D-Optimality for Regression Designs: A Review

R. C. St. John
Bowling Green State University
Department of Quantitative
Analysis and Control
Bowling Green, Ohio 43403

and

N. R. Draper
University of Wisconsin
Statistics Department
1210 W. Dayton Street
Madison, Wisconsin 53706

After stating the model and the design problem, we briefly present the results for regression design prior to the work of Kiefer and Wolfowitz. We then review the major results of Kiefer and Wolfowitz, particularly those on the theory of design, as well as the way the criterion has been extended to non-linear models. Finally, we discuss algorithms for constructing D-optimum designs.

KEY WORDS
D-Optimum
Design

1. Model and Notation

In this section we introduce our notation for the linear regression model, and the assumptions about the model and the design region. We consider the linear model

$$y_i = \mathbf{f}'(\mathbf{x}_i)\boldsymbol{\beta} + \epsilon_i, \qquad i = 1, 2, \cdots, n, \tag{1}$$

which we express in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2}$$

The vector $\mathbf{y}$ is an $n \times 1$ vector of observations; $\mathbf{X}$ is an $n \times p$ matrix, with row i containing $\mathbf{f}'(\mathbf{x}_i)$; $\mathbf{x}_i$ is a $q \times 1$ vector of predictor variables; $\mathbf{f}'(\mathbf{x}_i)$ is a $p \times 1$ vector which depends on the form of the response function assumed; $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters; $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of independently and identically distributed random variables, with mean zero and variance σ^2. The experimental region is denoted by χ, and we assume that χ is compact and that $f_i(\mathbf{x}_i)$'s are continuous on χ.

We assume that least squares estimates of the parameters $\boldsymbol{\beta}$ are to be obtained. These are given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{3}$$

and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{4}$$

Then, at point $\mathbf{x}\epsilon\chi$, the predicted response is

$$\hat{y}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\hat{\boldsymbol{\beta}}, \tag{5}$$

with variance

$$v(\hat{y}(\mathbf{x})) = \sigma^2\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}). \tag{6}$$

Received July, 1974; revised October, 1974

The design problem consists of selecting vectors $\mathbf{x}_i$, $i = 1, 2, \cdots, n$ from χ such that the design defined by these n vectors is, in some defined sense, optimal. By and large, solutions to this problem consist of developing some sensible criterion based on the model (2), and using it to obtain optimal designs.

2. Background

Smith (1918) was one of the first to state a criterion and obtain optimal experimental designs for regression problems. For polynomial regression of order $p - 1$ in one variable over the design region $\chi = [-1, 1]$, she proposed the criterion

$$\min_{\{\mathbf{x}_i, i=1,2,\cdots,n\}} \max_{\mathbf{x}\epsilon\chi} v(\hat{y}(\mathbf{x})). \tag{7}$$

This criterion was later called G-optimality by Kiefer and Wolfowitz (1959), and it has taken on considerable importance in the theory and construction of optimal designs. Using this criterion Smith obtained designs for various values of p. Guest (1958) showed that, for a polynomial of degree $p - 1$, the allocation of points according to Smith's min max criterion (7) could be obtained by finding the zeroes of the derivative of a Legendre polynomial.

Wald (1943) proposed the criterion of maximizing the determinant of $\mathbf{X}'\mathbf{X}$ as a means of maximizing the local power of the F-ratio for testing a linear hypothesis on the parameters of certain fixed-effects analysis of variance models. Mood (1946) proposed the same criterion for obtaining weighing designs. Kiefer and Wolfowitz (1959) later called this criterion D-optimality and extended its use to regression models in general. De la Garza (1954) showed that, for a polynomial of degree $p - 1$, there exists a design with p distinct points which has the same $\mathbf{X}'\mathbf{X}$ matrix as a design with more than p distinct points (Kiefer (1959) subsequently corrected De la Garza's proof). Hoel (1958) obtained the optimal allocation for a polynomial of degree $p - 1$ using both the min max criterion of Smith and the determinant criterion of Wald. He showed that the two criteria gave the same results, thus hinting at the equivalence theorem of Kiefer and Wolfowitz (1959), which we discuss in the next section.

Various other properties of the $\mathbf{X}'\mathbf{X}$ matrix were suggested as being an appropriate criterion for design. Elfving (1952) and Chernoff (1953) minimized the trace of $(\mathbf{X}'\mathbf{X})^{-1}$ to obtain regression designs. Ehrenfeld (1955) suggested that maximizing the minimum eigenvalue of $\mathbf{X}'\mathbf{X}$ be used as a design criterion.

3. Design Theory

Kiefer (1959, 1961a, 1961b, 1962 and 1970a) and Kiefer and Wolfowitz (1959, 1960) made a number of contributions to the theory of regression design. We discuss the two main results: the idea of design measure and their general equivalence theorem. A design can be viewed as a measure ξ on χ. Suppose we have an n-point design with n_i observations at the site $\mathbf{x}_i$ (note that $\Sigma n_i = n$). Let ξ be a probability measure on χ such that $\xi(\mathbf{x}_i) = 0$ if there are to be no observations at point $\mathbf{x}_i$, and such that $\xi(\mathbf{x}_i) = n_i/n$ if there are to be $n_i > 0$ observations at point $\mathbf{x}_i$. For a discrete n-point design, ξ takes on values which are multiples of $1/n$, and defines an *exact* design on χ. Removing the restriction that ξ be a multiple of $1/n$, we can extend this idea to a *design measure* which satisfies, in general:

$$\xi(\mathbf{x}) \geq 0, \mathbf{x} \varepsilon \chi$$

$$\int_\chi \xi(d\mathbf{x}) = 1. \tag{8}$$

With respect to model (2), we can define a matrix analogous to $\mathbf{X}'\mathbf{X}$ for design ξ. Let

$$m_{ij}(\xi) = \int_\chi f_i(\mathbf{x})f_j(\mathbf{x})\xi(d\mathbf{x}), \quad i, j = 1, 2, \cdots, p \tag{9}$$

and

$$\mathbf{M}(\xi) = [m_{ij}(\xi)].$$

Note that, for an exact design, $n\mathbf{M}(\xi) = \mathbf{X}'\mathbf{X}$. Similarly, a generalization of the variance function (6) is given by

$$d(\mathbf{x}, \xi) = \mathbf{f}'(\mathbf{x})\mathbf{M}^{-1}(\xi)\mathbf{f}(\mathbf{x}). \tag{10}$$

From this following general definitions of D-optimality and G-optimality.

Definition 1. ξ^* is D-optimal if and only if $\mathbf{M}(\xi^*)$ is nonsingular and

$$\max_\xi \det(\mathbf{M}(\xi)) = \det(\mathbf{M}(\xi^*)). \tag{11}$$

Definition 2. ξ^* is G-optimal if and only if

$$\min_\xi \max_{\mathbf{x}\varepsilon\chi} d(\mathbf{x}, \xi) = \max_{\mathbf{x}\varepsilon\chi} d(\mathbf{x}, \xi^*). \tag{12}$$

A sufficient condition for ξ^* to satisfy (12) is

$$\max_{\mathbf{x}\varepsilon\chi} d(\mathbf{x}, \xi^*) = p. \tag{13}$$

The equivalence of D- and G-optimality is established in the general equivalence theorem of Kiefer and Wolfowitz, which can be formally stated as follows:

Theorem 1. Conditions (11), (12), and (13) are equivalent, and the set of all ξ satisfying these conditions is convex.

The implication of this result is that we can use (13) to verify whether or not a specific design is D-optimal. That is, if (13) is satisfied, then the design is D-optimal. Note that D-optimality is essentially a parameter estimation criterion, whereas G-optimality is a response estimation criterion. The equivalence theorem says these two design criteria are identical *when the design is expressed as a measure on* χ.

As defined above, the criterion of D-optimality is applicable when all p parameters are to be estimated. However, in many cases one is interested in only $s < p$ parameters, the other parameters being nuisance parameters. Kiefer and Wolfowitz (1959) discussed this problem and Kiefer (1961a) gave definitions of D_s-optimality and G_s-optimality (i.e., optimality when interest is in $s < p$ parameters) which are analogous to the above definitions of D- and G-optimality. Kiefer (1961a, 1961b, and 1962) later extended the general equivalence theorem to this situation.

Karlin and Studden (1966a) gave various forms of the response model (1) for which the D-optimal designs were known. In addition, Karlin and Studden extended and simplified Kiefer's results for D_s- and G_s-optimality. In particular they discuss the problem of estimating $s < p$ parameters when $\mathbf{M}(\xi)$ is singular.

Atwood (1969) gave further results in the theory of D-optimality. In particular he gave more general definitions of D_s- and G_s-optimality, and he discussed the role of symmetry in optimal design.

Silvey and Titterington (1973) have given a geometric interpretation of optimal design. Using results from Strong Lagrangian Theory they established duality theorems which state the Kiefer–Wolfowitz equivalence theorem in terms of a design measure on the function space (the image of χ under the model). Moreover, they suggest that this duality may lead to efficient algorithms for constructing D-optimal designs, a point which we discuss below.

Sibson (1973) and (1974), following his contribution to the discussion of the paper of Wynn (1972), is developing linear programming algorithms for the explicit solution of D-optimality problems, based on these duality theorems. He has been able to avoid the unboundedness problems involved by imposing extra constraints on the total mass allocation to points, namely, that no point should have mass greater than $1/p$.

Kiefer (1974) has recently given further equivalence results between D-optimality and other design criteria. He also suggests construction methods based on these equivalence results.

For the sake of brevity we will hereafter discuss only the case of interest in all p parameters. Readers interested in the case where $s < p$ parameters are of interest should refer to the above papers (as well as recent results by Atwood (1973), Wynn (1970 and 1972), and Fedorov (1972)).

4. Design for Non-linear Models.

The form of model (1) is somewhat restrictive since it assumes that the response function is linear in the parameters. What can be done in the case of non-linear response function? Chernoff (1953) linearized the model using a first order Taylor series expansion about a preliminary value for $\boldsymbol{\beta}$, and he applied the maximum trace criterion to obtain locally optimum designs for this linearized model.

G. E. P. Box and Lucas (1959) also linearized models which are non-linear in the parameters. They applied the D-optimality criterion to obtain designs for the linearized models. Box and Lucas, using a geometric argument to obtain p-point designs for p-parameter models, showed that the locally D-optimal p-point design maximizes the volume of the simplex defined by the design points in the image (under the model) of the design region.

G. E. P. Box and Hunter (1965) showed that locally D-optimal designs for non-linear models maximize the posterior probability of the parameters at the least squares (maximum likelihood under normality) solution point. They extended this result to obtain sequential designs. Their design sequence selects, at each step, the point which maximizes (over the design region) the variance of the predicted response, based on a linearization of the non-linear model. Note that, at each step in the experimental sequence, the preliminary estimates of the parameters must be updated as data are obtained. That is, the posterior distribution on the parameters becomes the prior distribution for obtaining the next design point. This design sequence is essentially an application of Wynn's use of the Kiefer–Wolfowitz equivalence theorem to obtain designs. We will later discuss more general methods of obtaining D-optimal designs which are based on the same procedure as employed by Box and Hunter.

Atkinson and Hunter (1968) extend the Box–Lucas results to the case where the design is to have more than p points. Draper and Hunter (1967a and 1967b), discussed the problem of selecting prior distributions on the parameters to obtain designs for non-linear models.

M. J. Box (1968, 1970, 1971) has given a number of additional results for non-linear design, particularly on the subjects of replicated design points (see also Wynn (1972)), model inadequacy, and computing difficulties in obtaining D-optimal designs. Sredni (1970) partitioned the $\mathbf{X}'\mathbf{X}$ matrix into three components: (i) the determinant of $\mathbf{X}'\mathbf{X}$, which is proportional to the volume of a Highest Posterior Distribution (H.P.D.) region on $\boldsymbol{\beta}$; (ii) a function of the eigenvalues of $\mathbf{X}'\mathbf{X}$, which describes the conditioning of this H.P.D. region; and, (iii) a function of the eigenvectors of $\mathbf{X}'\mathbf{X}$, which describes the orientation of the H.P.D. region relative to the axes of the parameter space.

White (1973) recently extended the Kiefer–Wolfowitz equivalence theorem to design for non-linear models. White developed the information matrix (9) and the variance function (10) as functions of $\boldsymbol{\beta}$ (assumed values of the parameters) and proved an equivalence using these new functions.

5. Algorithms for Constructing D-Optimal Designs.

As we mentioned in section 1, the practical design problem consists of selecting an *exact* design, namely vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$, which define the experimental conditions for n experimental runs. In section 3 we reviewed the concept of using a measure ξ on the design region χ to define a *measure*, or *approximate* design.

Measure designs are of interest primarily because the D-optimal measure design provides the reference against which exact designs can be evaluated, and also because the points in an optimal exact design will often correspond to the points of support (points of positive measure) of the D-optimal measure design.

We first discuss algorithms for obtaining D-optimal designs. Kiefer and Wolfowitz (1959) suggested a game theory approach for constructing designs. However, this approach appears more useful for verifying the optimality of given designs. Kiefer (1961a and 1961b) gave examples of constructing D-optimal designs, but his methods did not lead to a general algorithm. Fedorov and his co-workers [Fedorov (1969a, 1969b, 1972), Fedorov and Dubova (1968)] appear to have been the first to develop a general algorithm for obtaining D-optimal designs.

Most of the original published results are in Russian. However, Fedorov's (1969b) book was translated by Studden and Klimko (1972), and our discussion of Fedorov's work is based on this translation.

Fedorov's algorithm for obtaining a D-optimal design allows the variance of the observation error (i.e., $v(\hat{y}(\mathbf{x}))$) to be a function of the point $\mathbf{x}$ at which the observation is made. We shall assume that this variance function is independent of $\mathbf{x}$ and $y(\mathbf{x})$, and present Fedorov's algorithm in this simpler form. The steps of the algorithm are as follows:

1. Let ξ_n be a nondegenerate (det $(\mathbf{M}(\xi_n)) > 0$) n-point design on χ, and let $i = n$.
2. Compute $\mathbf{M}(\xi_i)$ and $\mathbf{M}^{-1}(\xi_i)$.
3. Find $\mathbf{x}_i$ such that

$$\max_{\mathbf{x}} d(\mathbf{x}, \xi_i) = d(\mathbf{x}_i , \xi_i). \tag{14}$$

4. Let

$$\alpha_i = \frac{d(\mathbf{x}_i , \xi_i) - p}{p(d(\mathbf{x}_i , \xi_i) - 1)} \tag{15}$$

be the measure at point $\mathbf{x}_i$.

5. Let

$$\xi_{i+1} = (1 - \alpha_i)\xi_i + \alpha_i \epsilon_{\mathbf{x}_i} \tag{16}$$

be the new design measure (where $\epsilon_{\mathbf{x}_i}$ places measure one at point $\mathbf{x}_i$).

6. Repeat steps 2–5 until $d(\mathbf{x}_i , \xi_i)$ is sufficiently close to p.

In words, we can state these steps simply as follows:

1. Select any non-degenerate starting design.
2. Compute the dispersion matrix.
3. Find the point of maximum variance.
4. Add the point of maximum variance to the design, with measure proportional to its variance.
5. Update the design measure.

Fedorov also developed a numerical procedure for updating $\mathbf{M}(\xi_i)$ and $\mathbf{M}^{-1}(\xi_i)$ such that the inverse need be calculated only for $M(\xi_n)$. Fedorov proved that the sequence of designs defined by steps 1–6 converges (as $i \to \infty$) to the D-optimal design ξ^*.

The value of α_i given in (15) maximizes the increase in the value of the determinant at each step in the sequence. However, Fedorov (1972) indicated that any sequence of α_i satisfying

$$\sum_{i=1}^{\infty} \alpha_i = \infty, \qquad \lim_{i\to\infty} \alpha_i = 0 \tag{17}$$

will also guarantee convergence to the D-optimal design ξ^*.

Wynn (1970) proved the convergence of a similar algorithm, but with $\alpha_i = 1/(i + 1)$, that is, with all points in the design sequence having equal weight. (Pazman (1974) has given a different proof to Wynn's algorithm.) Obviously, such a choice for α_i does satisfy (17). Dykstra (1971), Hebble and Mitchell (1971), and Covey-Crump and Silvey (1970) suggested algorithms similar to Wynn's, but did not give convergence proofs.

Atwood (1973) has offered some improvements to Fedorov's algorithm. He made these suggestions:

1. It may be possible to achieve a greater increase in the value of the determinant by *removing* measure from a point already in the design ξ_i and distributing the measure removed among the remaining points in design ξ_i . If such an action would result in a greater increase in det $(\mathbf{M}(\xi_i))$ than would Fedorov's method of adding measure to the point satisfying (14), the point $\mathbf{x}_k$ from which measure is to be removed is given by

$$\min_{\{\mathbf{x}_j \mid \xi_i(\mathbf{x}_j) > 0\}} d(\mathbf{x}_j , \xi_i) = d(\mathbf{x}_k , \xi_i). \tag{18}$$

2. In the event that a number of different actions (for example, adding measure to a point of maximum variance or removing measure from a design point of minimum variance) would lead to identical increases in the value of det $(\mathbf{M}(\xi_i))$, choose the new design to be the equal-proportion convex combination of these potential designs. This procedure can greatly improve convergence when the D-optimal design is symmetric.
3. The maximum measure for any design point in ξ^* is $1/p$. If at any stage in the algorithm the measure at a design point is greater than $1/p$, allocate the excess measure to the other points in the design.

St. John (1973) has modified Atwood's first suggestion by distributing the measure removed only among the points of support for which $d(\mathbf{x}_i , \xi_i) > p$, and additionally, distributing the removed measure among *these* points in a manner proportional to $d(x_i , \xi_i) - p$. In this way the removed measure will be distributed according to "need". Atwood's second suggestion can also be improved through a better choice of α_i when ties occur (St. John (1973)).

These improvements by Atwood (1973) and St. John (1973) can be especially useful in eliminating undesirable points of support which were included in the initial design. In such cases, the rate of convergence of Fedorov's algorithm will be markedly improved.

Silvey and Titterington (1973) have recently outlined an algorithm for obtaining a D-optimal design measure on the dual space (i.e., on the space spanned by the response function specified in model (1)). Their method of construction differs from that developed by Fedorov in that the iteration scheme consists of double-loop procedure (i.e., one iterative procedure inside another iterative procedure). They are presently developing an efficient procedure for implementing this double-loop algorithm.

Wynn (1972) extended his algorithm for constructing D-optimal designs when one is interested in efficient estimates for $s < p$ parameters. His algorithm is based on the extended equivalence theorem given by Kiefer (1961a), Karlin and Studden (1966a, b) and Atwood (1969). Interested readers are referred to Wynn's paper, and to the spirited discussion that followed it. Further refinements of this algorithm were given by Wynn (1973).

Despite Fedorov's algorithm, D-optimality is not always an easy criterion to implement. Nalimov *et al* (1970) and M. J. Box and Draper (1971) point out that creating an exact design from the probability measure may require an extremely large number of experimental runs. It we restricted the design to a fixed number of experimental runs (that is, fixed n) it may be necessary to round off the measures in order to implement the D-optimal design. However, there may be other n-point designs with det $(\mathbf{X}'\mathbf{X})$ larger than the value of det $(\mathbf{X}'\mathbf{X})$ given by this rounded off design. Kiefer (1970a) suggested rounding-off procedures which yield n-point designs with det $(\mathbf{X}'\mathbf{X})$ close to the best possible value of det $(\mathbf{X}'\mathbf{X})$ for an n-point design. A drawback of his procedure is that the number of points in the design region with positive measure (that is, the number of points of support (or sites) of the design) may be larger than the value of n. In this case, the rounding-off procedure may eliminate points with small measure, thereby changing the nature of the design. For these reasons, it is desirable to search for algorithms which yield the n-point design with maximum value of det $(\mathbf{X}'\mathbf{X})$.

Let $\xi(n)$ denote an *exact* design with n points. Then, we denote the information matrix for this n point design as

$$n\mathbf{M}(\xi(n)) = \mathbf{X}_n{}'\mathbf{X}_n \,, \tag{19}$$

and the variance function (assuming $\sigma^2 = 1$ without loss of generality) of the predicted response as

$$v(\hat{y}(\mathbf{x})) = \mathbf{f}'(\mathbf{x})(\mathbf{X}_n{}'\mathbf{X}_n)^{-1}\mathbf{f}(\mathbf{x}). \tag{20}$$

Definition 3. $\xi^*(n)$ is *D_n-optimal* if and only if $\mathbf{X}_n{}^{*\prime}\mathbf{X}_n{}^*$ is nonsingular and the value of det$(\mathbf{X}_n{}^{*\prime}\mathbf{X}_n{}^*)$ is the maximum value over all possible designs $\xi(n)$.

Definition 4. $\xi^*(n)$ is *G_n-optimal* if and only if $\max_{\mathbf{x}}\ \mathbf{f}'(\mathbf{x})(\mathbf{X}_n{}^{*\prime}\mathbf{X}_n{}^*)^{-1}\mathbf{f}(\mathbf{x})$ is minimum over all possible designs $\xi(n)$.

Unfortunately, an equivalence theorem comparable to Theorem 1 is not possible for n-point designs. Kiefer (1965) presented a counter-example to the existence of such an equivalence theorem.

A number of authors obtained D_n-optimal designs for specific models using mathematical programming methods to optimize $|\mathbf{X}_n{}'\mathbf{X}_n|$. Designs using this approach were given by Box and Draper (1971), Atkinson (1969), Bloom, Pfaffenberger and Kochenberger (1972), Box (1966), Hartley and Ruud (1969), Mahoney (1970), and Newhardt and Bradley (1971). However, as general computing algorithms for constructing designs, these methods involve optimizing with respect to np variables and can give local solutions which are not optimal.

Since no equivalence theorem exists, the algorithms for obtaining D_n-optimal designs are somewhat different than Fedorov's D-optimal algorithm. There are two different algorithms for obtaining D_n-optimal designs. The first is due to Fedorov (1969b, 1972) and Van Schalkwyk (1971). The second is due to Wynn (1972) and Mitchell and Miller (1970).

Suppose we have an arbitrary n-point design, and we wish to find a D_n-optimal design. Fedorov suggested replacing one of the points in the design, say $\mathbf{x}_i$, with a point from the design space χ, say $\mathbf{x}$, such that the increase in the value of $|\mathbf{X}_n{}'\mathbf{X}_n|$ is maximum. Let $\xi_j(n)$ represent the exact n-point design measure after exchange j. Define

$$\Delta_j(\mathbf{x}_i\,, \mathbf{x}) = d_j(\mathbf{x}) - [d_j(\mathbf{x})d_j(\mathbf{x}_i) - d_j{}^2(\mathbf{x}, \mathbf{x}_i)] - d_j(\mathbf{x}_i), \tag{21}$$

where

$$d_j(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{M}^{-1}(\xi_j(n))\mathbf{f}(\mathbf{x}) \tag{22}$$

and

$$d_j(\mathbf{x}, \mathbf{x}_i) = \mathbf{f}'(\mathbf{x})\mathbf{M}^{-1}(\xi_j(n))\mathbf{f}(\mathbf{x}_i). \tag{23}$$

If $\Delta_j(\mathbf{x}_i\,, \mathbf{x}) > 0$, then, when point $\mathbf{x}_i$ is replaced by point $\mathbf{x}$,

$$|M(\xi_{j+1}(n))| = |\mathbf{M}(\xi_j(n))|\,[1 + \Delta_j(\mathbf{x}_i\,, \mathbf{x})]. \tag{24}$$

The algorithm is as follows:

1. Begin with an arbitrary n-point design $\xi_0(n)$.
2. Calculate $\mathbf{M}(\xi_j(n))$ and $\mathbf{M}^{-1}(\xi_j(n))$.
3. Select $\mathbf{x}_i$ and $\mathbf{x}$ which satisfy
$$\max_{\{\mathbf{x}_i|\xi_j(\mathbf{x}_i)>0\}}\ \max_{\mathbf{x}\in\chi} \Delta_j(\mathbf{x}_i\,, \mathbf{x}) \tag{25}$$
and replace $\mathbf{x}_i$ by $\mathbf{x}$.
4. Repeat steps 2 and 3 until $\Delta_j(\mathbf{x}_i\,, \mathbf{x})$ is sufficiently close to zero.

Fedorov also presented a numerical procedure for updating $\mathbf{M}(\xi_j(n))$ and $\mathbf{M}^{-1}(\xi_j(n))$. Van Schalkwyk (1971) presented a similar algorithm, except he assumed that the point to be removed from the design is the point $\mathbf{x}_i$ with minimum variance (i.e., the point $\mathbf{x}_i$ which minimizes (22)). This procedure does not always produce the maximum increase in the value of the determinant.

The other exchange algorithm, due independently to Wynn (1972) and to Mitchell and Miller (1970), does not attempt to maximize the increase in the value of the determinant at each step, but rather emulates the Wynn algorithm for finding a D-optimal design. The algorithm is as follows:

1. Begin with an arbitrary n-point design $\xi_0(n)$.
2. Find $\mathbf{x}_{n+1}$ such that

$$d(\mathbf{x}_{n+1}, \xi_i(n)) = \max_{\mathbf{x}\epsilon X} d(\mathbf{x}, \xi_i(n)) \tag{26}$$

and add $\mathbf{x}_{n+1}$ to the n-point design.

3. Find $\mathbf{x}_i{}^*$ such that

$$d(\mathbf{x}_i{}^*, \xi_i(n+1)) = \min_{1 \le i \le n+1} d(\mathbf{x}_i, \xi_i(n+1)) \tag{27}$$

and remove $x_i{}^*$ from the $(n+1)$-point design.

4. Repeat steps 2 and 3 until no increase in the value of $|\mathbf{M}(\xi_i(n))|$ is obtained by an exchange.

Step 3 is due entirely to Mitchell and Miller, and is a simplification of Wynn's procedure for dropping a point from the augmented design.

Both of the above exchange algorithms converge (since both are monotonic in the value of $|\mathbf{M}(\xi_i(n))|$ and both are bounded above by $|\mathbf{M}(\xi^*)|$), but neither is guaranteed to converge to the D_n-optimal design. As far as we know, there has been no detailed comparison of these two exchange algorithms. If the cost speed of a point exchange is assumed to be the same for both algorithms (which may not be the case), then Fedorov's algorithm has a faster convergence rate (by the nature of the algorithm), but this does not guarantee convergence to a better design. We also note that, given a solution design from Fedorov's exchange algorithm, the maximum-variance, minimum-variance exchange algorithm cannot improve *that particular design* (this is obvious from the function $\Delta_i(\mathbf{x}_i, \mathbf{x})$). However, given an equivalent arbitrary starting design $\xi_0(n)$, we do not know which algorithm gives a better design, or which is more efficient if both give the same result. Another serious drawback is that there is no way of verifying whether or not the solution design given by either algorithm is a D_n-optimal design.

Mitchell (1974a) has recently proposed a variation on this exchange algorithm. In essence, Mitchell's new algorithm allows the value of n, the number of design points, to increase or decrease in order to allow a better search of the design region. However, constraints are placed on the amount of change in the value of n, and the algorithm forces a return to an n point design.

None of these algorithms for obtaining exact designs are guaranteed to give a D_n-optimal design. As we have previously mentioned, there is no equivalence theorem for exact designs, and therefore the value of max $d(\mathbf{x}, \xi(n))$ cannot be used to prove that $\xi(n)$ is D_n-optimal. Atwood (1969) and Wynn (1970) obtained upper and lower bounds on the efficiency of exact designs, based on the value of $|\mathbf{M}(\xi^*)|$, (ξ^* the D-optimal design) and the value of $\max_{\mathbf{x}} d(\mathbf{x}, \xi(n))$. Atwood, in the discussion of Wynn's (1972) paper, gave a different form of upper bound on the efficiency of a D_n-optimal design. It appears that these bounds are useful for determining that an exact design is *not* D_n-optimal, but not for assuring that an exact design *is* D_n-optimal. At present it appears that guaranteeing D_n-optimality can only be done by an exhaustive search of all possible designs.

The criterion of D-optimality has been applied to obtain (or justify) designs for analysis of variance situations, as originally done by Wald (1943), and to obtain weighing designs, a problem discussed by Mood (1946). Hill, Hunter and Wichern (1968) combined the criterion of D-optimality with a model discrimination criterion. Stigler (1971) extended this criterion to allow for determining lack of fit. Viort (1972a and 1972b) developed a procedure for applying D-optimality to obtain designs for dynamic models.

This paper is a condensed, updated version of St. John (1971) and Draper and St. John (1974). In particular, St. John (1971) gave an extensive review of D-optimality for weighing problems and for analysis of variance problems. The authors wish to thank the referees for several useful additions to this paper.

6. Acknowledgments

N. R. Draper was partially sponsored at the Mathematics Research Center, University of Wisconsin, by the United States Army under Contract No. DA-31-124-ARO-D-462, and partially sponsored at the Statistics Department, University of Wisconsin, by the United States Navy through the Office of Naval Research under Contract No. N-00014-67A-0128-0017.

R. C. St. John was partially sponsored at the University of Wisconsin by the Madison Academic Computer Center and the Wisconsin Alumni Research Foundation, and partially sponsored at Bowling Green State University by the Faculty Research Committee.

Bibliography

[1] Andrukovich, P. F., Golikova, T. I., and Kostina, C. G. (1969). Nearly D-optimal second-order designs on a hypercube. (In Russian). *New Ideas in Experimental Design.* Edited by V. V. Nalimov, Izd-vo "*Nauka*", Moscow, USSR.

[2] Atkinson, A. C. (1969). Constrained maximization and the design of experiments. *Technometrics, 11*, 616–618.

[3] Atkinson, A. C. (1972). Planning experiments to determine inadequate regression models. *Biometrika, 59*, 275–293.

[4] Atkinson, A. C. (1973). Multifactor second-order designs for cuboidal regions. *Biometrika, 60*, 15–19.

[5] Atkinson, A. C., and Hunter, W. G. (1968). The design of experiments for parameter extimation. *Technometrics, 10*, 271–289.

[6] Atwood, C. L. (1969). Optimal and efficient design of experiments. *Ann. Math. Statist., 40*, 1570–1602.

[7] Atwood, C. L. (1973). Sequences converging to D-optimal designs of experiments. *The Annals of Statist., 1*, 342–352.

[8] Beggs, William J. (1966). The Kiefer approach to experimental design. *Master's Essay, Department of Statistics, University of Wisconsin*, Madison, Wisconsin.

[9] Behnken, D. W. (1964). Estimation of copolymer reactivity rates: an example of nonlinear estimation. *J. of Polymer Science, Part A, 2*, 645–668.

[10] Bloom, T., Pfaffenberger, R., and Kochenberger, G. (1972). The efficient construction of optimal response surface designs based upon the $|\mathbf{X}'\mathbf{X}|$ criterion. *Presented at the 1972 American Statistical Assn. Meeting*, Montreal, Canada, August 14–18.

[11] Box, G. E. P. (1957). Use of statistical methods in the elucidation of basic mechanisms. *Bull. Inter. Statist. Inst., Part 3, 36*, 215–225.

[12] Box, G. E. P. and Hill, W. J. (1967). Discrimination among mechanistic models. *Technometrics, 9*, 57–71.

[13] Box, G. E. P., and Hunter, W. G. (1965). Sequential design of experiments for nonlinear models. *IBM Scientific Computing Symposium in Statist.*, 113–137.

[14] Box, G. E. P. and Lucas, H. L. (1959). Design of experiments in non-linear situations. *Biometrika, 46*, 77–90.

[15] Box, M. J. (1966). A comparison of several current optimization methods, and the use of transformations in constrained problems. *Computer Journal, 9*, 67–77.

[16] Box, M. J. (1968a). The use of designed experiments in nonlinear model building. *The Future of Statist.*, 241–268.

[17] Box, M. J. (1968b). The occurrence of replications in optimal designs of experiments to estimate parameters in nonlinear models. *J. Roy. Statist. Soc., B30*, 290–302.

[18] Box, M. J. (1969). The design of experiments and nonlinear model-building under a variety of error structures. *Ph.D. Thesis*, University of London, London, England.

[19] Box, M. J. (1970). Some experiences with a nonlinear experimental design criterion. *Technometrics, 12*, 569–589.

[20] Box, M. J. (1971a). Simplified experimental design. *Technometrics, 13*, 19–31.

[21] Box, M. J. (1971b). An experimental design criterion for precise estimation of a subset of the parameters in a nonlinear model. *Biometrika, 58*, 149–153.

[22] Box, M. J. and Draper, N. R. (1971). Factorial designs, the $|\mathbf{X}'\mathbf{X}|$ criterion, and some related matters. *Technometrics, 13*, 731–742.

[23] Brodsky, V. Z., and Golikova, T. I. (1971). D-optimal designs for a special case of linear regression. *Preprint No. 14 LSM, Izd-vo Moscow State University*, Moscow, USSR.

[24] Chernoff, H. (1953). Locally optimum designs for estimating parameters. *Ann. Math. Statist., 24*, 586–602.

[25] Clark, V. (1965). Choice of levels in polynomial regression with one or two variables. *Technometrics, 7*, 325–333.

[26] Covey-Crump, P. A. K., and Silvey, S. D. (1970). Optimal regression designs with previous observations. *Biometrika, 57*, 551–566.

[27] De La Garza, A. (1954). Spacing of information in polynomial regression. *Ann. Math. Statist., 25*, 123–130.

[28] Draper, N. R. and Hunter, W. G. (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika, 53*, 525–533.

[29] Draper, N. R. and Hunter, W. G. (1967a). The use of prior distributions in the design of experiments for parameter estimation in nonlinear situations. *Biometrika, 54*, 147–153.

[30] Draper, N. R. and Hunter, W. G. (1967b). The use of prior distributions in the design of experiments for parameter extimation in nonlinear situations—multiresponse case. *Biometrika, 54*, 662–665.

[31] Draper, N. R. and St. John, R. C. (1974). Models and designs for experiments with mixtures: I, Background material. *Mathematics Research Center Technical Summary Report No. 1435*, University of Wisconsin.

[32] Dubova, I. S., Ptskialadze, T. V., and Fedorov, V. V. (1970). Tables of D-optimum designs I (discrete D-optimum designs). *Preprint No. 16, Izd-vo Moscow State University*, Moscow, USSR.

[33] Dykstra, Otto, Jr. (1971). The augmentation of experimental data to maximize $|\mathbf{X}'\mathbf{X}|$. *Technometrics, 13*, 682–688.

[34] Ehrenfeld, S. (1955). On the efficiencies of experimental designs. *Ann. Math. Statist., 26*, 247–255.

[35] Ehrenfeld, S. (1956). Complete class theorems in experimental design. *Proceedings of the Third Berkeley Symposium on Mathematical Statist. and Probability, 1*, 57–68.

[36] Elfving, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist., 23*, 255–262.

[37] Elfving, G., (1959). Design of linear experiments. *Cramer Festschrift 58*, Wiley, New York.

[38] Farrell, R. H., Kiefer, J., and Walbran, A. (1967). Optimum multivariate designs. *Proceedings of the Fifth Berkeley Symposium of Mathematical Statist. and Probability, 1*, 113–138.

[39] Fedorov, V. V. (1968). Design of experiments in the case of simultaneous measurements of several response surfaces. *Preprint No. 3 LSM, Izd-vo Moscow State University*, Moscow, USSR.

[40] Fedorov, V. V. (1969a). Properties and methods for construction of point optimal designs in regression experiments. *Preprint No. 5 LSM, Izd-vo Moscow State University*, Moscow, USSR.

[41] Fedorov, V. V. (1969b). Theory of optimal experiments. *Preprint No. 7 LSM, Izd-vo Moscow State University*, Moscow, USSR.

[42] Fedorov, V. V. (1969c). Sequential methods for design of experiments in the study of the mechanism of a phenomenon. *New Ideas in Experimental Design.* Edited by V. V. Nalimov, Izd-vo "*Nauka*", Moscow, USSR.

[43] Fedorov, V. V. (1971a). The design of experiments in the multiresponse case. *Theory of Probability and its Applications, 16* (No. 2), 323–332.

[45] Fedorov, V. V. (1971b). Design of experiments when the response surface is nonlinear in the parameters. *Preprint No. 21, Izd-vo Moscow State University*, Moscow, USSR.

[46] FEDOROV, V. V. (1972). *Theory of Optimal Experiments.* Translated and edited by W. J. Studden, and E. M. Klimko, Academic Press, New York.
[47] FEDOROV, V. V., and DUBOVA, I. S. (1968). Methods for constructing optimal designs in regression experiments. *Preprint No. 4 LSM, Izd-vo Moscow State University*, Moscow, USSR.
[48] FEDOROV, V. V., and DUBOVA, I. S. (1972). Tables of optimum designs II (saturated D-optimal designs on a cube). *Issue No. 40, Izd-vo Moscow State University*, Moscow, USSR.
[49] FEDOROV, V. V., and MALYUTOV, M. B. (1972). Optimal designs in regression problems. *Mathematische Operationsfurschung und Statistik, 4*, 237–324.
[50] GAYLOR, D. W., and MERRILL, J. A. (1968). Augmenting existing data in multiple regression. *Technometrics, 10*, 73–81.
[51] GOLIKOVA, T. I., and MIKESHINA, N. G. (1969). Properties of D-optimal designs and the structure of quasi-D-optimal designs. *Information Materials.* Edited by S. I. Samoilenko, 18–26.
[52] GOLIKOVA, T. I., MIKESHINA, N. G., NALIMOV, V. V., and VESELAYA, G. N. (1967). Construction of nearly D-optimal second order designs on a cube. (In Russian). *Zavodskaya Laboratoriya, XXXIII*, 847–859.
[53] GUEST, P. G. (1958). The spacing of observations in polynomial regression. *Ann. Math. Statist., 29*, 294–299.
[54] GUEST, P. G. (1961). *Numerical Methods of Curve Fitting.* Cambridge University Press.
[55] HARTLEY, H. O. and RUUD, P. G. (1969). Computer optimization of second order response surface designs. *Statistical Computing.* Academic Press, New York, 441–462.
[56] HEBBLE, T. L., and MITCHELL, T. J. (1971). Repairing response surface designs. *Technometrics, 14*, 767–779.
[57] HILL, W. J., HUNTER, W. G., and WICHERN, D. W. (1968). A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics, 10*, 145–160.
[58] HILL, W. J., and HUNTER, W. G. (1974). Design of experiments for subsets of parameters. *Technometrics, 16*. 425–434.
[59] HOEL, P. G. (1958). Efficiency problems in polynomial estimation. *Ann. Math. Statist., 29*, 1134–1145.
[60] HOEL, P. G. (1961). Some properties of optimal spacing in polynomial estimation. *Ann. Inst. Stat. Math., 13*, 1–8.
[61] KARLIN, S., and STUDDEN, W. J. (1966a). Optimal experimental designs. *Ann. Math. Statist., 37*, 783–815.
[62] KARLIN, S., and STUDDEN, W. J. (1966b). *Techebycheff Systems: With Applications in Analysis and Statist.*, New York: Wiley Interscience.
[63] KIEFER, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *Ann. Math. Statist., 20*, 675–699.
[64] KIEFER, J. (1959). Optimum experimental designs. *J. Roy. Statist., Soc., B21*, 272–319.
[65] KIEFER, J. (1961a). Optimum designs in regression problems, II. *Ann. Math. Statist., 32*, 298–325.
[66] KIEFER, J. (1961b). Optimum experimental designs V, with applications to rotatable designs. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statist. and Probability, 1*, 381–405.
[67] KIEFER, J. (1962). Two more criteria equivalent to D-optimality of designs. *Ann. Math. Statist., 33*, 792–796.
[68] KIEFER, J. (1970a). The role of symmetry and approximation in exact design optimality. *Statistical Decision Theory and Related Topics*, 109–118, Academic Press, New York.
[69] KIEFER, J. (1970b). Optimum Experimental Designs. *Proc. Intl. Congress Math.*, Nice, *3*, 249–254, Gauthier-Villars, Paris.
[70] KIEFER, J. (1972). Exact optimality theory for designs. *IMS Bull., 1*, 182.
[71] KIEFER, J. (1973). Panel discussion on optimality criteria for design. *Proceedings of March 1973 Fort Collins Conference.*
[72] KIEFER, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Math. Statist., 2*, 849–879.
[73] KIEFER, J., and WOLFOWITZ, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist., 30*, 271–294.
[74] KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Canad. J. Math., 12*, 363–366.
[75] KONO, K. (1962). Optimum designs for quadratic regression on k-cube. *Mem. Fac. Sci., Kyushu University of Science A*, 114–122.
[76] LINDLEY, D. V. (1956). On a measure of the information provided by an experiment. *Ann. Math. Statist., 27*, 968–1005.
[77] MALYUTOV, M. B., and FEDOROV, V. V. (1971). On the design of experiments taking into account inadequacy of the assumed model. *Preprint No. 18, Izd-vo Moscow State University*, Moscow, USSR.
[78] MINNICH, G. M. (1972). Some considerations in the choice of experimental designs for dynamic models. *Ph.D. Thesis*, Department of Statist., University of Wisconsin, Madison, Wisconsin.
[79] MITCHELL, T. J. (1974a). An algorithm for the construction of D-optimal experimental designs. *Technometrics, 16*, 203–211.
[80] MITCHELL, T. J. (1974b). Computer construction of D-optimal first-order designs. *Technometrics, 16*, 211–220.
[81] MITCHELL, T. J., and MILLER, F. L., JR. (1970). Use of design repair to construct designs for special linear models. *Math. Div. Ann. Progr. Rept. (ORNL-4661), Oak Ridge National Laboratory*, Oak Ridge, Tennessee, 130–131.
[82] MAHONEY, F. (1970). The design of experiments for regression analysis by computer search procedures. *Ph.D. Thesis*, Rensselaer Polytechnic Institute.
[83] MOOD, A. M. (1946). On Hotelling's weighing problem. *Ann. Math. Statist., 17*, 432–446.
[84] NALIMOV, V. V., editor. (1969). *New Ideas in Experimental Design—A Collection of Articles.* Izd-vo "*Nauka*", Moscow, USSR.
[85] NALIMOV, V. V., and GOLIKOVA, T. I. (1971). Logical foundations of the planning of experiments. *Preprint No. 20, Izd-vo Moscow State University*, Moscow, USSR.
[86] NALIMOV, V. V., GOLIKA, T. I., and MIKESHINA, N. G. (1970). On practical use of the concept of D-optimality *Technometrics, 12*, 799–812.
[87] NANDI, H. K. (1950). On the efficiency of experimental designs. *Calcutta Stat. Assn. Bull., 3*, 167–171.
[88] NEUHARDT, J. B. and BRADLEY, H. E. (1971). Selection of multi-factor experimental arrangements with resource constraints. *J. Amer. Statist. Assoc., 66*, 618–621.
[89] PAZMAN, A. (1974). A convergence theorem in the theory of D-optimum experimental designs. *Ann. Statist., 2*, 216–218.
[90] PESOTCHINSKY, L. L. (1972). D-optimum and quasi-D-

optimum design for quadratic regression on a cube and a ball. *Preprint No. 42 LSM, Izd-vo Moscow State University*, Moscow, USSR.

[91] St. John, R. C. (1971). D-optimality, a review. *Master's Essay, Department of Statist., University of Wisconsin*, Madison, Wisconsin.

[92] St. John, R. C. (1973). Models and designs for experiments with mixtures. *Ph.D. Thesis*, Department of Statist., University of Wisconsin, Madison, Wisconsin.

[93] Shaw, D. E. (1973). The effects of nonlinearity in regression models. *Ph.D. Thesis*, Department of Statist., University of Wisconsin, Madison, Wisconsin.

[94] Sibson, R. (1973). DA-optimality and duality. *Proc. European Meeting Statisticians, Budapest*, 1972. (In press).

[95] Sibson, R. (1974). Cutting plane algorithms for D-optimal designs. (To appear).

[96] Silvey, S. D., and Titterington, D. M. (1973). A geometric approach to optimal design theory. *Biometrika, 60*, 21–32.

[97] Skwish, J. (1969). On the robustness of optimal designs in polynomial regression. (Abstract). *Ann. Math. Statist., 40*, 1161.

[98] Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika, 12*, 1–85.

[99] Sredni, J. (1970). Problems of design, estimation, and lack of fit in model building. *Ph.D. Thesis*, Department of Statist., University of Wisconsin, Madison, Wisconsin.

[100] Srivastava, J. N., and McDonald, L. (1971). On the costwise optimality of certain hierarchical and standard multiresponse models under the determinant criterion. *J. of Multivariate Analysis, 1*, 118–128.

[101] Stigler, S. M. (1971). Optimal experimental design for polynomial regression. *J. Amer. Statist. Assoc., 66*, 311–318.

[102] Stone, M. (1959). Application of a measure of information to the design and comparison of regression experiments. *Ann. Math. Statist., 30*, 55–70.

[103] Turing, A. M. (1948). Rounding-off errors in matrix processes. *Quart. J. Mech. Applied Math., 1*, 287–308.

[104] Uranisi, J. (1964). Optimum design for the special cubic regression on the q-simplex, *Mathematical reports, General Education Department, Kyushu University, 1*, 7–12.

[105] Van Schalkwyk, D. J. (1971). On the design of mixture experiments. *Ph.D. Thesis*, University of London, London, England.

[106] Viort, B. (1972a). D-optimal designs for dynamic models, Part I. Theory. *Technical Report No. 314, Department of Statist., University of Wisconsin*, Madison, Wisconsin.

[107] Viort, B. (1972b). D-optimal designs for dynamic models, Part II. Applications. *Technical Report No. 316, Department of Statist., University of Wisconsin*, Madison, Wisconsin.

[108] Wald, A. (1943). On the efficient design of statistical investigations. *Ann. Math. Statist., 14*, 134–140.

[109] Wheeler, R. E. (1973). Efficient experimental design. *Presented at the Gordon Research Conference on Statist. in Chemistry and Chemical Engineering*, New Hampton, New Hampshire, July 9–13, 1973.

[110] White, L. V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika, 60*, 345–348.

[111] Whittle, P. (1973). Some general points in the theory of optimum experimental design. *J. Roy. Statist. Soc., B35*, 123–130.

[112] Wynn, H. P. (1970). The sequential generation of D-optimum experimental designs. *Ann. Math. Statist., 41*, 1655–1664.

[113] Wynn, H. P. (1972). Results in the theory and construction of D-optimum experimental designs. (With discussion). *J. Roy. Statist. Soc., B34*, 133–147 and 170–186.

[114] Wynn, H. P. (1973). Simple conditions for optimum experimental design algorithms. *Presented at the Symposium on Linear Models*, Colorado Springs, Colorado, March 23, 1973.

This is an updated version of a bibiography prepared by R. C. St. John (1971).