*The Annals of Statistics*
1987, Vol. 15, No. 4, 1593–1603

# AN APPLICATION OF THE KIEFER–WOLFOWITZ EQUIVALENCE THEOREM TO A PROBLEM IN HADAMARD TRANSFORM OPTICS[1]

BY CHING-SHUI CHENG

*University of California, Berkeley*

Let $\Omega = \{\mathbf{x} = (x_1, \dots, x_n)^T: 0 \le x_i \le 1\}$ be the unit cube in $R^n$. For any probability measure $\xi$ on $\Omega$, let $\mathbf{M}(\xi) = \int_\Omega \mathbf{x}\mathbf{x}^T \xi(d\mathbf{x})$. Harwit and Sloane (1976) conjectured that if $\mathbf{X}^*$ is the incidence matrix of a balanced incomplete block design (BIBD) with $n$ treatments and $n$ blocks of size $(n+1)/2$, then $\mathbf{X}^*$ minimizes $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1}$ over the $n \times n$ matrices with entries $0 \le x_{ij} \le 1$. This arises from a problem in spectroscopy. In order to solve the conjecture, we consider the more general problem of maximizing $j_a(\mathbf{M}(\xi))$ over the probability measures on $\Omega$ for $-\infty \le a \le 1$, where $j_0(\mathbf{M}(\xi)) = \{\det \mathbf{M}(\xi)\}^{1/n}$, $j_{-\infty}(\mathbf{M}(\xi))$ = the minimum eigenvalue of $\mathbf{M}(\xi)$ and $j_a(\mathbf{M}(\xi)) = \{n^{-1}\operatorname{tr}[\mathbf{M}(\xi)]^a\}^{1/a}$ for other $a$'s. A complete solution is obtained by using the equivalence theorem in optimal design theory. Let $\xi_k$ be the uniform measure on the vertices of $\Omega$ with $k$ coordinates equal to 1. Then depending on the value of $a$, optimality is attained by $\xi_k$ or a mixture of $\xi_k$ and $\xi_{k+1}$ with $k \ge [(n+1)/2]$. Optimal $\xi$'s with a smaller support can be found by using BIBDs. It follows that if $n$ is odd and $\mathbf{X}^*$ is the block-treatment incidence matrix of a BIBD with $n$ treatments and $N$ blocks of size $(n+1)/2$, then $\mathbf{X}^*$ minimizes $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^a$ for all $a < 0$ and maximizes $\det(\mathbf{X}^T\mathbf{X})$ and $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^a$ for all $0 < a \le 1 - \ln(n/2+1)/\ln(n+1)$ over the $N \times n$ matrices with entries $0 \le x_{ij} \le 1$. Similar results are derived for the even case and the incidence matrices of BIBDs of larger block sizes.

**1. Introduction.** The purpose of this paper is to use the equivalence theorem in the approximate theory of optimal design to solve some outstanding problems in Hadamard transform optics and spring balance weighing design.

Hotelling proved in 1944 that if $\mathbf{X}$ is an $N \times n$ matrix $(N \ge n)$ with $-1 \le x_{ij} \le 1$ for all $i$, $j$, then all the diagonal entries of $(\mathbf{X}^T\mathbf{X})^{-1}$ are $\ge N^{-1}$. This implies that

$$\operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1} \ge n/N. \tag{1.1}$$

It also follows from Hotelling's proof that equality holds in (1.1) if and only if $\mathbf{X}^T\mathbf{X} = N\mathbf{I}_n$, i.e., all the entries of $\mathbf{X}$ are 1 or $-1$, and any two columns of $\mathbf{X}$ are orthogonal. When $N = n$, these are the well known *Hadamard matrices*.

What is the analogue of (1.1) when the entries of $\mathbf{X}$ are restricted to be between 0 and 1? This more difficult problem arises from the method of multiplexing in spectroscopy. A detailed discussion of the applications to Hadamard transform optics in spectroscopy and imaging can be found in Harwit and Sloane (1979). A brief introduction follows.

Received June 1986; revised March 1987.

[1]Research supported by National Science Foundation grant DMS-85-02784.

*AMS* 1980 *subject classifications*. Primary 62K05; secondary 62K10.

*Key words and phrases*. Balanced incomplete block design, equivalence theorem, Hadamard matrix, Hadamard transform optics, optimal design, spring balance weighing design.

Let $\mathbf{D}_{N,n}$ (respectively, $\overline{\mathbf{D}}_{N,n}$) be the collection of all the $N \times n$ matrices with entries $0 \le x_{ij} \le 1$ (respectively, $x_{ij} = 0$ or 1). Suppose an experimenter wants to determine the spectrum of a beam of light. When the noise is independent of the strength of the incident signal, more accurate estimates may be obtained by combining different frequency components in groups rather than measuring the intensity of each frequency component separately. An optical separator (such as a prism) separates the different frequency components of the incident light and focuses them into different locations (slots) of a mask. Each slot can be open or closed. If a slot is open, then the light is transmitted; if it is closed, then the light is absorbed. The total intensity of the frequency components not blocked is then measured. If there are $n$ frequency components and $N$ different measurements are to be made, then we have the model

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e}, \tag{1.2}$$

where $\mathbf{y}$ is the $N \times 1$ random vector of measurements, $\mathbf{w} = (w_1, w_2, \dots, w_n)^T$ is the vector of the intensities of the $n$ frequency components, $\mathbf{X} \in \overline{\mathbf{D}}_{N,n}$ and $\mathbf{e}$ is the random error with $\mathbf{E}(\mathbf{e}) = 0$ and $\operatorname{cov}(\mathbf{e}) = \sigma^2 \mathbf{I}_N$. The $(i, j)$th entry of $\mathbf{X}$ is equal to 0 if in the $i$th measurement, the $j$th frequency component is blocked, and is equal to 1, otherwise. One important question is how to design the mask (i.e., to choose $\mathbf{X}$). This is precisely what the statisticians called *spring balance weighing designs* [Raghavarao (1971), Chapter 17]. Furthermore, if the slots can be partially open, then the entries of the design matrix are real numbers between 0 and 1, i.e., $\mathbf{X} \in \mathbf{D}_{N,n}$. Under model (1.2), the least squares estimator of $\mathbf{w}$ is $\hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ with dispersion matrix $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$. A useful criterion is to choose an $\mathbf{X} \in \mathbf{D}_{N,n}$ (or $\overline{\mathbf{D}}_{N,n}$) such that $\sum_{i=1}^n \operatorname{var}(\hat{w}_i) = \sigma^2 \operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1}$ is minimized. Such an $\mathbf{X}$ is called *A-optimal.*

More generally, for $a \in (-\infty, 1]$, $a \neq 0$, one can consider the *maximization* of $j_a(\mathbf{X}^T\mathbf{X}) \equiv (n^{-1}\sum_{i=1}^n \lambda_i^a)^{1/a}$ over $\mathbf{D}_{N,n}$, where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\mathbf{X}^T\mathbf{X}$; the solutions are called $j_a$-optimal [Pukelsheim (1980)]. For $a < 0$, this is to minimize $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^a$, while for $a > 0$, $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^a$ is maximized. It is clear that $j_{-1}$-optimality is equivalent to $A$-optimality. Since $\lim_{a \to 0} j_a(\mathbf{X}^T\mathbf{X}) = \{\det(\mathbf{X}^T\mathbf{X})\}^{1/n}$ and $\lim_{a \to -\infty} j_a(\mathbf{X}^T\mathbf{X}) =$ the minimum eigenvalue of $\mathbf{X}^T\mathbf{X}$, the $j_a$-family also covers the well known $D$- and $E$-criteria as limiting cases. For this reason, we shall refer to the $D$- and $E$-criteria as $j_0$- and $j_{-\infty}$-criteria, respectively. The $j_a$-family is more general than the $\Phi_p$-criteria ($p \ge 0$) introduced by Kiefer (1974). Since $\Phi_p$-optimality is equivalent to $j_{-p}$-optimality, the $\Phi_p$-family corresponds to the $j_a$-criteria with $a \in [-\infty, 0]$. We also point out that the $j_a$-criteria are defined for $a \le 1$ only, since $(n^{-1}\sum_{i=1}^n \lambda_i^a)^{1/a}$ is no longer a concave function of $\mathbf{X}^T\mathbf{X}$ when $a > 1$.

Very little is known about the solution of the preceding problem. Harwit and Sloane (1976) conjectured that for any $\mathbf{X} \in \mathbf{D}_{n,n}$ (i.e., $N = n$),

$$\operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1} \ge 4n^2/(n+1)^2. \tag{1.3}$$

[Also see Sloane (1979).] They also noted that equality holds in (1.3) if $\mathbf{X}$ is an $S$-matrix. An $S$-matrix $\mathbf{S}_n$ of order $n$ is an $n \times n$ matrix obtained by deleting the first row and column of a normalized $(n+1) \times (n+1)$ Hadamard matrix and then changing 1's to 0's and $-1$'s to 1's. It is easy to see that

$$\mathbf{S}_n^T\mathbf{S}_n = \{(n+1)/4\}(\mathbf{I}_n + \mathbf{J}_n),$$

where $\mathbf{J}_n$ is the $n \times n$ matrix with all entries equal to 1, i.e., $\mathbf{S}_n$ is the incidence matrix of a balanced incomplete block design with $n$ treatments and $n$ blocks of size $(n+1)/2$. An $S$-matrix $\mathbf{S}_n$ is known to be $D$-optimal over $\overline{\mathbf{D}}_{n,n}$ [Raghavarao (1971), page 320]. If Harwit and Sloane's conjecture holds, then $\mathbf{S}_n$ is also $A$-optimal over $\mathbf{D}_{n,n}$. In the present paper, among other results, we shall show that $\mathbf{S}_n$ is not only $A$- but also $\Phi_p$-optimal over $\mathbf{D}_{n,n}$ for all $p \geq 0$.

Harwit and Sloane also asked if the minimum of $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1}$ over $\mathbf{D}_{n,n}$ can always be attained by a matrix in $\overline{\mathbf{D}}_{n,n}$. As noted earlier, if we restrict ourselves to optimizing over the matrices with 0–1 entries only, then we are in the domain of spring balance weighing designs. Recently, Jacroux and Notz (1983) made important contributions to the determination of optimal spring balance weighing designs. Their main results are:

1. For odd $n$, if there is an $\mathbf{X} \in \overline{\mathbf{D}}_{N,n}$ such that

$$\mathbf{X}^T\mathbf{X} = (4n)^{-1}N(n+1)(\mathbf{I}_n + \mathbf{J}_n),$$

then $\mathbf{X}$ is $A$-, $D$- and $E$-optimal over $\overline{\mathbf{D}}_{N,n}$.

2. For even $n$,

(a) if there is an $\mathbf{X} \in \overline{\mathbf{D}}_{N,n}$ such that

$$\mathbf{X}^T\mathbf{X} = \{N/4(n-1)\}\{n\mathbf{I}_n + (n-2)\mathbf{J}_n\},$$

then $\mathbf{X}$ is $A$- and $E$-optimal over $\overline{\mathbf{D}}_{N,n}$;

(b) if there is an $\mathbf{X} \in \overline{\mathbf{D}}_{N,n}$ such that

$$\mathbf{X}^T\mathbf{X} = \{4(n+1)\}^{-1}N(n+2)(\mathbf{I}_n + \mathbf{J}_n),$$

then $\mathbf{X}$ is $D$-optimal over $\overline{\mathbf{D}}_{N,n}$.

Letting $N = n$ in result 1, one obtains the $A$-, $D$- and $E$-optimality of $\mathbf{S}_n$ over $\overline{\mathbf{D}}_{n,n}$. However, this does not quite solve Harwit and Sloane's problem. Jacroux and Notz's method involves discrete optimization and depends heavily on the fact that the matrix entries are either 0's or 1's. They also had to deal with the three criteria $A$-, $D$- and $E$- separately. In this paper, we shall extend the preceding results to the whole $j_a$-family and at the same time strengthen the optimality to all of $\mathbf{D}_{N,n}$. The key idea is to use the approach of approximate designs due to Kiefer and Wolfowitz. There we have the celebrated equivalence theorem which is a powerful tool for constructing and verifying optimal designs. Once it is realized that the equivalence theorem can be used to solve the problem, the proof becomes very simple and stronger results can be obtained. The readers are referred to Kiefer (1974) for an extensive treatment of the approximate theory of optimal design.

This paper is organized as follows. In Section 2, the problem is formulated in terms of approximate designs. The equivalence theorem is stated there. A quick application shows how Jacroux and Notz's results can be proved by using the equivalence theorem. A complete solution of optimal approximate designs is given in Section 3, followed by some discussions and applications. The proof is presented in Section 4.

**2. Approximate theory.** Let $\Omega$ be the unit cube $\{\mathbf{x} = (x_1, \ldots, x_n)^T: 0 \le x_i \le 1\}$ in $R^n$ and $\Xi$ be the set of all the probability measures on the Borel subsets of $\Omega$. For any $\xi \in \Xi$, define the *information matrix*

$$\mathbf{M}(\xi) = \int_\Omega \mathbf{x}\mathbf{x}^T \xi(d\mathbf{x}).$$

Let $j_0(\mathbf{M}(\xi)) = \{\det \mathbf{M}(\xi)\}^{1/n}$,

$$j_{-\infty}(\mathbf{M}(\xi)) = \text{the minimum eigenvalue of } \mathbf{M}(\xi)$$

and for $-\infty < a \le 1$, $a \ne 0$, let

$$j_a(\mathbf{M}(\xi)) = \left\{n^{-1} \sum_{i=1}^{n} \lambda_i^a(\xi)\right\}^{1/a},$$

where $\lambda_1(\xi), \ldots, \lambda_n(\xi)$ are the eigenvalues of $\mathbf{M}(\xi)$. For $a \in [-\infty, 1]$, we shall solve the problem of

$$(*) \qquad \text{maximizing } j_a(\mathbf{M}(\xi)) \text{ over } \xi \in \Xi.$$

For any $\mathbf{X} \in \mathbf{D}_{N,n}$, write

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_N^T \end{pmatrix}, \quad \text{where } \mathbf{x}_i \in \Omega.$$

Then

$$\mathbf{X}^T\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^T.$$

If $\xi_{\mathbf{X}}$ is the probability measure on $\Xi$ which assigns probability $1/N$ to each $\mathbf{x}_i$, then

$$\mathbf{M}(\xi_{\mathbf{X}}) = N^{-1}\mathbf{X}^T\mathbf{X}.$$

Therefore, if $\xi_{\mathbf{X}}$ is $j_a$-optimal over $\Xi$, then $\mathbf{X}$ is $j_a$-optimal over $\mathbf{D}_{N,n}$. The matrices in $\mathbf{D}_{N,n}$ are called *exact designs*, while the measures $\xi$ in $\Xi$ are called *approximate designs*.

The equivalence theorem was first proved by Kiefer and Wolfowitz (1960) for the $D$-criterion. Kiefer (1974) extended it to general criteria. Other versions can be found, for instance, in Pukelsheim (1980) and Pukelsheim and Titterington (1983). When specialized to the $j_a$-criteria and to the present setting, it is reduced to the following form:

THEOREM 2.1 (Equivalence theorem for the $j_a$-criteria). *Suppose* $\mathbf{M}(\xi^*)$ *is positive definite. Then for any* $-\infty < a < 1$, *the following three statements are* *equivalent*:

(a) $\xi^*$ *maximizes* $j_a(\mathbf{M}(\xi))$ *over* $\xi \in \Xi$.
(b) $\xi^*$ *minimizes* $\max_{\mathbf{x}\in\Omega}\mathbf{x}^T\{\mathbf{M}(\xi)\}^{a-1}\mathbf{x}$ *over* $\xi \in \Xi$.
(c) $\max_{\mathbf{x}\in\Omega}\mathbf{x}^T\{\mathbf{M}(\xi^*)\}^{a-1}\mathbf{x} = \operatorname{tr}\{\mathbf{M}(\xi^*)\}^a$.

*Moreover, an optimal* $\xi^*$ *must be supported on points where the maximum of* $\mathbf{x}^T\{\mathbf{M}(\xi^*)\}^{a-1}\mathbf{x}$ *over* $\Omega$ *is attained and all the optimal* $\xi^*$*'s have the same information matrix* $\mathbf{M}(\xi^*)$.

Usually, the equivalence theorem can be applied in the following way. The last statement in Theorem 2.1 about the support of an optimal $\xi^*$ can be used to guess a good candidate; condition (c), which is easier to check than (a), can then be used to verify its optimality.

As an application, let us first see how Jacroux and Notz's results can be proved by using Theorem 2.1. Consider the case where $N$ is odd. If $\mathbf{X}$ is an $N \times n$ (0,1) matrix such that $\mathbf{X}^T\mathbf{X} = (4n)^{-1}N(n+1)(\mathbf{I}_n + \mathbf{J}_n)$, then $\mathbf{M}(\xi_{\mathbf{X}}) = N^{-1}\mathbf{X}^T\mathbf{X} = (4n)^{-1}(n+1)(\mathbf{I}_n + \mathbf{J}_n)$. In order to show that $\xi_{\mathbf{X}}$ is $j_a$-optimal over $\Xi$, it is enough to verify

$$\max_{\mathbf{x}\in\Omega}\mathbf{x}^T\left\{(4n)^{-1}(n+1)(\mathbf{I}_n+\mathbf{J}_n)\right\}^{a-1}\mathbf{x} = \operatorname{tr}\left\{(4n)^{-1}(n+1)(\mathbf{I}_n+\mathbf{J}_n)\right\}^{a},$$

i.e.,

$$\max_{\mathbf{x}\in\Omega}\mathbf{x}^T(\mathbf{I}_n+\mathbf{J}_n)^{a-1}\mathbf{x} = (4n)^{-1}(n+1)\operatorname{tr}(\mathbf{I}_n+\mathbf{J}_n)^{a}. \tag{2.1}$$

Since $(\mathbf{I}_n+\mathbf{J}_n)^a = \mathbf{I}_n - n^{-1}\{1-(n+1)^a\}\mathbf{J}_n$, the right side of (2.1) is $(4n)^{-1}(n+1)\{n-1+(n+1)^a\}$. The proof is finished by showing

$$\max_{\mathbf{x}\in\Omega}\mathbf{x}^T\left[\mathbf{I}_n - n^{-1}\left\{1-(n+1)^{a-1}\right\}\mathbf{J}_n\right]\mathbf{x} = (4n)^{-1}(n+1)\left\{n-1+(n+1)^a\right\}.$$

It is easy to see that if $a \le 1 - \ln(n/2+1)/\ln(n+1)$, then the maximum of $\mathbf{x}^T[\mathbf{I}_n - n^{-1}\{1-(n+1)^{a-1}\}\mathbf{J}_n]\mathbf{x}$ over $\mathbf{x}\in\Omega$ is attained at vertices of $\Omega$ with $(n+1)/2$ coordinates equal to 1 and $(n-1)/2$ coordinates equal to 0. For such an $\mathbf{x}$, a simple computation shows that

$$\mathbf{x}^T\left[\mathbf{I}_n - n^{-1}\left\{1-(n+1)^{a-1}\right\}\mathbf{J}_n\right]\mathbf{x} = (4n)^{-1}(n+1)\left\{(n-1)+(n+1)^a\right\}.$$

So condition (c) in Theorem 2.1 is satisfied and $\xi_{\mathbf{X}}$ is $j_a$-optimal over $\Xi$ for all $-\infty < a \le 1 - \ln(n/2+1)/\ln(n+1)$. By passing to the limit ($a \to -\infty$), we see that $\xi_{\mathbf{X}}$ is also $j_{-\infty}$-optimal. Consequently, $\mathbf{X}$ is $j_a$-optimal over all the $N \times n$ matrices with $0 \le x_{ij} \le 1$ for all $a \le 1 - \ln(n/2+1)/\ln(n+1)$; in particular, it is $A$-, $D$- and $E$-optimal over the $N \times n$ (0,1) matrices. By letting $N = n$ and $a = -1$, we have also proven Harwit and Sloane's conjecture for the case where $n$ is odd. Jacroux and Notz's results for even $N$ can be handled in the same fashion.

The preceding proof is so simple partly because we already have the right candidate. In case no good candidate is available, as commented earlier, the equivalence theorem can often provide some good guess. In the next section, we shall give a complete solution to problem (∗). In the proof, one can also see how the designs found by Jacroux and Notz emerge naturally from the equivalence theorem.

**3. Main results.** We first state and discuss the solution of problem (∗). The proof is deferred to the next section.

For any integer $k$ such that $[(n+1)/2] \le k \le n$, where $[x]$ is the integral part of $x$, let $\xi_k$ be the uniform measure on all the vertices of $\Omega$ with $k$ coordinates equal to 1 and $n-k$ coordinates equal to 0. Define two functions $f$ and $g$ on the integers $[(n+1)/2] \le k \le n$ by

$$(3.1)\quad g(k) = \begin{cases} \dfrac{\ln\{(2k+1-n)/(2k+1)\}}{\ln\{k(n-1)/(n-k)\}} + 1, & \text{if } \left[\dfrac{(n+1)}{2}\right] \le k \le n-1, \\ 1, \quad \text{if } k = n, & \end{cases}$$

and

$$(3.2)\qquad f(k) = \begin{cases} \dfrac{\ln\{(2k-1-n)/(2k-1)\}}{\ln\{k(n-1)/(n-k)\}} + 1, \\ \qquad \text{if } \left[\dfrac{(n+1)}{2}\right] + 1 \le k \le n-1, \\ -\infty, \quad \text{if } k = \left[\dfrac{(n+1)}{2}\right], \\ 1, \quad \text{if } k = n. \end{cases}$$

Then it is clear that $-\infty = f([(n+1)/2]) < g([(n+1)/2]) < f(k) < g(k) < f(k+1) < g(k+1) < 1 = f(n) = g(n)$ for all $[(n+1)/2] + 1 \le k \le n-2$. So $\{[f(k), g(k)]\}_{k=[(n+1)/2]}^{n}$ and $\{(g(k), f(k+1))\}_{k=[(n+1)/2]}^{n-1}$ together form a partition of $[-\infty, 1]$. We state our main result as

THEOREM 3.1. *Suppose $k$ is an integer such that $[(n+1)/2] \le k \le n$. For $f(k) \le a \le g(k)$, the design $\xi_k$ is $j_a$-optimal over $\Xi$, while for $k < n$ and $g(k) < a < f(k+1)$, the mixture $\varepsilon\xi_k + (1-\varepsilon)\xi_{k+1}$, with $\varepsilon$ given by*

$$(3.3)\quad \frac{(k+1)^2(n-1) - (k+1)(n-k-1)\{(2k+1-n)/(2k+1)\}^{1/(a-1)}}{(2k+1)(n-1) + (2k+1-n)\{(2k+1-n)/(2k+1)\}^{1/(a-1)}},$$

*is $j_a$-optimal over $\Xi$.*

Since $\{[f(k), g(k)]\}_{k=[(n+1)/2]}^{n}$ and $\{(g(k), f(k+1))\}_{k=[(n+1)/2]}^{n-1}$ together form a partition of $[-\infty, 1]$, the preceding theorem provides $j_a$-optimal designs for all $a \in [-\infty, 1]$. Each optimal design is a $\xi_k$ or a mixture of $\xi_k$ and $\xi_{k+1}$ with

$k \geq [(n+1)/2]$. It can be seen that the $\varepsilon$ defined in (3.3) is a strictly increasing function of $a \in (g(k), f(k+1))$ and that as $a$ ranges from $g(k)$ to $f(k+1)$, the $\varepsilon$'s cover all of (0,1). Therefore, any mixture of $\xi_k$ and $\xi_{k+1}$ is $j_a$-optimal for a single value of $a$ in$(g(k), f(k+1))$, whereas except for $k = n$, $\xi_k$ is optimal for infinitely many $a$'s. As $a$ increases from $-\infty$ to 1, the support of an optimal design shifts from the vertices of $\Omega$ with $[(n+1)/2]$ 1's to those with more 1's. At $a = 1$, the optimal design puts all the mass on the point with all the coordinates equal to 1.

Among the optimal designs in Theorem 3.1, $\xi_{[(n+1)/2]}$ is the most important. It is $j_a$-optimal for the widest range of $a$'s which also includes the more important and commonly used criteria. For instance, when $n$ is odd, $f((n+1)/2) = -\infty$ and $g((n+1)/2) = 1 - \ln(n/2+1)/\ln(n+1)$. So $\xi_{(n+1)/2}$ is $j_a$-optimal for all $a \in [-\infty, 1 - \ln(n/2+1)/\ln(n+1)]$. Since $1 - \ln(n/2+1)/\ln(n+1) > 0$, $\xi_{(n+1)/2}$ is $j_a$-optimal for all $a \in [-\infty, 0]$, i.e., it is optimal with respect to Kiefer's $\Phi_p$-criteria for all $p \geq 0$; in particular, it is $A$-, $D$- and $E$-optimal. When $n$ is even, $g(n/2) = 1 - \ln(n+1)/\ln(n-1)$. So $\xi_{n/2}$ is $j_a$-optimal for all $a \in [-\infty, 1 - \ln(n+1)/\ln(n-1)]$. Unlike the odd case, this does not quite cover all the $j_a$-criteria with nonpositive $a$'s. The right end $1 - \ln(n+1)/\ln(n-1)$ is less than, although quite close to 0. As a consequence, $\xi_{n/2}$ is not $D$-optimal. However, since $1 - \ln(n+1)/\ln(n-1) \geq -1$ if $n \geq 3$, $\xi_{n/2}$ is $A$-optimal for all even $n \geq 4$; it is also $E$-optimal for all even $n$. Due to the importance of the $\Phi_p$-criteria, we state the results for $\Phi_p$-optimal designs in the following corollary.

COROLLARY 3.2. (i) *If $n$ is odd, then $\xi_{(n+1)/2}$ is $\Phi_p$-optimal over $\Xi$ for all $0 \leq p \leq \infty$.*

(ii) *If $n$ is even, then $\xi_{n/2}$ is $\Phi_p$-optimal over $\Xi$ for all $p \geq \ln(n+1)/\ln(n-1) - 1$. For $0 \leq p < \ln(n+1)/\ln(n-1) - 1$, $\varepsilon\xi_{n/2} + (1-\varepsilon)\xi_{n/2+1}$ is $\Phi_p$-optimal over $\Xi$, where*

$$\varepsilon = \frac{(n+2)^2(n-1) - (n^2-4)(n+1)^{1/(p+1)}}{4\{n^2 + (n+1)^{1/(p+1)} - 1\}}. \tag{3.4}$$

As a matter of fact, we first obtained our results for the $\Phi_p$-criteria. The distinction between the even and odd dimensions as exhibited in Corollary 3.2 was somewhat puzzling at the time. It was Professor F. Pukelsheim's suggestion to look into the $j_a$-family which led to Theorem 3.1 and a much better understanding of the problem.

It was pointed out earlier that when $n$ is even, $\xi_{n/2}$ is $A$- and $E$-optimal, but not $D$-optimal. Letting $p = 0$ in (3.4), we have $\varepsilon = (n+2)/2(n+1)$. So $\xi_D \equiv \{(n+2)/2(n+1)\}\xi_{n/2} + \{n/2(n+1)\}\xi_{n/2+1}$is $D$-optimal. This design can be described in simple terms. Since $\xi_{n/2}$ is supported on $\binom{n}{n/2}$ points, $\xi_D$ assigns to each vertex of $\Omega$ with $n/2$ 1's a weight of $(n+2)\left\{2(n+1)\binom{n}{n/2}\right\}^{-1}$, which is equal to $\left\{\binom{n}{n/2} + \binom{n}{n/2+1}\right\}^{-1}$. Similarly, it can be shown that $\xi_D$ also assigns a weight of $\left\{\binom{n}{n/2} + \binom{n}{n/2+1}\right\}^{-1}$ to each vertex of $\Omega$ with $n/2 + 1$ 1's. Accordingly, $\xi_D$ is simply the uniform measure on all the vertices of $\Omega$ with $n/2$ and $n/2 + 1$ 1's. We state this as

COROLLARY 3.3. *If n is even, then the uniform measure $\xi_D$ on all the vertices of $\Omega$ with $n/2$ and $n/2 + 1$ 1's is D-optimal over $\Xi$.*

The solutions to the approximate design problem (*) can be used to derive upper bounds on $j_a(\mathbf{X}^T\mathbf{X})$ for $\mathbf{X} \in \mathbf{D}_{N,n}$. This is because if $\xi^*$ is $j_a$-optimal over $\Xi$, then $j_a(\mathbf{X}^T\mathbf{X}) \le N^{-1}j_a(\mathbf{M}(\xi^*))$ for all $\mathbf{X} \in \mathbf{D}_{N,n}$. It is easy to see that when $n$ is odd, $\mathbf{M}(\xi_{(n+1)/2}) = (n+1)(4n)^{-1}(\mathbf{I}_n + \mathbf{J}_n)$, and when $n$ is even, $\mathbf{M}(\xi_{n/2}) = \{4(n-1)\}^{-1}\{n\mathbf{I}_n + (n-2)\mathbf{J}_n\}$. Therefore, we have, for $a = -1$:

COROLLARY 3.4. *For any $\mathbf{X} \in \mathbf{D}_{N,n}$ with $N \ge n$,*

$$\operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1} \ge N^{-1}\operatorname{tr}\big(\mathbf{M}(\xi_{(n+1)/2})\big)^{-1} = 4n^3/N(n+1)^2, \quad \textit{if n is odd}$$

*and*

$$\operatorname{tr}(\mathbf{X}^T\mathbf{X})^{-1} \ge N^{-1}\operatorname{tr}\big(\mathbf{M}(\xi_{n/2})\big)^{-1} = 4(n^2 - 2n + 2)/nN, \quad \textit{if n is even and } n \ge 4.$$

Lower bounds for other criteria can be similarly derived. Note that the inequality in Corollary 3.4 for odd $n$ reduces to (1.3) when $N = n$.

The designs listed in Theorem 3.1 are not the only $j_a$-optimal designs. Optimal designs with smaller supports can be found by using BIBDs. It is clear that if $\mathbf{N}$ is the block-treatment incidence matrix of a BIBD with $n$ treatments and block size $k$, then the uniform measure on all the row vectors of $\mathbf{N}$ has the same information matrix as $\xi_k$. Thus in Theorem 3.1, we can replace $\xi_k$ with the uniform measure on the row vectors of the block-treatment incidence matrix of an *arbitrary* BIBD with $n$ treatments and block size $k$. A $D$-optimal design for even $n$ with a smaller support than the design described in Corollary 3.3 can be constructed as follows. Let $d_1$ and $d_2$ be BIBDs with $n$ treatments and block sizes $n/2$ and $n/2 + 1$, respectively. If $r_1 = r_2$, where $r_i$ is the number of replications of each treatment in $d_i$, $i = 1, 2$, then it can be seen that $\xi_D$ has the same information matrix as the uniform measure on the row vectors of $\binom{\mathbf{N}_1}{\mathbf{N}_2}$, where $\mathbf{N}_i$ is the block-treatment incidence matrix of $d_i$. In view of the previous observations, we have

COROLLARY 3.5. (i) *If n is odd and $\mathbf{X}^*$ is the block-treatment incidence matrix of a* BIBD *with n treatments and N blocks of size $(n+1)/2$, then $\mathbf{X}^*$ minimizes* $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^a$ *for all $a < 0$ and maximizes* $\det(\mathbf{X}^T\mathbf{X})$ *and* $\operatorname{tr}(\mathbf{X}^T\mathbf{X})^a$ *for all $0 < a \le 1 - \ln(n/2+1)/\ln(n+1)$ over $\mathbf{X} \in \mathbf{D}_{N,n}$.*

(ii) *If $n$ is even and* $\mathbf{X}^*$ *is the block-treatment incidence matrix of a* BIBD *with $n$ treatments and $N$ blocks of size $n/2$, then* $\mathbf{X}^*$ *minimizes* $\mathrm{tr}(\mathbf{X}^T\mathbf{X})^a$ *over* $\mathbf{X} \in \mathbf{D}_{N,n}$ *for all* $a \le 1 - \ln(n+1)/\ln(n-1)$.

(iii) *Suppose $n$ is even. Let $d_1$ be a* BIBD *with $n$ treatments and $N_1$ blocks of size $n/2$ and $d_2$ be a* BIBD *with $n$ treatments and $N_2$ blocks of size $n/2 + 1$ such that $r_1 = r_2$. Then* $\mathbf{X}^* = \begin{pmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \end{pmatrix}$ *is D-optimal over* $\mathbf{D}_{N_1+N_2,\,n}$, *where* $\mathbf{N}_i$ *is the block-treatment incidence matrix of $d_i$.*

(iv) *For* $[(n+1)/2] < k < n$, *the block-treatment incidence matrix* $\mathbf{X}^*$ *of a* BIBD *with $n$ treatments and $N$ blocks of size $k$ maximizes* $\mathrm{tr}(\mathbf{X}^T\mathbf{X})^a$ *over* $\mathbf{X} \in \mathbf{D}_{N,n}$ *for all* $a \in [f(k), g(k)]$, *where $f(k)$ and $g(k)$ are defined in* (3.1) *and* (3.2).

The results by Jacroux and Notz (1983) quoted in Section 1 are special cases of Corollary 3.5.

**4. Proof of Theorem 3.1.** The case $a = 1$ in Theorem 3.1 is trivial, so we shall only consider the case $-\infty < a < 1$. One can prove the theorem by computing the information matrices of the optimal designs listed there and verifying condition (c) in Theorem 2.1. This of course does not give the complete picture. Instead we shall actually show how these designs are derived; particularly how the weight $\varepsilon$ in (3.3) is determined.

First of all, notice that problem $(*)$ is permutation invariant. For any permutation $\pi$ of integers $1, 2, \ldots, n$ and any $\mathbf{x} = (x_1, \ldots, x_n)^T \in \Omega$, let $\pi(\mathbf{x}) = (x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)})^T$. Then obviously $j_a\{\mathbf{M}(\xi_\pi)\} = j_a\{\mathbf{M}(\xi)\}$, where $\xi_\pi$ is defined by $\xi_\pi(B) = \xi(\pi(B))$ for any Borel subset $B$ of $\Omega$. Since $j_a$ is concave for $a \le 1$, problem $(*)$ must have a permutation-invariant solution, i.e., there exists a $j_a$-optimal $\bar{\xi}^*$ such that $\bar{\xi}^* = \bar{\xi}^*_\pi$ for all $\pi$. For such $\bar{\xi}^*$, $\mathbf{M}(\bar{\xi}^*)$ must also be permutation-invariant, i.e., $\mathbf{M}(\bar{\xi}^*) = \alpha\mathbf{I}_n + \beta\mathbf{J}_n$ for some constants $\alpha$ and $\beta$. By Theorem 2.1, all the optimal designs have the same information matrix; it follows that $\mathbf{M}(\xi^*) = \alpha\mathbf{I}_n + \beta\mathbf{J}_n$ for any $j_a$-optimal $\xi^*$, permutation-invariant or not. This yields a substantial reduction of the problem. Now we only have to determine $\alpha$ and $\beta$.

By Theorem 2.1, an optimal $\xi^*$ must be supported on points where the maximum of $\phi(\mathbf{x}) \equiv \mathbf{x}^T\{\mathbf{M}(\xi^*)\}^{a-1}\mathbf{x}$ over $\Omega$ is attained. For any $1 \le i \le n$, if we fix all the coordinates of $\mathbf{x}$ except $x_i$, then $\phi(\mathbf{x})$ is a quadratic function of $x_i$ with positive leading coefficient, whose maximum over $[0, 1]$ is attained at 0 or 1. Therefore, $\xi^*$ must be supported on the points with all the coordinates equal to 0 or 1, i.e., vertices of $\Omega$.

Now we determine the number of zero coordinates a point in the support of an optimal $\xi^*$ can have. Let $\mathbf{x}$ be a vertex of $\Omega$ in the support of $\xi^*$ with $k$ 1's and $(n-k)$ 0's. Since $\{\mathbf{M}(\xi^*)\}^{a-1} = \{\alpha\mathbf{I}_n + \beta\mathbf{J}_n\}^{a-1} = \alpha^{a-1}\mathbf{I}_n - n^{-1}\{\alpha^{a-1} - (\alpha + n\beta)^{a-1}\}\mathbf{J}_n$,

$$\phi(\mathbf{x}) = \alpha^{a-1}\mathbf{x}^T\mathbf{x} - n^{-1}\left\{\alpha^{a-1} - (\alpha + n\beta)^{a-1}\right\}\mathbf{x}^T\mathbf{J}_n\mathbf{x} \tag{4.1}$$

$$= \alpha^{a-1}k - n^{-1}\left\{\alpha^{a-1} - (\alpha + n\beta)^{a-1}\right\}k^2.$$

The maximum of (4.1) is attained at the integer(s) closest to $k^* \equiv n\alpha^{a-1}/2\{\alpha^{a-1} - (\alpha + n\beta)^{a-1}\}$. Clearly $k^* > n/2$. Therefore, an optimal $\xi^*$ must be supported on vertices of $\Omega$ with at least $[(n+1)/2]$ coordinates equal to 1. In fact, since (4.1) is a quadratic function of $k$, one concludes that

if $k^* > [k^*] + \frac{1}{2}$, then $\xi^*$ must be supported on vertices of $\Omega$ with $[k^*] + 1$ 1's;

if $k^* < [k^*] + \frac{1}{2}$, then $\xi^*$ must be supported on vertices of $\Omega$ with $[k^*]$ 1's;

if $k^* = [k^*] + \frac{1}{2}$, then $\xi^*$ must be supported on vertices of $\Omega$ with $[k^*]$ and $[k^*] + 1$ 1's.

The solution is clear now. Since $\mathbf{M}(\xi^*)$ must be of the form $\alpha\mathbf{I}_n + \beta\mathbf{J}_n$, in view of the previous results on the support of $\xi^*$, one can try the uniform measures $\xi_k$ or mixtures of $\xi_k$ and $\xi_{k+1}$ with $k \geq [(n+1)/2]$.

A straightforward computation shows that

$$\mathbf{M}(\xi_k) = n^{-1}(n-1)^{-1}k(n-k)\left\{\mathbf{I}_n + (n-k)^{-1}(k-1)\mathbf{J}_n\right\}.$$

Therefore,

$$\begin{aligned}\{\mathbf{M}(\xi_k)\}^a &= n^{-a}(n-1)^{-a}k^a(n-k)^a \\ &\quad\times\left\{\mathbf{I}_n - n^{-1}\left[1 - \{k(n-1)/(n-k)\}^a\right]\mathbf{J}_n\right\}.\end{aligned}$$

Let $h(t) = \mathbf{x}^T\{\mathbf{M}(\xi_k)\}^{a-1}\mathbf{x}$, where $\mathbf{x}$ is a vertex of $\Omega$ with $t$ 1's. The right side of Theorem 2.1(c), $\operatorname{tr}\{\mathbf{M}(\xi_k)\}^a$, is $n^{-a}(n-1)^{-a}k^a(n-k)^a[n-1+\{k(n-1)/(n-k)\}^a]$, which can be shown to equal $h(k)$. Therefore, condition (c) in Theorem 2.1 holds for $\xi_k$, i.e., $\xi_k$ is $j_a$-optimal if and only if

$$h(k) \geq \max\{h(k-1), h(k+1)\}. \tag{4.2}$$

By direct computation, (4.2) is equivalent to

$$(2k-1-n)/(2k-1) \leq \{k(n-1)/(n-k)\}^{a-1} \leq (2k+1-n)/(2k+1).$$

Taking logarithms, we conclude that $\xi_k$ is $j_a$-optimal for all $a \in [f(k), g(k)]$.

For $g(k) < a < f(k+1)$, we try mixtures of $\xi_k$ and $\xi_{k+1}$. Let $\xi(\varepsilon) = \varepsilon\xi_k + (1-\varepsilon)\xi_{k+1}$. Then

$$\begin{aligned}\{\mathbf{M}(\xi(\varepsilon))\}^{a-1} &= \{(k+1)(n-k-1) + \varepsilon(2k+1-n)\}^{a-1} \\ &\quad\times\{n(n-1)\}^{1-a}\left[\mathbf{I}_n - n^{-1}\left\{1 - (K(\varepsilon))^{a-1}\right\}\mathbf{J}_n\right],\end{aligned} \tag{4.3}$$

where

$$K(\varepsilon) = 1 + n\{k(k+1) - 2\varepsilon k\}/\{(k+1)(n-k-1) + \varepsilon(2k+1-n)\}. \tag{4.4}$$

Let $H(t) = \mathbf{x}^T\{\mathbf{M}(\xi(\varepsilon))\}^{a-1}\mathbf{x}$, where $\mathbf{x}$ is a vertex of $\Omega$ with $t$ 1's. Recall that an optimal design $\xi^*$ must be supported on points where the maximum of $\mathbf{x}^T\{\mathbf{M}(\xi^*)\}^{a-1}\mathbf{x}$ is attained. Because $\xi(\varepsilon)$ is supported on vertices of $\Omega$ with $k$ 1's as well as those with $k+1$ 1's, if $\xi(\varepsilon)$ is $j_a$-optimal, then we must have

$H(k) = H(k+1)$. Comparing $H(k)$ with $H(k+1)$, we are led to

$$K(\varepsilon) = \{(2k+1-n)/(2k+1)\}^{1/(a-1)}, \tag{4.5}$$

where $K(\varepsilon)$ is defined in (4.4). Solving (4.5), we obtain the $\varepsilon$ given in (3.3). It is easy to see that the condition $g(k) < a < f(k+1)$ implies $0 < \varepsilon < 1$, so $\xi(\varepsilon)$ is a legitimate probability measure.

To finish the proof, we must show that $\xi(\varepsilon)$ satisfies condition (c) in Theorem 2.1. Since $H(t)$ is a quadratic function of $t$ with negative leading coefficient and $H(k) = H(k+1)$, the maximum of $H(t)$ over the integers is attained at $t = k$ and $k+1$. It follows that $\max_{\mathbf{x} \in \Omega} \mathbf{x}^T\{\mathbf{M}(\xi(\varepsilon))\}^{a-1}\mathbf{x} = H(k)$. A routine but somewhat tedious computation, with the help of (3.3) and (4.5), shows that $H(k) = \operatorname{tr}\{\mathbf{M}(\xi(\varepsilon))\}^a$. Therefore, $\xi(\varepsilon)$ satisfies condition (c) in Theorem 2.1 and the proof is completed.

**Acknowledgment.** The author is grateful to Professor F. Pukelsheim for helpful discussions.

## REFERENCES

HARWIT, M. and SLOANE, N. J. A. (1976). Masks for Hadamard transform optics, and weighing designs. *Appl. Optics* **15** 107–114.

HARWIT, M. and SLOANE, N. J. A. (1979). *Hadamard Transform Optics*. Academic, New York.

HOTELLING, H. (1944). Some improvements in weighing and other experimental techniques. *Ann. Math. Statist.* **15** 297–306.

JACROUX, M. and NOTZ, W. (1983). On the optimality of spring balance weighing designs. *Ann. Statist.* **11** 970–978.

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

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math.* **14** 363–366.

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

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

RAGHAVARAO, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments*. Wiley, New York.

SLOANE, N. J. A. (1979). Multiplexing methods in spectroscopy. *Math. Mag.* **52** 71–80.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720