots, \hat{\theta}_k$, standardized relative to sample size n and model variance σ^2. It follows from our Corollary 4 that, when $l = k$, the A-optimal weights for θ are

$$w_i = \frac{\sqrt{b_{ii}}}{\sum_{j=1}^{k}\sqrt{b_{jj}}}, \qquad i = 1, \dots, k, \tag{4}$$

and the minimum value of (3) is

$$\frac{1}{k}\left(\sum_{j=1}^{k}\sqrt{b_{jj}}\right)^2, \tag{5}$$

where $b_{11}, \dots, b_{kk}$ are the diagonal elements of the positive-definite $k \times k$ matrix

$$B = (XX')^{-1}.$$

A direct proof of this special case is easy. For the full parameter θ all weights w_i must be positive, and differentiation of the objective function

$$\frac{1}{k}\operatorname{trace} M(w)^{-1} = \frac{1}{k}\operatorname{trace} \Delta_w^{-1}(XX')^{-1} = \frac{1}{k}\sum_{i=1}^{l}\frac{b_{ii}}{w_i}$$

leads to the optimal solution (4) and the minimum value (5).

For a parameter subsystem $K'\theta$ the objective function (3) turns into

$$\frac{1}{k}\operatorname{trace} K'M(\xi)^{-}K,$$

and an optimal design may have some weights that vanish. Instead of augmenting the calculus approach by a discussion of the boundary behaviour, we derive the general result from the equivalence theorem.

3. ϕ-optimal weights. Let $\mathscr{M}$ the set of *all* moment matrices on the regression range $\mathscr{X} = \{x_1, \dots, x_l\}$. Given a moment matrix $M \in \mathscr{M}$, the *information matrix* for $K'\theta$ is

$$C_K(M) = \min\{LML' : L \in \mathbb{R}^{s \times k},\ LK = I_s\}, \tag{6}$$

where the minimum is taken relative to the Loewner ordering [see Gaffke and Pukelsheim (1988)]. In particular, if the first s out of the k components of θ are of interest, then $K' = (I_s, 0)$, and (6) becomes

$$C_K(M) = M_{11} - M_{12} M_{22}^{-} M_{21}. \tag{7}$$

In general, the subsystem $K'\theta$ is identifiable under M if and only if the range of M includes the range of K, and then (6) is the positive-definite $s \times s$ matrix

$$C_K(M) = (K'M^{-}K)^{-1}. \tag{8}$$

Expressions (7) and (8) are invariant to the choice of the generalized inverses M_{22}^{-} and M^{-} of M_{22} and M, respectively.

If all weights are positive, then a convenient generalized inverse of $M(w) = X'\Delta_w X$ is the Moore–Penrose inverse

$$M(w)^{+} = X'(XX')^{-1}\Delta_w^{-1}(XX')^{-1}X.$$

This is the starting point for Torsney (1981) and Kitsos, Titterington and Torsney (1988). Let A^{-} denote the *set* of all generalized inverses of a matrix A. With no restrictions on the weights, we obtain the inclusion relation

$$X'(XX')^{-1}\Delta_w^{-}(XX')^{-1}X \subseteq M(w)^{-}. \tag{9}$$

Hence upon introducing the $l \times s$ matrix

$$V = (XX')^{-1}XK, \tag{10}$$

the information matrix in (8) takes the form

$$C_K(M(w)) = (V'\Delta_w^{-}V)^{-1}. \tag{11}$$

Now let ϕ be an information function on the set of nonnegative-definite $s \times s$ matrices. The *design problem* then is:

$$\text{Maximize } \phi(C_K(M(w))) \quad \text{for } w \in R^l \quad \text{subject to } \min_{i=1}^{l} w_i \geq 0, \qquad \sum_{i=1}^{l} w_i = 1.$$

The following theorem describes the optimal solution in terms of a fix point relation for the weights w_i. It makes use of the *polar* ϕ^0 of the information function ϕ [see Pukelsheim and Titterington (1983), page 1062].

THEOREM. *Let the weight vector w be such that $K'\theta$ is identifiable under $M(w)$, that is, the range of $M(w)$ contains the range of K, and let C be the information matrix $C_K(M(w))$ from* (11), *with V given by* (10).

Then w is ϕ-optimal for $K'\theta$ if and only if there exists a nonnegative-definite $s \times s$ matrix D with the properties that

$$(12) \qquad \phi(C)\phi^0(D) = \operatorname{trace} CD = 1$$

and that the weights satisfy

$$(13) \qquad w_i = \sqrt{a_{ii}} \quad \textit{for all } i = 1, \dots, l,$$

where $a_{11}, \dots, a_{ll}$ are the diagonal elements of the nonnegative-definite $l \times l$ matrix

$$(14) \qquad A = VCDCV'.$$

PROOF. For the direct part assume w to be ϕ-optimal for $K'\theta$. By Theorem 5 of Pukelsheim (1980), there exist a nonnegative-definite $s \times s$ matrix D fulfilling (12) and a generalized inverse $G \in M^-$ such that, for all $i = 1, \dots, l$,

$$(15) \qquad x_i'G'KCDCK'Gx_i \le 1.$$

In the case that x_i is a support point of the design ξ, that is, $w_i > 0$, (15) holds with equality and the product $K'Gx_i$ is invariant to the choice of $G \in M^-$. A convenient choice is

$$(16) \qquad G = X'(XX')^{-1}\Delta_w^+(XX')^{-1}X$$

from (9). Furthermore, observe that $x_i = X'e_i$, where e_i is the Euclidean unit vector of $\mathbb{R}^k$ with ith entry unity and 0's elsewhere. Thus (15) yields

$$1 = e_i'\Delta_w^+VCDCV'\Delta_w^+e_i = \frac{a_{ii}}{w_i^2},$$

and leads to relation (13).

In the case that x_i is *not* a support point of ξ, that is, $w_i = 0$, (15) becomes useless and (13) merely holds because of the identifiability assumption. Namely, without loss of generality let $w_1, \dots, w_r > 0 = w_{r+1} = \cdots = w_l$. Then the range of $M(w)$ is spanned by $x_1, \dots, x_r$. Since it includes the range of K, there exists some $r \times s$ matrix H such that

$$\begin{aligned} K &= (x_1, \dots, x_r)H \\ &= (x_1, \dots, x_r, x_{r+1}, \dots, x_l)\begin{pmatrix} H \\ 0 \end{pmatrix} \\ &= X'\begin{pmatrix} H \\ 0 \end{pmatrix}. \end{aligned}$$

For $i > r$ we now get $V'e_i = K'X'(XX')^{-1}e_i = (H', 0)e_i = 0$ and $a_{ii} = 0$. Hence $w_i = 0 = \sqrt{a_{ii}}$ again verifies (13).

For the converse part choose G as in (16). Then the arguments from the direct part also show that the quadratic form in (15) equals 1 or 0 according to whether x_i is a support point or not. Hence inequality (15) is satisfied, and optimality follows. □

For c-optimality, the matrix V of (15) becomes the vector

$$v = (XX')^{-1}Xc. \tag{17}$$

COROLLARY 1. *The weight vector w is optimal for $c'\theta$ if and only if*

$$w_i = \frac{|v_i|}{\sum_{j=1}^{l}|v_j|}, \qquad i = 1, \ldots, l,$$

with vector v given by (17), *and in this case the optimal value is*

$$c'M(\xi)^{-}c = \left(\sum_{j=1}^{l} |v_j|\right)^2.$$

PROOF. From $C = 1/D = (c'M(w)^{-}c)^{-1}$ we get $A = (c'M(w)^{-}c)^{-1}vv'$. Therefore $a_{ii} = (c'M(w)^{-}c)^{-1}v_i^2$ and $1 = \Sigma w_i = (c'M(w)^{-}c)^{-1/2}\Sigma|v_i|$. □

A proof of Corollary 1, assuming all weights to be positive, is reported in Kitsos, Titterington and Torsney (1988), Section 6.1, but the result first appeared in Torsney (1981). We note that it can be taken further for Fellman (1974), Theorem 3.1.4 and subsequent remark, established that there must always exist a c-optimal design satisfying our condition that $x_1, \ldots, x_l$ be linearly independent. Combining the two results, we have the following corollary.

COROLLARY 2. *There always exists a c-optimal design on linearly independent support points whose optimal weights, given the support points, can be determined explicitly according to Corollary* 1.

4. ϕ_p-optimal weights. An important family of criteria are the matrix means $\phi_p(C)$ with $p \in [-\infty, 1]$, that is, the p-mean of the eigenvalues of the positive-definite matrix C. The familiar D-, A- and E-criteria are included as ϕ_0, ϕ_{-1} and $\phi_{-\infty}$.

COROLLARY 3. *Assume $p \in (-\infty, 1]$. Then the weight vector w is ϕ_p-optimal for $K'\theta$ if and only if*

$$w_i = \frac{\sqrt{b_{ii}}}{\sum_{j=1}^{l}\sqrt{b_{jj}}} \quad \text{for } i = 1, \ldots, l, \tag{18}$$

where $b_{11}, \ldots, b_{ll}$ are the diagonal elements of the nonnegative-definite $l \times l$ matrix

$$B = VC^{p+1}V', \tag{19}$$

with matrix V given by (10), *and in this case the optimal value is*

$$\phi_p(C_K(M(w))) = \left(\frac{1}{s}\left(\sum_{j=1}^{l}\sqrt{b_{jj}}\right)^2\right)^{1/p}.$$

PROOF. According to Lemma 3 in Pukelsheim (1980), page 355, the unique solution to (12) is $D = C^{p-1}/\text{trace}\, C^p$. Hence A and B of (14) and (19) are proportional, $A = B/\text{trace}\, C^p$, and the result follows. □

For $p = -1$, that is, for the A-criterion, the matrix B in (19) simplifies to VV', and we obtain the explicit formula on A-optimality mentioned in Section 2.

COROLLARY 4. *The weight vector* w *is* ϕ_{-1}*-optimal for* $K'\theta$ *if and only if formula* (18) *holds with*

$$B = VV';$$

and in this case the optimal value is

$$\phi_{-1}(C_K(M(\xi))) = \frac{s}{\left(\sum_{j=1}^{l}\sqrt{b_{jj}}\right)^2}. \tag{20}$$

For $p = 0$, that is, for the D-criterion, the matrix B in (19) is

$$B = VCV' = V(V'\Delta_w^{-}V)^{-1}V', \tag{21}$$

whence A in (14) becomes $B/\text{trace}\, C^0 = B/s$. It is known that, in the Loewner ordering, B in (21) can be estimated from above by Δ_w. From $w_i = \sqrt{b_{ii}/s} \le \sqrt{w_i/s}$ we thus obtain

$$w_i \le \frac{1}{s}, \qquad i = 1, \ldots, l.$$

This bound is due to Atwood (1973) who proved it for an arbitrary number of support points. When $s = k$ our derivation yields $B = \Delta_w$ and $w_i = 1/k$, which is in line with our crucial restriction (2).

For the general parameter p formula (18) suggests the iterative procedure to choose, in step $r + 1$, the weight $w_i^{(r+1)}$ proportional to $\sqrt{b_{ii}(w^{(r)})}$. This procedure appears as case $\lambda = 1/2$ in the family of algorithms introduced by Silvey, Titterington and Torsney (1978). Namely, if ϕ is differentiable and the weights are positive, then the function $\Phi(w) = \phi((V'\Delta_w^{-1}V)^{-1})$ is differentiable, with partial derivatives $d_i = \partial\Phi/\partial w_i = \phi(C)a_{ii}/w_i^2$. Hence, with $\lambda = 1/2$, our procedure requires $w_i^{(r+1)}$ to be proportional to

$$\sqrt{a_{ii}(w^{(r)})} = w_i^{(r)}d_i^{\lambda}. \tag{22}$$

Torsney (1981, 1983) proved monotonicity of this iteration, with $\lambda = 1/2$, in the case of $p = -1$. This choice of λ relates to an earlier result of Fellman

(1974), Theorem 3.1.5. Torsney also explored monotonicity of (22) for a general criterion ϕ, establishing a sufficient condition for monotonicity of the choice $\lambda = 1/(t+1)$ when the criterion is homogeneous of degree $-t$. More recently, Fellman (1989) reports a numerical study on the range of values of λ for which (22) is monotonic in the case $p = -1$, while Torsney (1988) generalizes (22) to the iteration that $w_i^{(r+1)}$ is proportional to $w_i^{(r)} f(d_i, \lambda)$, where $f(d,\lambda)$ is a positive increasing function of d.

5. Examples. First, we consider the line fit model $x(t) = (1, t)'$, with $t \in [-1, 1]$. The regression range is

$$\mathscr{X} = \left\{ \begin{pmatrix} 1 \\ t \end{pmatrix} : t \in [-1, 1] \right\} \subseteq \mathbb{R}^2,$$

and every optimal design is supported by the extreme regression vectors

$$x_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad x_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

of $\mathscr{X}$, as follows from Theorem 2.1 of Ehrenfeld (1956). With $X = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ the vector v in (17) becomes

$$v = (X^{-1})'c = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} c_1 - c_2 \\ c_1 + c_2 \end{pmatrix}.$$

Hence the optimal weights w_1 and w_2 for $c'\theta$ are proportional to $|c_1 - c_2|$ and $|c_1 + c_2|$. If $c_1 = \pm c_2$, then one weight is 0.

Second, we consider the parabola fit model $x = (1, t, t^2)$ with $t \in [-1, 1]$. We choose $K' = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. By symmetry considerations we can restrict attention to the designs on the regression vectors $x(t)$ with $t = \pm 1, 0$ [cf. Pukelsheim (1987), page 213]. Hence we have

$$X = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \qquad X^{-1} = \frac{1}{2}\begin{pmatrix} 0 & 2 & 0 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{pmatrix}.$$

The diagonal elements of $(XX')^{-1} = (X^{-1})'X^{-1}$ are $1/2, 2, 1/2$, whence the A-optimal design for θ has weights $w_{\pm 1} = 1/4$ and $w_0 = 1/2$. The diagonal elements of $VV' = (X^{-1})'KK'X^{-1}$ are $1/2, 1, 1/2$, whence the A-optimal design for the linear and quadratic coefficient has weights $w_{\pm 1} = (\sqrt{2} - 1)/2$ and $w_0 = 2 - \sqrt{2}$, as given by Pukelsheim (1980), Example 6.2.4.

Third, we consider a polynomial fit model of degree d, $x(t) = (1, t, \dots, t^d)'$, with $t \in [-1, 1]$. A reasonable support is formed by the quantiles of the arcsin distribution, $x_i = x(s_i)$ with

$$s_i = \sin\left(\left(\frac{i}{d} - \frac{1}{2}\right)\pi\right), \qquad i = 0, \dots, d, \tag{23}$$

as pointed out by Fedorov (1972) for the D-criterion. His arguments actually apply to the full family of matrix means [see Pukelsheim (1992)].

TABLE 1

A-optimal arcsin support designs

3	−1	−0.5	0.5											0.10541
	0.158	0.342	0.342	0.158										
4	−1	−0.707	0	0.707	1									0.02613
	0.114	0.250	0.271	0.250	0.114									
5	−1	−0.809	−0.309	0.309	0.809	1								$0.0^2 6019$
	0.090	0.194	0.217	0.217	0.194	0.090								
6	−1	−0.866	−0.5	0	0.5	0.866	1							$0.0^2 1320$
	0.074	0.157	0.176	0.186	0.176	0.157	0.074							
7	−1	−0.901	−0.623	−0.223	0.223	0.623	0.901	1						$0.0^3 2798$
	0.063	0.132	0.146	0.159	0.159	0.146	0.132	0.063						
8	−1	−0.924	−0.707	−0.383	0	0.383	0.707	0.924	1					$0.0^4 5774$
	0.055	0.114	0.124	0.136	0.142	0.136	0.124	0.114	0.055					
9	−1	−0.940	−0.766	−0.5	−0.174	0.174	0.5	0.766	0.940	1				$0.0^4 1168$
	0.048	0.100	0.108	0.118	0.126	0.126	0.118	0.108	0.100	0.048				
10	−1	−0.951	−0.809	−0.588	−0.309	0	0.309	0.588	0.809	0.951	1			$0.0^5 2323$
	0.043	0.089	0.095	0.104	0.111	0.115	0.111	0.104	0.095	0.089	0.043			
11	−1	−0.959	−0.841	−0.655	−0.415	−0.142	0.142	0.415	0.655	0.841	0.959	1		$0.0^6 4561$
	0.039	0.080	0.085	0.092	0.099	0.104	0.104	0.099	0.092	0.085	0.080	0.039		
12	−1	−0.966	−0.866	−0.707	−0.5	−0.259	0	0.259	0.5	0.707	0.866	0.966	1	$0.0^7 8856$
	0.036	0.074	0.077	0.083	0.089	0.094	0.095	0.094	0.089	0.083	0.077	0.074	0.036	

Left column: degree d of the polynomial fit. Middle columns: arcsin support points s_i from (23) in upper row and their A-optimal weights w_i from (4) in lower row, for $i = 0, \dots, d$. Right column: value (20) of the A-criterion.

TABLE 2

A-optimal designs on $[-1, 1]$

3	−1	−0.464	0.464	1										0.10661
	0.150	0.350	0.350	0.150										
4	−1	−0.677	0	0.677	1									0.02650
	0.104	0.250	0.290	0.250	0.104									
5	−1	−0.789	−0.291	0.291	0.789	1								$0.0^{2}6107$
	0.080	0.187	0.233	0.233	0.187	0.080								
6	−1	−0.853	−0.479	0	0.479	0.853	1							$0.0^{2}1340$
	0.065	0.147	0.185	0.205	0.185	0.147	0.065							
7	−1	−0.892	−0.604	−0.212	0.212	0.604	0.892	1						$0.0^{3}2839$
	0.055	0.121	0.149	0.175	0.175	0.149	0.121	0.055						
8	−1	−0.918	−0.692	−0.369	0	0.369	0.692	0.918	1					$0.0^{4}5860$
	0.047	0.102	0.123	0.148	0.159	0.148	0.123	0.102	0.047					
9	−1	−0.935	−0.754	−0.485	−0.167	0.167	0.485	0.754	0.935	1				$0.0^{4}1185$
	0.042	0.089	0.104	0.125	0.140	0.140	0.125	0.104	0.089	0.042				
10	−1	−0.948	−0.799	−0.574	−0.299	0	0.299	0.574	0.799	0.948	1			$0.0^{5}2358$
	0.037	0.078	0.090	0.107	0.123	0.130	0.123	0.107	0.090	0.078	0.037			
11	−1	−0.957	−0.834	−0.643	−0.404	−0.138	0.138	0.404	0.643	0.834	0.957	1		$0.0^{6}4630$
	0.034	0.070	0.079	0.092	0.107	0.117	0.117	0.107	0.092	0.079	0.070	0.034		
12	−1	−0.964	−0.860	−0.697	−0.489	−0.252	0	0.252	0.489	0.697	0.860	0.964	1	$0.0^{7}8989$
	0.031	0.064	0.070	0.081	0.094	0.105	0.109	0.105	0.094	0.081	0.070	0.064	0.031	

Left column: degree d of the polynomial fit. Middle columns: A-optimal support points in upper row and their A-optimal weights in lower row, for $i = 0, \ldots, d$. Right column: optimal value of the A-criterion.

TABLE 3
A-efficiencies [in %] of the arcsin support designs σ_A^d relative to the optimal designs τ_A^d

3	4	5	6	7	8	9	10	11	12
98.878	98.623	98.569	98.548	98.537	98.531	98.527	98.525	98.523	98.522

The A-optimal weights are readily computed from Corollary 4, and are exhibited in Table 1. Because the support points are the quantiles of the arcsin distribution, we call these designs *A-optimal arcsin support designs* σ_A^d.

In the line fit model ($d = 1$) and in the parabola fit model ($d = 2$), the arcsin support designs are overall optimal, with optimal values 1 and 0.375, respectively. In general, when the support points are allowed to vary over the experimental domain $[-1, 1]$, the A-optimal design τ_A^d for the full parameter vector θ in a dth degree polynomial model is distinct from the arcsin support design σ_A^d. The designs τ_A^d are given in Table 2, and were computed with the FORTRAN program Polyplan of Preitschopf (1989).

Up to degree $d \leq 12$ the efficiency

$$e_A(\sigma_A^d) = \frac{\phi_{-1}(M(\sigma_A^d))}{\phi_{-1}(M(\tau_A^d))}$$

of the arcsin support designs relative to the optimum value stays above 98.5%. The exact numbers are given in Table 3.

Finally, we present an example that demonstrates the loss of efficiency that is due to our restriction to linearly independent regression vectors. Consider the model for linear regression over the two-dimensional cube, $x \in \mathscr{X} = [0, 1]^2$. The A-optimal design for θ puts mass 0.42265 on the vertices $\binom{1}{0}$ and $\binom{0}{1}$, and 0.1547 on $\binom{1}{1}$ [see Cheng (1987), page 1598]. The A-optimal criterion value is 0.5359. Thus the A-optimal design sits on three support points, in the two-dimensional regression range $[0, 1]^2$.

We now check the linearly independent subsets of $\left\{\binom{1}{0}, \binom{0}{1}, \binom{1}{1}\right\}$. The first two vectors, by symmetry, receive equal weight 0.5, and the A-criterion has value 0.5. Hence this design has an A-efficiency of 93%. Alternatively, the vectors $\binom{1}{0}$ and $\binom{1}{1}$ get weights 0.5858 and 0.4142, respectively, with A-criterion value equal to 0.3431. The A-efficiency is merely 64%. It would seem that the loss of symmetry is responsible for the sharp decline in the objective function.

REFERENCES

ATWOOD, C. L. (1973). Sequences converging to D-optimal designs of experiments. *Ann. Statist.* **1** 342–352.

CHENG, C.-S. (1987). An application of the Kiefer–Wolfowitz equivalence theorem to a problem in Hadamard transform optics. *Ann. Statist.* **15** 1593–1603.

EHRENFELD, S. (1956). Complete class theorems in experimental design. In *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 57–67. Univ. California Press, Berkeley.

FEDEROV, V. V. (1972). *Theory of Optimal Experiments*. Academic, New York.

FELLMAN, J. (1974). On the allocation of linear observations. *Soc. Sci. Fenn. Comment. Phys.-Math.* **44** 27–77.

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

GAFFKE, N. and PUKELSHEIM, F. (1988). Admissibility and optimality of experimental designs. In *Model-Oriented Data Analysis. Proc. IIASA Workshop on Data Analysis, Eisenach 1987. Lecture Notes in Econom. and Math. Systems.* **297** 37–43. Springer, New York.

KARLIN, S. and STUDDEN, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Wiley, New York.

KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2** 849–879.

KITSOS, C. P., TITTERINGTON, D. M. and TORSNEY, B. (1988). An optimal design problem in rhythmometry. *Biometrics* **44** 657–671.

PREITSCHOPF, F. (1989). Bestimmung optimaler Versuchspläne in der polynomialen Regression. Ph.D. dissertation, Univ. Augsburg.

PUKELSHEIM, F. (1980). On linear regression designs which maximize information. *J. Statist. Plann. Inference* **4** 339–364.

PUKELSHEIM, F. (1987). Information increasing orderings in experimental design theory. *Internat. Statist. Rev.* **55** 203–219.

PUKELSHEIM, F. (1992). *Optimality Theory of Experimental Design in Linear Models*. Wiley, New York.

PUKELSHEIM, F. and TITTERINGTON, D. M. (1983). General differential and Lagrangian theory for optimal experimental design. *Ann. Statist.* **11** 1060–1068.

SILVEY, S. D., TITTERINGTON, D. M. and TORSNEY, B. (1978). An algorithm for optimal designs on a finite design space. *Comm. Statist. A—Theory Methods* **7** 1379–1389.

STUDDEN, W. J. (1977). Optimal designs for integrated variance in polynomial regression. In *Statistical Decision Theory and Related Topics II* (S. S. Gupta and D. S. Moore, eds.) 411–420. Academic, New York.

TORSNEY, B. (1981). Algorithms for a constrained optimization problem with applications in statistics and optimum design. Ph.D. dissertation, Univ. Glasgow.

TORSNEY, B. (1983). A moment inequality and monotonicity of an algorithm. In *Proceedings of the International Symposium on Semi-infinite Programming and Applications, Univ. Texas at Austin. Lecture Notes in Econom. and Math. Systems* (K. O. Kortanek and A. V. Fiacco, eds.) **215** 249–260. Springer, New York.

TORSNEY, B. (1988). Computing optimizing distributions with applications in design, estimation and image processing. In *Optimal Design and Analysis of Experiments* (Y. Dodge, V. V. Fedorov and H. P. Wynn, eds.) 361–370. North-Holland, Amsterdam.

INSTITUT FÜR MATHEMATIK
UNIVERSITÄT AUGSBURG
D-8900 AUGSBURG
GERMANY

DEPARTMENT OF STATISTICS
UNIVERSITY OF GLASGOW
GLASGOW, G12 8QW
SCOTLAND