NOT FOR QUOTATION
WITHOUT PERMISSION
OF THE AUTHOR

DUALITY OF OPTIMAL DESIGNS FOR MODEL DISCRIMINATING AND PARAMETER ESTIMATION PROBLEMS

V. Fedorov
V. Khaborov

September 1984
WP-84-77

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

ACKNOWLEDGMENT

I am grateful to Lucy Tomsits for editing and typing this paper.

ABSTRACT

Attempts to find out relations between different criteria of optimality have a long history descending to the fifties (Kiefer, 1958; Stone, 1958). This paper mainly deals with analysis of relations between most widespread criteria used in estimation problems and some criteria for discriminating experiments which belong to T-criteria family (Atkinson and Fedorov, 1974).

CONTENTS

DUALITY OF OPTIMAL DESIGNS FOR MODEL DISCRIMINATING AND PARAMETER ESTIMATION PROBLEMS

V. Fedorov and V. Khaborov

1. INTRODUCTION

The main object of this paper is the optimal designs for experiments which can be described by linear regression models:

$$y_{ij} = \vartheta_t^T f(x_i) + \varepsilon_{ij} = \eta(x, \vartheta_t) + \varepsilon_{ij}. \tag{1}$$

Vector x_i describes the conditions which the i-th set of observations are made under ($j = \overline{1, n_i}$, $i = \overline{1, n}$, $\sum_{i=1}^{n} n_i = N$) and the value of its components can be chosen (controlled) by an experimenter: $x_i \in X \subset R^k$ where X is a compact. Components of the vector $\vartheta \in R^m$ are unknown parameters and the subscript "t" points out their true values. Components of vector $f(x_i)$ are given basic functions, which are continuous on the compact X.

The errors ε_{ij} are assumed to be random, identically independently distributed with zero mean and finite variance which, without losing generality, can be chosen equal to 1. These assumptions on errors are sufficient in the case when the estimation problem is under consideration, but for discriminating experiments the normality of their distribution will be assumed in what follows below. Some more general situations can be considered similarly to (Fedorov, 1980; Denisov, Fedorov, and Khaborov, 1981).

The set of values

$$\xi_N = \{p_i, x_i\}, \quad p_i = n_i / N, \sum_{i=1}^{n} p_i = 1$$

is a design of an experiment. Fractions p_i can be considered as measures prescribed to points x_i and variations of these measures must be proportional to N^{-1} in experimental practice.

To deal with discrete measures in optimization problems, one should apply to very complicated mathematical technique. The problem essentially can be simplified if the discreteness is neglected and any probabilistic measure $\xi(dx)$ on X is considered as some experimental design. Corresponding designs are called approximate or continuous. In this paper, they will be referred as "designs".

For a comparatively large N, it is not a problem to construct an appropriate discrete approximation of any measure $\xi(dx)$ especially if one takes into account that almost for all widely used criteria of optimality, there exist optimal designs with finite number of supporting points (points of concentration of measure $\xi(dx)$; see for instance, Fedorov, 1972; and section 3 of this paper). Formally, the construction of optimal designs can be considered as an optimization problem in the space of probabilistic measures:

$$\xi^* = \text{Arg} \inf_{\xi} \Psi(\xi)$$

where the optimality criterion Ψ is defined by objectives of an experimenter and is usually a convex function of ξ.

In the parameter estimation problem, the dependence of Ψ from ξ can be expressed through elements of Fisher's information matrix:

$$\Psi(\xi) = \Psi[M(\xi)], \tag{2}$$

where

$$M(\xi) = \int_X f(x) f^T(x) \xi(dx)$$

This matrix in the regular case is inverse to the normalized dispersion (variance-covariance) matrix:

$$D(\xi) = M^{-1}(\xi), \;\; D(\hat{\vartheta}) = N^{-1} D(\xi),$$

$\hat{\vartheta}$ is the (best linear unbiased) estimator of ϑ. In the case of discriminating (or more accurately, model testing) experiments, the structure of Ψ is slightly more complicated. For instance, when there are two rival models:

$$\eta(x, \vartheta) = \begin{cases} \eta_1(x_1 \vartheta_1) \\ \eta_2(x_1 \vartheta_2) \end{cases}$$

the design problem can be described by the following optimization problem (Atkinson and Fedorov, 1975).

$$\xi^* = \text{Arg} \sup_{\xi} \inf_{\substack{\vartheta_1 \in \Omega_1 \\ \vartheta_2 \in \Omega_2}} \int_X [\eta_1(x, \vartheta_1) - \eta_2(x, \vartheta_2)]^2 \xi(dx). \tag{3}$$

Very often some nondegenerate regression function $\eta(\vartheta,x)$ is compared with zero hypothesis and in this case, (3) transforms in the more simple problem

$$\xi^* = \text{Arg} \sup_{\xi} \Phi(\xi),$$

where

$$\Phi(\xi) = -\Psi(\xi) = \inf_{\vartheta\in\Omega} \int_X \eta^2(x,\vartheta_1)\xi(dx), \tag{4}$$

which will be mainly considered in the following sections.

2. EQUIVALENCY OF DIFFERENT DESIGN CRITERIA

In this section, the equivalency between some criteria corresponding to model testing experiments and experiments oriented to parameter estimation will be analyzed. The majority of results are based on the well-known results from the theory of extrema of quadratic forms.

1. Let us start with the most evident and simple case when an experimenter is interested in some linear combination $c^T\vartheta$ of unknown parameters. For interpolation or extrapolation problem $c = f(x_0)$, where x_0 is the point of interest. Then if he wants to estimate $c^T\vartheta$ the criterion

$$\Psi(\xi) = c^T M^-(\xi)c, \tag{5}$$

where "−" means pseudo-inverse matrix, can be used. If the significance of $c^T\vartheta_t$ is tested, then

$$\Phi(\xi) = \inf_{(c^T\vartheta)^2\geq 1} \int_X \eta^2(x,\vartheta)\xi(dx). \tag{6}$$

It is easy to check out that in (6), instead of 1, any positive constant can be taken without influence of an optimal design if $\eta(x,\vartheta)$ depends linearly of ϑ. The similar fact will take place for the criteria considered below and it will be used without any comments.

It is natural to suggest that

$$\{\xi:\ c^T M^-(\xi)c < \infty\} \neq \phi \tag{7}$$

for any pseudo-inverse matrix. Of course, if (7) takes place, then for any optimal design ξ^* (very often the solution of (2) is not unique), $c^T M^-(\xi^*)c < \infty$, or in another words, we assume that $c^T\vartheta$ is estimable in the experiments defined by ξ^*. It will be useful to note that the necessary and sufficient condition of the estimability of $c^T\vartheta$ is the following equality (see for instance, Rao, 1973):

$$c^T(I-M^-(\xi)M(\xi)) = 0 \tag{8}$$

for any pseudo-inverse matrix. The designs satisfying to (8) will be called *regular*.

Consider now criterion (6) more detailly. Due to (2), one has

$$\int_X \eta^2(x,\vartheta)\xi(dx) = \vartheta^T M(\xi)\vartheta$$

and (6) transforms to

$$\Phi(\xi) = \inf_{(c^T\vartheta)^2 \geq 1} \vartheta^T M(\xi)\vartheta \tag{9}$$

It is obvious that all optimal designs ξ^* for (9) coincide with the optimal designs for of the more simple problem

$$\inf_{c^T\vartheta=1} \vartheta^T M(\xi)\vartheta \tag{10}$$

Taking into account the condition (8) and using the standard Lagrangian technique, one can get

$$\inf_{c^T\vartheta=1} \vartheta^T M(\xi)\vartheta = c^T M^-(\xi)c, \tag{11}$$

with

$$\vartheta^* = M^-(\xi)c$$

From the last equation, it immediately follows that regular optimal designs (in other words, the solutions of (2)) are the same both for criteria (5) and (6). In this sense these criteria are equivalent (compare with Kiefer's equivalency theorem, Fedorov, 1972). The equivalency property is useful in several aspects:

— It helps an experimenter, ensuring him that he can solve two statistical problems simultaneously;

— In numerical construction of optimal designs, it gives possibility to choose the most convenient algorithm, because dependently on $f(x)$,X and c either optimization problem (5) or (6) can be more simple;

— In theoretical analysis of optimal designs, sometimes it is convenient to relay between (5) and (6).

2. If in the model testing case, there is some prior information on the parameters ϑ_t described by prior distribution function, $F_0(d\vartheta)$, then it is reasonable to use the mean of the noncentrality parameter as a criterion of optimality:

$$\Phi_0(\xi) = \int_\Omega \int_X \eta^2(x,\vartheta)\xi(dx)F_0(d\vartheta). \tag{12}$$

If the distribution $F_0(d\vartheta)$ has a dispersion matrix equals to D_0 then (12) can be transformed to

$$\Phi_0(\xi) = \text{tr} \;\; D_0 M(\xi) \tag{13}$$

In practice, the knowledge of D_0 is problematic and one can relax this demand and assume that only the determinant value of a dispersion matrix are given to be greater than $d > 0$. In this case, the criterion

$$\Phi(\xi) = \inf_{|D_0| \geq d} \;\; \text{tr} \; D_0 M(\xi) \tag{14}$$

can be the point of an interest. If the matrix $M(\xi)$ is nonsingular, then the infrenum in (13) can be found easily (compare with Fedorov, 1981)

$$\Phi(\xi) = md^{\frac{1}{m}} |M(\xi)|^{\frac{1}{m}} \tag{15}$$

Evidently the maximization of (15) is equivalent to the maximization of $|M(\xi)|$, or in other words, criterion (14) is equivalent to D-criterion:

$$\Psi(\xi) = |M(\xi)|^{-1} \tag{16}$$

This criteria is one of the most widely used criteria in the estimation problem. Some properties of D-optimal designs connected with model testing were discussed by Kiefer (1958) and Stone (1958). The above result gives additional explanation of the relation between D-criterion and the model testing problem. In the next section even more startling example illuminating this relation will be considered.

3. Let us start with a very natural criteria for model testing problem:

$$\Phi(\xi) = \inf_{\sup_{x \in V} \varphi^2(x,\vartheta) \geq 1} \;\; \int_X \eta^2(x,\vartheta)\xi(dx), \tag{17}$$

which in the linear case takes the form:

$$\Phi(\xi) = \inf_{\sup_{x \in V} (\vartheta^T q(x))^2 \geq 1} \;\; \vartheta^T M(\xi)\vartheta.$$

It is not difficult to check the following chain of equalities:

$$\inf_{\sup_{x\in U}(\vartheta^T q(x))^2\geq 1} \vartheta^T M(\xi)\vartheta = \sup_{x\in U} \inf_{(\vartheta^T q(x))^2\geq 1} \vartheta^T M(\xi)\vartheta$$

$$= \sup_{x\in U} q^T(x)M^{-1}(\xi)q(x),$$

where, of course, $M^{-1}(\xi)$ exists for any design with $\Phi(\xi) > 0$ or $\Psi(\xi) = |M^{-1}(\xi)| < \infty$.

The first equality follows from the inclusion:

$$\{\vartheta: \sup_{x\in U} (\vartheta^T q(x))^2 \geq 1\} \supset \{\vartheta=\vartheta(x): (\vartheta^T q(x))^2\geq 1,\ x\in U\},$$

the second one is the corollary of the result of section 1.

The criteria

$$\Psi(\xi) = \sup_{x\in U} q^T(x)M^{-1}(\xi)q(x)$$

belongs to the family of g-criteria (see for instance, Fedorov (1981)). When $U = X$ and $q(x) = f(x)$, one can get even stronger result because of the criteria $|M(\xi)|^{-1}$ and $\sup_{x\in U} f^T(x)M^{-1}(\xi)f(x)$ are equivalent in the case of continuous designs due to Kiefer-Wolfowitz's theorem (see for example, Fedorov (1972)). This fact leads to the equivalency of (16) and (17) immediately.

4. The equivalency of some criteria can be achieved with the help of the well-known result on eigenvalues of matrices (Rao, 1973). Let M be a symmetric matrix and C be a positively definite matrix. If $\lambda_1 \geq \ldots \geq \lambda_m$ are the roots of $|M-\lambda C| = 0$ then

$$\inf_{\vartheta} \frac{\vartheta^T M\vartheta}{\vartheta^T C\vartheta} = \lambda_m . \tag{18}$$

From this relation, the equivalency of the following two criteria immediately occurs:

$$\Psi(\xi) = \lambda_m^{-1}(\xi)$$

and

$$\Phi(\xi) = \inf_{\vartheta^T C\vartheta \geq 1} \int \eta^2(x,\vartheta)\xi(dx)$$

When $C = I_m$, then $\Psi(\xi)$ is the popular E-criteria in the design theory.

The results of sections 1-4 can be summarized in

THEOREM 1. The following criteria are equivalent on the set of regular designs

1) $c^T m^-(\xi)c$ and $\inf_{(c^T\vartheta)^2 \geq \delta} \gamma(\xi,\vartheta)$,

2) $|M^{-1}(\xi)|$ and $\inf_{|D_o| \geq d} \int \gamma(\xi,\vartheta)F_o(d\vartheta)$,

3) $\sup_{x \in U}(q^T(x)M^{-1}(\xi)q(x)$ and $\inf_{\sup_{x \in U}(q^T(x)\vartheta)^2 \geq \delta} \gamma(\xi,\vartheta)$

4) $\lambda_1[B^T M^{-1}(\xi)B]$ and $\inf_{\vartheta^T BB^T\vartheta \geq \delta} \gamma(\xi,\vartheta)$, where $\delta > 0$ and

$\gamma(\xi,\vartheta) = \int_X (\vartheta^T f(x))^2 \xi(dx)$.

3. SOME PROPERTIES OF OPTIMAL DESIGNS

Theorem 1 allows some new results on the properties of optimal designs to be achieved or illuminate some of the known results both for parameter estimation and model testing problems. In application, the number of supporting points in an optimal design is one of the prime interests; the lesser the number, the simpler it is to realize in practice the corresponding optimal designs.

The results on the number of supporting points can be achieved by switching between the following two theorems (see for example, Stone (1958); Fedorov (1972); Denisov, Fedorov, and Khaborov (1981)).

THEOREM 2. In design problem (2) there exists optimal design containing no more than $m(m+1)/2$ supporting points.

THEOREM 3. In design problem (4) there exists optimal design containing no more than $(m+1)$ supporting points if Ω is a compact and convex set. If additionally in (4),at least k of constraints are active for optimal designs, then there exists optimal design containing no more than $(m-k+1)$ supporting points.

It should be noted that if the conditions of *Theorem 3* are fulfilled, it gives more strong result than *Theorem 2*.

Example 1

Consider the first case from *Theorem 1* concerned with the extrapolation problem. There exist some results on the number of supporting points in this case which are rather complicated in proving and are significantly based on the structure of basic functions $f(x)$ (see for instance, Fedorov (1972)). From *Theorem 3*, it follows that for criteria under consideration, there exist optimal design containing no more than m supporting points. To get this result, it is necessary to take into account that the design problem

$$\sup_{\xi} \inf_{(c^T\vartheta)^2 \geq \delta} \gamma(\xi, \vartheta)$$

is equivalent to

$$\sup_{\xi} \inf_{c^T\vartheta \geq \delta} \gamma(\xi,\vartheta)$$

due to evenness, both functions $(c^T\vartheta)^2$ and $\gamma(\xi,\vartheta)$ and that for any design ξ, the constraint $c^T\vartheta \geq \delta$ is active.

It is useful to note that the result does not depend on the dimension of x.

Example 2

In spite of the similarity of the model testing criteria from point 3 and 4 of *Theorem 1* to the one considered previously, it is not possible to get analogous results here. It is the matter of fact that the sets $\sup_{x \in X} (f^T(x)\vartheta)^2 \geq \delta$ or $\vartheta^T BB^T \vartheta \geq \delta$ are not convex, and therefore the result of *Theorem 3* cannot be applied. Naturally, the result of *Theorem 2* happens to be true but the bound $n_0 = m(m+1)/2$ for the number of supporting points is not very efficient and often cannot satisfy an experimenter. In these cases, more detailed analysis could be don with the help of the so-called equivalency theorems. These theorems can be formulated (see Fedorov, 1980) for both sets of optimality criteria (for parameter estimation and model testing problems).

THEOREM 4 (estimation problem). A necessary and sufficient condition for a design ξ^* to be optimal is the fulfillment of the inequality

$$\varphi(x,\xi^*) \geq \mathrm{tr}\, M \frac{\partial \Psi}{\partial M} \Big|\, M = M(\xi^*), \quad x \in X,$$

where

$$\varphi(x,\xi) = f^T(x)\frac{\partial\Psi}{\partial M}f(x).$$

If $\int_{X^1} \xi^*(dx) > 0$ then the function $\varphi(x,\xi^*)$ achieves its lower bound on the set of X^1. Naturally the existence of derivatives $\frac{\partial\Psi}{\partial M}$ is suggested.

THEOREM 5 (model testing problem). A necessary and sufficient condition for a design ξ^* to be optimal is the existence of such measure $\mu^*(d\vartheta)$ that

$$\gamma(x,\xi^*) \leq \Phi(\xi^*,\vartheta^*)$$

where

$$\gamma(x,\xi^*) = \int (f^T(x)\vartheta^*)^2\mu^*(d\vartheta)$$

and the measure μ^* is defined on the set

$$\Omega^* = \{\vartheta^*: \vartheta^* = \text{Arg}\inf_{\vartheta\in\Omega}\Psi(\xi^*,\vartheta)\}, \int_{\Omega^*}\mu^*(d\vartheta) = 1.$$

If $\int_{X^1} \xi^*(dx) > 0$ then the function $\gamma(x,\xi^*)$ achieves its upper bound on the set of X^1.

Note that in *Theorem 5* the convexity of Ω is not assumed.

Consider the polynomial regression $(f^T(x)=1,x,\ldots,x^{m-1}, |x| \leq 1)$ and prove by two different ways that the number of supporting points in the optimal design for case 3 from *Theorem 1* equals m. Let us start with *Theorem 4* repeating the well-known approach (see for instance, Fedorov (1972)). It is more convenient to put here $\Psi(\xi) = -\ln|M(\xi)|$. In this case, $-\varphi(x,\xi) = f^T(x)M^{-1}(\xi)f(x)$ and it is the polynomial of degree $2m-2$.

Evidently, this polynomial can achieve its maxima on interval $|x| \leq 1$ no more than in m points. So the number of supporting points n_0 due to *Theorem 4* cannot exceed m. If $n_0 < m$, then $|M(\xi)| = 0$. Therefore for the optimal design $n_0 = m$.

Apply now to *Theorem 5* to get the same result. It is clear that the function $\gamma(x,\xi)$ is a polynomial of degree less or equal $2m-2$. Repeating the last part of the previous proof, one gets $n_0 = m$.

Example 3

Consider now case (4) from *Theorem 1* with $B = I$ for the sake of simplicity. *Theorem 4* cannot be used here without additional considerations because generally, optimal designs can have nonunique largest eigenvalue $\lambda[M^{-1}(\xi^*)]$ and function $\Psi(M)$ is nondifferentiable in this case. But *Theorem 5* works here and similar to the previous case, one gets $n_0 = m$ for one dimensional polynomial regression of degree $m - 1$.

Note that *Theorem 5* becomes more convenient to use particularly when rank $B < m$ and one faces the nondifferentiability of $\Psi[M]$ in most cases due to the possible singularity of the information matrix $M(\xi^*)$.

4. NUMERICAL PROCEDURES

Theorem 1 enables one to choose between principally different algorithms of the numerical construction of optimal designs. The first set of algorithms based on *Theorem 4* and their description can be found in [Fedorov (1972); Silvy, (1980)]. The algorithms related to model testing criteria were described in [Atkinson and Fedorov (1975); Denisov, Fedorov and Khaborov (1981)].

REFERENCES

Atkinson, A., and V. Fedorov (1975) The Design of Experiments for Discriminating Between two Rival Models. *Biometrika* 62, 57-70.

Denisov, V., V. Fedorov, and V. Khaborov (1981) Tchebysheff's Approximation in the Problem of Asymptotical Locally-Optimal Designs Construction for Discriminating Experiments, in "Linear and Nonlinear Parameterization in Experimental Design Problems". *Problems in Cybernetics*, V. Fedorov and V. Nalimov (eds.), pp.3-10 (in Russian). Moscow.

Fedorov, V. (1972) *Theory of Optimal Experiments*. New York: Academic Press.

Fedorov, V. (1980) Design of Model Testing Experiments, in *Symposia Mathematica*, pp.171-180. Bologna.

Fedorov, V. (1980) Convex Design Theory. *Math. Operationsforsch. Statist., Ser. Statist.* 11:403-413.

Fedorov, V. (1981) Active Regression Experiments, edited by V. Penenko. *Mathematical Methods in Experimental Design*, pp. 19-73, Novoskibirsk, Nauka (in Russian).

Kiefer, J. (1958) On the Nonrandomized Optimality and Randomized Nonoptimality of Symmetrical Designs. *Am. Math. Statist.* 29:675-699.

Rao, C.R. (1973) *Linear Statistical Inference and its Application*. New York: J. Wiley and Sons.

Silvy, S.D. (1980) *Optimal Design*. London: Chapmann and Hill.

Stone, M. (1958) Application of a Measure of Information to the design and comparison of Regression Experiments. *Am. Math. Stat.* 29:55-70.