# Constrained Optimization of Experimental Design

Dennis Cook & Valery Fedorov†

School of Statistics, University of Minnesota

† Currently at Computer Science and Mathematics Division, Oak Ridge National Laboratory, U.S.A.

Published in:

*Statistics* 26 (1995) 129-178

**Abstract**

This is an attempt to discuss various approaches developed in experimental design when some constraints are imposed. The constraints may be on the total cost of the experiment, the location of the supporting point, the value of the auxiliary objective functions, and so on. The basic idea of the paper is that corresponding optimization problems can be imbedded in the convex theory of experimental design. Part 1 is concerned with the properties of optimal designs, while Part 2 is devoted mainly to numerical methods.

We have tried to avoid detail emphasizing ideas rather than technicalities. This is not intended as a literature review. The authors subjectively, surely left many excellent papers behind.

# 1 Equivalence Theorems

## 1.1 Introduction

Experimental design problems considered in this paper are basically related to the standard linear regression model:

$$y_{ij} = \eta(x_i, \theta) + \epsilon_{ij}, \tag{1}$$

$$i = 1, \ldots, n, \qquad j = 1, \ldots, r_i, \qquad \Sigma r_i = N,$$

$$\eta(x, \theta) = \theta^T f(x)$$

where $\theta \epsilon R^m$ are unknown parameters, $f^T(x) = (f_1(x), \ldots, f_m(x))$ are given functions, supporting points $x_i$ can be chosen from some set $X, \varepsilon_{ij}$ are uncorrelated random errors with zero means and variances equal to 1.

For the best linear unbiased estimator of unknown parameters the accumulated "accuracy" is described by the information matrix:

$$M(\xi) = \Sigma p_i f(x_i) f^T(x_i), \qquad p_i = r_i/N,$$

which is completely defined by the design $\xi = \{x_i, p_i\}_1^N$. In the context of the convex design theory:

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

where $\xi(dx)$ is a probability measure ("large sample" or continuous) design with the supporting set belonging to $X : supp\ \xi \subset X$. The subscript corresponding to the area of integration will be used only when it is essential for understanding. Assume that $\Psi$ is an objective function. A design

$$\xi^* = \arg\min_{\xi} \Psi[M(\xi)], \qquad \int \xi(dx) = 1, \tag{2}$$

is called $(\Psi)$ optimal.

In the traditional case minimization has to be over the set of all possible probability measures $\Xi$ with supporting sets belonging to $X$. For a practitioner it means that one has typically to find an optimal design with a given number of observations, i.e. $\sum_{i=1}^n r_i = N$ or $\sum_{i=1}^n p_i = 1$. Of course, the reality can be worse and more constraints can be imposed.

For instance, the cost of observations may depend upon $x$ with the total cost of the experiment not to exceed some level.

Sometimes together with the parameters of model (1) one may wish to estimate the parameters of some competing model. Then in addition to (2) it is reasonable to demand that the information matrix for the competing model is not very "small".

In the observation network optimization problem for example it is usually not reasonable to locate several sensors at the same site nor with very small distances between them. This leads to the restriction of the number of sensors per square unit. In terms of continuous designs it means that the density of the design measure has to be restricted.

Thus additionally to (2) one has to consider design problems when the structure of $\Xi$ is more complicated than in the traditional experimental design theory, which is very briefly surveyed in Section 2. In Sections 3–5 the various types of constraints are considered, and the main focusing being on the similarities among the corresponding results rather than on the discrepancies between them.

### 1.2 Standard Equivalence Theorem

The optimization problem (2) has been intensively studied since Kiefer's pioneering paper (1959). In this section we shall summarize the major properties of the traditional optimal designs.

Assume that:

(a) $X$ is compact;

(b) $f(x)$ are continuous functions in $X, f \epsilon R^m$;

(c) $\Psi(M)$ is a convex function;

(d) there exists a real number $q$ such that

$$\{\xi : \Psi[M(\xi)] \leq q < \infty\} = \Xi(q) \neq \emptyset;$$

(e) for any $\xi \epsilon \Xi(q)$ and $\bar{\xi} \epsilon \Xi$:

$$\Psi[(1-\alpha)M(\xi) + \alpha M(\bar{\xi})] = \Psi[M(\xi)] + \alpha \int \psi(x,\xi)\bar{\xi}(dx) + \tau(\alpha,\xi,\bar{\xi}),$$

where $\tau(\alpha,\xi,\bar{\xi}) = o(\alpha)$.

Assumptions (c) and (e) are most essential and restrictive for the theory. Fortunately the majority of the popular optimality criteria satisfy them, such as D– and linear criteria with regular optimal designs. But there exist some natural and widely used criteria which do not satisfy (e) (for instance, the minimax ones). One can face similar troubles even for "good" criteria when an optimal design happens to be singular; see Silvey (1980), Pukelsheim (1980), for details and

We shall use notation $\Psi(\xi)$ for $\Psi[M(\xi)], \Psi^*$ for $\Psi(\xi^*)$ and $\min_x, \min_\xi, \int$, and so on, instead of $\min_{x\epsilon X}, \min_{\xi\epsilon\Xi}, \int_X$, respectively, if it does not lead to ambiguity.

Assumptions (a) and (b) can be avoided in the most general approach to the convex design, see Ermakov (1983), but this leads to some technical difficulties and inconveniences which are annoying while not essential (see Fedorov and Uspensky (1975) for "regularization" results). Therefore we prefer to work with (a) and (b) enforced.

**THEOREM 1.** If (a)–(e) hold, then

1. For any optimal design there exists a design with the same information matrix and containing no more than $n = m(m+1)/2$ supporting points.

2. A necessary and sufficient condition for a design $\xi^*$ to be optimal is fulfillment of the inequality

$$\min_x \psi(x, \xi^*) \geq 0. \tag{3}$$

3. The set of optimal designs is convex.

4. $\psi(x, \xi^*)$ achieves zero almost everywhere in $supp\ \xi^*$.

**Proof.** The proof of this theorem is well known and is only sketched here to clarify the main ideas which are used in the subsequent sections. The proof of the first part of the theorem is based on Caratheodory's theorem and on the fact that any information matrix can be considered as an element of the convex hull of the elementary information matrices:

$$\mathbf{m}(x) = f(x) f^T(x) \epsilon R^{m(m+1)/2}, x \epsilon X.$$

Necessity and sufficiency of (3) follows from the fact that the following inequality

$$\min_{\xi \epsilon \Xi} \lim_{\alpha \to 0} \frac{\Psi[(1-\alpha)M(\xi^*) + \alpha M(\xi)] - \Psi[M(\xi^*)]}{\alpha} \geq 0 \tag{4}$$

is a necessary and sufficient condition of optimality of $\xi^*$. If (e) holds, then (4) can be easily transformed to

$$\min_\xi \int \psi(x, \xi^*) \xi(dx) \geq 0 \tag{5}$$

and subsequently

$$\min_x \psi(x, \xi^*) \geq 0. \tag{6}$$

The third part of the theorem follows directly from the convexity of the objective function.

Integration of both parts of the expression from (e) with respect to $\xi^*(dx)$ confirms the final section of the theorem.

Theorem 1 is the basic one in convex design theory and its various modifications have been extensively discussed in the statistical literature: see Fedorov (1972), Fedorov and Malyutov (1972), Whittle (1973), Silvey (1980).

**Example 1.**
For the D–criterion, $\Psi(M) = -ln|M|$ and

$$\psi(x, \xi) = m - d(x, \xi),$$

where

$$d(x, \xi) = f(x)^T M^{-1}(\xi) f(x).$$

Point 2 of the theorem can be reformulated in the more traditional form:

The following problems:

$$\xi^* = \arg\max_\xi \ ln|M(\xi)|$$

$$\xi^* = \arg\min_{\xi} \max_{x} d(x,\xi)$$

$$\max_{x} d(x,\xi^*) = m$$

are equivalent.

This is Kiefer's celebrated equivalence theorem.

### 1.3 Linear Constraints

Constraints linear with the respect to the design measure mainly arise in experiments when the cost of an observation depends upon controlled variables. The corresponding optimization problem can be stated now in the following way:

$$\xi^* = \arg\min_{\xi} \Psi[M(\xi)] \tag{7}$$

$$s.t. \int \xi(dx) = 1, \qquad C(\xi) = \int \phi(x)\xi(dx) \leq 0, \tag{8}$$

where $\phi = (\phi_1, \ldots, \phi_l)^T$. In what follows the first constraint out of (8) will be usually skipped.

**Example 2.**
Let the functions $\zeta_\alpha(x), \alpha = 1, \ldots, l$, be the costs when observation is taken at point $x$. Assume that the total cost for a particular $\alpha$ can not exceed $C_\alpha$. Then

$$\Sigma r_i \zeta_\alpha(x_i) \leq C_\alpha, \qquad \alpha = 1, \ldots, l,$$

where $r_i$ is the number of observations at point $x_i$.

For continuous designs the latter inequality takes the form:

$$\int \phi(x)\xi(dx) \leq 0,$$

where $\phi_\alpha(x) = \zeta_\alpha(x) - c_\alpha,\ c_\alpha = C_\alpha/N$.

**Example 3.**
In model testing experiments (see Atkinson &Fedorov, 1975; Fedorov & Khabarov, 1986) when there are two rival response models, i.e.

$$\eta(x,\theta) = \begin{cases} \eta_1(x,\theta_1) \\ \eta_2(x,\theta_2) \end{cases}$$

the design problem consists of maximization of some discrepancy measure between these competing models. This case can be transformed to the case when a response function $\eta(x,\theta)$ is compared with the totally null hypothesis $\eta_2(x,\theta_2) \equiv 0$.

If there is some prior information on the parameters $\theta$ described by a prior distribution function $F(d\theta)$, then one can consider the following discrepancy measure

$$\int\int \eta^2(x,\theta)F(d\theta)\xi(dx)$$

which has to be maximized in the model testing experiments.

If the parameter estimation is considered as the primary problem, but the model testing problem is still of interest, then one may consider (7) and (8) with

$$\phi(x) = c - \int \eta^2(x,\theta)dF(\theta),$$

where $c$ is the guaranteed value of the discrepancy measure. In the case of linear response models:

$$\phi(x) = c - f^T(x)Df(x),$$

where

$$D = \int \theta\theta^T dF(\theta).$$

We consider the optimization problem (7) under assumptions (a)–(e) in addition to the condition $(b')$ $\phi(x)$ are continuous for all $x\epsilon X$.

**THEOREM 2.** If (a)–(e) and $(b')$ hold, then

1. For any optimal design there exists a design with the same information matrix and containing no more than $m(m+1)/2 + l$ supporting points.
2. A necessary and sufficient condition for a design $\xi^*$ to be optimal is fulfillment of the inequality

$$\min_x \; q(x,u^*,\xi^*) \geq 0,$$

where

$$q(x,u,\xi) = \psi(x,\xi) + u^T\phi(x),$$

$$u^* = \arg\max_{u\epsilon U}\;\min_x\; q(x,u,\xi^*),$$

$$U = \{u : u\epsilon R^l, u_\alpha \geq 0\}.$$

3. The set of optimal designs is convex.
4. $q(x,u^*,\xi^*)$ achieves zero almost everywhere in *supp* $\xi^*$.

**Proof.** To prove the first part of the theorem it is sufficient to notice that any couple $M(\xi), C(\xi)$ belongs to the convex hull of

$$\{m(x),\phi(x)\}\epsilon R^{m(m+1)/2+l}, x\epsilon X.$$

and then apply Caratheodory's theorem.

To prove the second part of the theorem one has to add to (5) the constraints (8):

$$\min_{\xi} \int \psi(x,\xi^*)\xi(dx) \geq 0 \tag{9}$$

where the minimum is taken over all designs such that

$$\int \phi(x)\xi(dx) \leq 0. \tag{10}$$

The fulfillment of inequalities (9) and (10) are necessary and sufficient condition for the optimality of $\xi^*$. But unlike the standard case there is generally no single point design (see (6)) simultaneously satisfying both of them.

The Lagrangian technique (see, for instance, Laurent (1972), Ch. 7) leads to the duality of optimization problem (9), (10) with the following maxi–min problem:

$$\max_{u \epsilon U} \min_{\xi} \int q(x,u,\xi^*)\xi(dx),$$

or equivalently

$$\max_{u \epsilon U} \min_{x} \int q(x,u,\xi^*)$$

(see, for instance, Fedorov and Gaivoronski (1984)) confirming the assertion of the theorem. The proof of two last parts of the theorem is identical to the standard case.

**Note 1.** The existence of a solution of (9), (10) with no more than ($l$+1) supporting points follows from Caratheodory's theorem.

**Note 2.** In the frame of Example 2 constraints (8) means that one is constrained both in the number of observations:

$$\sum_{i=1}^{n} r_i = N \Rightarrow \int \xi(dx) = 1$$

and in their total cost ($l = 1$)

$$\sum_{i=1}^{n} r_i\zeta(x_i) \leq C \Rightarrow \int \phi(x)\xi(dx) \leq 0 \text{ or} \int \zeta(x)\xi(dx) \leq c.$$

It is essential that there are two linear constraints. If the total number of observations is not restricted, then the transformation

$$\xi'(dx) = \phi(x)\xi(dx), \qquad f'(x) = \phi^{-1/2}(x)f(x)$$

returns us to the standard case (2), compare with Chernoff (1972, p. 16), Kiefer (1959).

**Example 4.**

Let us consider the design problem for a one–dimensional polynomial response and the D–criterion:

$$f_\alpha(x) = x^{\alpha-1}, |x| \leq 1, \Psi(M) = -ln|M|, \alpha = 1,\ldots,m$$

with linear constraints:

$$\int_{-1}^{1} \phi(x)\xi(dx) \leq 0,$$

and let $\{f, \phi\}$ constitute a Chebyshev system on $|x| \leq 1$.

From Example 1 it follows that function

$$q(x, u, \xi) = m - \Sigma M^{-1}_{\alpha\beta} x^{\alpha-1} x^{\beta-1} + u^T \phi(x),$$

where $M^{-1}_{\alpha\beta}$ are elements of the inverse matrix $M^{-1}$, is a linear combination of $2m + l$ Chebyshev's functions with some nonzero coefficients. Therefore (see, for instance, Karlin and Studden, 1966) this function has no more than $2m + l$ roots and subsequently has no more than $m + l/2$ (if $l$ is even) or $m + (l+1)/2$ (if $l$ is odd) minima on the interval $|x| \leq 1$. But in accordance with Theorem 2 they have to coincide with the support points. So for this case the number of support points is essentially less than $m(m + 1)/2$.

## 1.4 Nonlinear Convex Constraints and Linearization

The approach considered in the previous section can be used for the more general design problem:

$$\xi^* = \arg\min_{\xi} \Psi[M(\xi)], \tag{11}$$

$$s.t. \Phi(\xi) \leq 0, \qquad \Phi \epsilon R^l. \tag{12}$$

Assume additionally to (a)–(e),(b$'$) that:

(c$'$)$\Phi(\xi)$ are convex;

(e$'$)$\Phi[(1-\alpha)\xi + \alpha\bar{\xi}] = \Phi(\xi) + \alpha \int \phi(x, \xi)\bar{\xi}(dx) + \tau(\alpha, \xi, \bar{\xi})$, where $\tau_k(\alpha, \xi, \bar{\xi}) = o(\alpha),\ k = 1, \ldots, l, \xi$ and $\bar{\xi}$ are defined in (e) with $\Xi(q)$ and $\Xi$ satisfying (10).

The analysis of (11), (12) is mainly based on ideas of Theorem 2 and on the possibility of linearization of $\Phi(\xi)$ near an optimal design (compare with Gaivoronski, 1984 and Lee, 1988).

All the final results can be described by Theorem 2 with functions $\phi(x, \xi)$ defined in (e$'$) and $\Phi(\xi) \leq 0$. Of course, it is assumed that the set of $\xi$ satisfying (12) is not empty and all constraints are active. We shall refer to Theorem 2 in the case of nonlinear constraints, but one has to remember about assumptions (c$'$),(e$'$).

**Example 5.**
Let $\Psi = -ln|M|$ and $\phi_k = -ln|M_k| - c_k$, where

$$M_k(\xi) = \int f_k(x) f_k^T(x) \xi(dx).$$

Then the design problem corresponds to the case when one wishes to find a D–optimal design for the response $\theta^T f(x)$ and to be sure that this design is efficient for some alternative responses $\theta_k^T f_k(x)$. Taking into account that:

$$\psi(x, \xi) = m - d(x, \xi),$$

$$\phi_k(x,\xi) = m_k - d_k(x,\xi),$$

$$d_k(x,\xi) = f_k^T(x)M_k^{-1}(\xi)f_k(x),$$

and assuming that $f(x)$ and $f_k(x), k = 1,\ldots,l$ are continuous in $X$ and $c_k$ are not very small (to provide fulfillment of (d) together with (10)) it is not difficult to check the validity of Theorem 2. From this theorem it follows that a necessary and sufficient condition for optimality of $\xi^*$ is existence of $u^*\epsilon U$ such that

$$d(x,\xi^*) + \Sigma u_k^* d_k(x,\xi^*) \leq m + \Sigma u_k^* m_k$$

while $\Phi(\xi) = 0$;

almost everywhere in *supp* $\xi^*$

$$d(x,\xi^*) + \Sigma u_k^* d_k(x,\xi^*) = m + \Sigma u_k^* m_k.$$

A number of similar examples for various optimality criteria can be found in Lee, 1988.

Theorems 1 - 2 can be considered as specific cases of the Kuhn–Tucker Theorem, and, of course, its versions and generalizations of this theorem can help to extend the previous results. For instance, the Kuhn–Tucker Theorem for the case with a continuum of constraints (see Pshenichnyi, 1971, Ch. 5.2) allows analysis of the following design problems:

$$\xi^* = \arg\min_{\xi} \ \Psi[M(\xi)], \tag{13}$$

$$s.t. \Phi(\xi,\lambda) \leq 0, \quad \Phi\epsilon R^1, \quad \text{for all } \lambda\epsilon\Lambda \subset R^l. \tag{14}$$

Let there be in addition to the previous assumptions:

(c$''$)$\Phi(\xi,\lambda)$ is convex for all $\lambda\epsilon\Lambda$ and $\Lambda$ is compact;

(e$''$)$\Phi[(1-\alpha)\xi + \alpha\bar{\xi},\lambda] = \Phi(\xi,\lambda) + \alpha\int\phi(x,\xi,\lambda)\bar{\xi}(dx) + \tau(\alpha,\xi,\bar{\xi},\lambda)$, where $\tau(\alpha,\xi,\bar{\xi},\lambda) = O(\alpha)$. Then the above mentioned theorem leads to

**THEOREM 3.** A necessary and sufficient condition for a design $\xi^*$ to be optimal is the existence of such $u^*\epsilon U$ and $\lambda_k^*\epsilon\Lambda$ that:

$$\min_x \ q(x,u^*,\xi^*) \geq 0, \tag{15}$$

where $q(x,u,\xi) = \psi(x,\xi) + u^T\phi(x,\xi), \Phi(\xi^*,\lambda^*) = 0,$

$\phi_k(x,\xi) = \phi(x,\xi,\lambda_k^*), U = \{u : u\epsilon R^{l+1}, u_k \geq 0\}, k = 1,\ldots,l+1.$

Assume that all $\lambda_k^*$ are known. Then Theorem 3 states that (12), (14) are equivalent to (11), (12) with $\Phi(\xi) = (\Phi(\xi,\lambda_1^*),\ldots,\Phi(\xi,\lambda_{l+1}^*))^T$. This makes it evident how to use the results of the previous theorems.

**Example 6.**
As in Example 4 suppose one wishes to find a D–optimal design for the response $\theta^T f(x)$ and to be sure that this design is efficient for some competing nonlinear response $\eta(x,\lambda)$, i.e.

$$-ln|M(\xi,\lambda)| \leq c,$$

$$M(\xi,\lambda) = \int f(x,\lambda) f^T(x,\lambda)\xi(dx), f(x,\lambda) = \partial\eta(x,\lambda)/\partial\lambda, \lambda\epsilon\Lambda \subset R^1.$$

The combination of the results of Example 4 and Theorem 3 transforms (15) to:

$$d(x,\xi^*) + \Sigma u_k^* d_k(x,\xi^*) \leq m + l\Sigma u_k^*,$$

$$d_k(x,\xi) = f^T(x,\lambda_k^*) M^{-1}(\xi,\lambda_k^*) f(x,\lambda_k^*), k = 1,\ldots,l+1,$$

where

$$\lambda_k^* \epsilon \Lambda : -ln|M(\xi^*,\lambda)| = c$$

If $f(x) = f(x,\lambda')$, where $\lambda'$ is a prior value of the parameters $\lambda$, this example can be considered as a particular case of the design problem for the nonlinear response, see Atkinson & Fedorov, 1988.

### 1.5 Directly Constrained Design Measures

A number of experimental design problems can be formulated as optimization problems with explicitly bounded measures (see Fedorov, 1986, 1989, Wynn, 1982):

$$\xi^* = \arg \min_{\xi} \ \Psi[M(\xi)] \tag{16}$$

$$s.t. \ \ \xi(dx) \leq \Phi(dx), \ \int \Phi(dx) = C \geq 1. \tag{17}$$

**Example 7.**
Let $N(\Delta X)$ be the number of sensors per $\Delta X$. Then for sufficiently large total number of sensors $N$ one can use the following approximation

$$N(\Delta X) \simeq \xi(\Delta X) \ \ N \simeq \zeta(X)\Delta X N,$$

where $\zeta(X)$ can be considered as a density function. This density may be chosen by practitioner under some constraints (for instance, in some areas the density of sensors must be half than in others.) Thus one comes to the necessity to introduce (17).

As in the moment space theory (see Fedorov, 1989, Karlin and Studden, 1966, Krein and Nudelman, 1973) $\xi^*$ can be called a $\Psi,\Phi$–optimal design. Assume additionally to (a)–(e) that:

(f) $\Phi(dx)$ is atomless, i.e. $\lim_{n\to\infty} \int_{\Delta X_n} \Phi(dx) = 0$, for any diminishing sequence $\Delta X_n$.

Sets $\Xi$ and $\Xi(q)$ in (e) have to satisfy (17).

Let $\bar{\Xi}$ be a set of design measures such that $\xi(\Delta x) = \Phi(\Delta x)$ for any $\Delta X \subset supp\ \xi$.

A function $\phi(x, \xi)$ is said to separate sets $X_1$ and $X_2$ with respect to the measure $\Phi(dx)$ if for any two sets $\Delta X_1 \epsilon X_1$ and $\Delta X_2 \epsilon X_2$ with equal nonzero measures:

$$\int_{\Delta X_1} \phi(x, \xi)\Phi(dx) \leq \int_{\Delta X_2} \phi(x, \xi)\Phi(dx).$$

**THEOREM 4.** If assumptions (a)–(f) hold, then:

1. $\xi^* \epsilon \Xi$ exists.
2. A necessary and sufficient condition of $\Psi, \Phi$–optimality of $\xi^* \epsilon \Xi$ is that $\psi(x, \xi^*)$ separates $X^* = supp\ \xi^*$ and $X \backslash X^*$.

**Proof.** The results of the theorem are strongly related to the moment spaces theory, and the proof is based on the corresponding ideas.

The existence of an optimal design follows from the compactness of the set of information matrices. The compactness of the latter is provided by (a), (b) and (f). The fact that at least one optimal design belongs to $\bar{\Xi}$ is the corollary of Liapunov's Theorem on the range of a vector measure (see Karlin and Studden, 1966, Ch. VIII).

Necessity follows from the fact that if there exist $\Delta X_1 \subset X^*$ and $\Delta X_2 \subset X \backslash X^*$ with equal nonzero measures such that:

$$\int_{\Delta X_1} \psi(x, \xi^*)\Phi(dx) > \int_{\Delta X_2} \psi(x, \xi^*)\Phi(dx),$$

then deletion of the first set from the supporting set with the subsequent inclusion of the second one causes the decrease of $\Psi$. This contradicts the optimality of $\xi^*$.

Now assume that $\xi^* \epsilon \bar{\Xi}$ is nonoptimal and $\xi \epsilon \bar{\Xi}$ is optimal, i.e.:

$$\Psi[M(\xi^*)] > \Psi[M(\xi)] + \delta, \qquad \delta > 0.$$

Let $\bar{\xi} = (1 - \alpha)\xi^* + \alpha\xi$, then:

$$\Psi[M(\bar{\xi})] \leq (1 - \alpha)\Psi[M(\xi^*)] + \alpha\Psi[M(\xi)]$$

$$< (1 - \alpha)\Psi[M(\xi^*)] + \alpha\{\Psi[M(\xi^*)] - \delta\} = \Psi[M(\xi^*)] - \alpha\delta. \tag{18}$$

Simultaneously:

$$\Psi[M(\bar{\xi})] = \Psi[M(\xi^*)] + \alpha \int_X \psi(x, \xi^*)\xi(dx) + o(\alpha)$$

Let $\qquad supp\ \xi = (X^* \backslash D) \cup E, D \subset X^*, E \subset (X \backslash X^*), E \cap D = \emptyset$

Then $$\int_E \Phi(dx) = \int_D \Phi(dx)$$

and $$\int_X \psi(x,\xi^*)\xi(dx) = \int_{X^*} \psi(x,\xi^*)\Phi(dx)$$
$$+ \int_E \psi(x,\xi^*)\Phi(dx) - \int_D \psi(x,\xi^*)\Phi(dx).$$

From assumption (e) it follows that

$$\int_{X^*} \psi(x,\xi^*)\Phi(dx) = \int_X \psi(x,\xi^*)\xi^*(dx) = 0,$$

and from the statement of the theorem it follows that

$$\min_{x\epsilon E} \psi(x,\xi^*) \geq \max_{x\epsilon D} \psi(x,\xi^*).$$

Therefore
$$\int_X \psi(x,\xi^*)\xi(dx) \geq 0$$

Subsequently

$$\Psi[M(\bar{\xi})] \geq \Psi[M(\xi^*)] + o(\alpha) \tag{19}$$

The comparison of (18) and (19) gives a contradiction, and this completes the proof.

**Note 3.** The comparison of Theorems 1, 2, and 4 gives a hint how the latter one can be generalized when to constraints (17) one adds (8), or (12), or (14). For this purpose the function $\psi(x,\xi)$ has to be replaced by a corresponding function $q(x,u,\xi) = \psi(x,\xi) + u^T\phi(x,\xi)$.

## 1.6 Marginally Restricted Designs

Assume that in regression model (1) the vector of independent variables consists of two parts, $x$ and $t$, and that only $x$ can be controlled by design. The design measure $\xi$ still has to be defined on $X \otimes T$, however. Unlike the previous approach when only the compactness of the design region was assumed, the vector of design variables is not constrained to a structured set. Both $X$ and $T$ are assumed to be compact.

While the design variables associated with $t$ cannot be controlled, we suppose that the corresponding coordinates of the support points of the final design are predetermined. This leads to the following optimization problem (See Cook and Thibodeau, 1980; Atkinson and Fedorov, 1988; Huang and Hsu, 1993):

$$\xi^* = \arg\min_{\xi} \Psi[M(\xi)] \tag{20}$$

$$s.t. \int_X \xi(dx,dt) = \tau(dt) \tag{21}$$

where $\tau(dt)$ is a given measure and $\int_T \tau(dt) = 1$.

If $x$ is replaced by $(x,t)$ in assumptions (a)–(e), then the following theorem follows from the ideas of Section 1.2.

**THEOREM 5.** If assumptions (a)–(e) hold with $x$ replaced with $(x,t)$, then a necessary and sufficient condition for the optimality of $\xi^*(dx,dt)$ is that

$$\min_x \; \psi(x,t,\xi^*) \geq 0 \quad \text{for all} \quad t \epsilon T \tag{22}$$

and

$$\psi(x,t,\xi^*) = 0 \;\; \text{when} \;\; x\epsilon \; supp \; \xi^*(dx|t),$$

where

$$\xi(dx|t) = \frac{\xi(dx,dt)}{\tau(dt)}$$

provided $\tau > 0$.

**Example 8.** In the case of the D–optimality crieterion,

$$\psi(x,t,\xi) = \int_X d(x,t,\xi)\xi(dx|t) - d(x,t,\xi),$$

where

$$d(x,t,\xi) = f^T(x,t,\xi)M^{-1}(\xi)f(x,t,\xi),$$

or a necessary and sufficient condition for D–optimality of $\Xi^*$ is that

$$\max_x \; d(x,t,\xi) \leq \int_X \; d(x,t,\xi)\xi(dx|t),$$

for any $t\epsilon T$.

# 2 Numerical Procedures of the First Order

## 2.1 Standard Case

The main objective of this section is to discuss the numerical methods for finding a solution of the optimization problem (2).

In what follows we presume that assumptions (a)–(d) are held. Instead of (e) the similar, but more restrictive assumption have to be made:

(e $_0$) for any $\xi\epsilon\Xi(q)$ and $\bar{\xi}\epsilon\Xi$:

$$\Psi[(1-\alpha)M(\xi) + \alpha M(\bar{\xi})] = \Psi[M(\xi)] + \alpha \int \psi(x,\xi)\bar{\xi}(dx) + \tau[\xi(\alpha)],$$

where $\tau[\xi(\alpha)] \leq \alpha^2 K_q, K_q > 0$ (i.e. the first order approximation holds uniformly with respect to $\bar{\xi}$).

We shall consider the iterative algorithms of the following type:

$$\xi_{s+1} = (1-\alpha_s)\xi_s + \alpha_s \xi^s, \tag{23}$$

where

$$\xi^s = \arg \min_{\xi} \Psi[(1-\alpha_s)\xi_s + \alpha_s \xi]. \tag{24}$$

When the first order approximation (see assumption (e$''$)):

$$\xi^s = \arg \min_{\xi} \{\Psi[M(\xi_s)] + \alpha_s \int \psi(x,\xi_s)\xi(dx)\}$$

is substituted into (21), then the corresponding algorithm is called "the first order algorithm".

The simplicity and popularity of the first order algorithms is explained by the fact that a design $\xi^s$ does not depend upon $\alpha$ and

$$\xi^s = \arg \min_{\xi} \int \psi(x,\xi_s)\xi(dx). \tag{25}$$

Optimization problem (22) is linear with respect to $\xi$ thus is easier than the original problem (21). Moreover, when there are no constraints imposed on $\xi$ then

$$\xi^s = \xi(x_s), \quad x_s = \arg \min_{x} \psi(x,\xi_s) \tag{26}$$

where $\xi(x)$ is a design measure completely atomized at point $x$.

Approximations (21)–(23) allow the development of a number of the first order algorithms. All of them can be imbedded in the following scheme:

i. There is a design $\xi_s \epsilon \Xi_q$. Find

$$x_s = \arg \min_{x} \psi(x,\xi_s).$$

ii. Choose $0 \le \alpha_s \le 1$ and construct

$$\xi_{s+1} = (1-\alpha_s)\xi_s + \alpha_s \xi(x_s).$$

The choice of a sequence $\alpha_s$ defines a variety of the algorithms (see Atkinson & Donev, 1992, Cook & Nachtsheim, 1989, Fedorov 1972, 1975, Silvey, 1980).

The following three sequences $\alpha_s$ are most popular (especially in the theoretical considerations):

1. $\lim \quad \alpha_s = 0, \quad \sum_{s=0}^{\infty} \alpha_s = \infty;$
2. $\alpha_s = \arg\ \min_\alpha \ \Psi[M(\xi_{s+1}(\alpha))], \quad$ where $\xi_{s+1}(\alpha) = (1-\alpha)\xi_s + \alpha\xi(x_s)$
3. $\alpha_s = \begin{cases} \alpha_{s-1}, & \text{if } \Psi[M(\xi_s(\alpha_{s-1}))] < \Psi[M(\xi_{s-1})]; \\ \alpha_{s-1}/\gamma,\ \gamma > 1 & \text{otherwise.} \end{cases}$

Proving the convergence of the algorithm corresponding to any of these sequences is rather standard (see Fedorov, 1975, 1986, Wu & Wynn, 1978) for the optimization theory.

For instance, let us prove the (weak) convergence of the algorithm for $\{\alpha_s\}$ defined by rule (2).

**Lemma 1.** If (a)–($e_0$) hold and $\xi, \bar{\xi} \epsilon \Xi_q$,
then

$$\min_x \ \psi(x,\xi) \leq \int \psi(x,\xi)\bar{\xi}(dx) \leq \Psi[M(\bar{\xi})] - \Psi[M(\xi)].$$

**Proof.** The result is a corollary of the convexity of $\Psi(M)$.

**THEOREM 6.** If (a)–(d), ($e_0$) hold, then for $\{\alpha_s\}$ defined by rule (2) the iterative procedure (i), (ii) converges:

$$\lim_{s \to \infty} \ \Psi(\xi_s) = \Psi^*. \tag{27}$$

where $\Psi^*$ is the minimal value of $\Psi$.

**Proof.** By the definition $\{\Psi(\xi_s)\}$ is monotonously decreasing and therefore it converges:

$$\lim_{s \to \infty} \ \Psi(\xi_s) = \Psi_* \tag{28}$$

Assume that $\Psi_*$ is not optimal value of the objective function. Then

$$\Psi_* - \Psi^* = \delta > 0.$$

It means that for any s (see Theorem 1):

$$\min_x \ \psi(x,\xi_s) \leq -\delta. \tag{29}$$

Rule (2) together with assumption ($e''$) provide that

$$\Psi(\xi_s) - \Psi(\xi_{s+1}) \geq \delta/4K_q > 0$$

and subsequently

$$\lim_{s \to \infty} \ \Psi(\xi_s) = -\infty.$$

contradicting to (25) and that proves (26).

Iterative procedure (i), (ii) with any rule (1)–(3) guarantees the convergence of the corresponding algorithms, but they are rather slow and serve more to the theory than to the practice.

The significant improvement in the rate of convergence is observing when in iterative procedure (i), (ii) is modified in the following way (see Atwood, 1973, Fedorov, 1975):

$$x_s = \arg\ \min\ [\psi(x_s^+,\xi_s), -\psi(x_s^-,\xi_s)], \tag{30}$$

where

$$x_s^+ = \arg\ \min_{x \epsilon X}\ \psi(x,\xi_s) \text{ and } x_s^- = \arg\ \min_{X \setminus X_s}\ \psi(x,\xi_s), X_s = supp\ \xi_s.$$

and

$$\alpha_s = \begin{cases} \gamma_s, & x_s = x_s^+ \\ -\min[\gamma_s, \quad p_s/(1-p_s)], & x_s = x_s^- \end{cases} \tag{31}$$

where $\{\gamma_s\}$ has to obey one of the rules (1)–(3) and $p_s$ is the measure of a point $x_s^- \epsilon X_s$.

The iterative procedure (i), (ii) & (30), (31) is usually complemented by some clustering rule, for instance:

all supporting points $x_i \epsilon X_{s-1}$ satisfying the inequality

$$||x_i - x_s^+|| \leq d \tag{32}$$

have to be merged with the point $x_s^+$, i.e. all their measures have to be transferred to $x_s^+$.

One can introduce "forward" and "backward" excursions when correspondingly $n^+$ steps operating with "new" points $x_s^+$ and $n^-$ steps handling points $x_s^-$ are doing subsequently; the numbers $n^+$ and $n^-$ are called a "length" of an excursion (see Mitchell, 1974).

Computations became simpler when in (27) one looks for $x_s^+$ which is only an approximate solution of the corresponding optimization problem, i.e.

$$\psi(x_s^+, \xi_s) = \min_x \; \psi(x, \xi_s) + \delta_s, \tag{33}$$

where $\lim_{s \to \infty} \quad \delta_s = 0$.

All these changes or improvements do not destroy the convergence of the iterative procedure.

## 2.2 Linear Constraints. Direct First Order Algorithm.

For the optimization problem (7) & (8), an iterative procedure very similar to the procedure (i) and (ii) (with all improvements discussed at the end of the previous section if one needs them) can be used to construct $\xi_*$:

(i′) There is a design $\xi_s \epsilon \Xi_q$. Find

$$\xi^s = \arg \min_\xi \int \psi(x, \xi_s)\xi(dx), \tag{34}$$

$$s.t. \int \zeta(x)\xi(dx) \leq c.$$

(ii′) Choose $0 \leq \alpha_s \leq 1$ and construct

$$\xi_{s+1} = (1 - \alpha_s)\xi_s + \alpha_s \xi^s \tag{35}$$

It is important to point out that similar to the unconstrained case the transition from (21) to (22) is still possible here, but the transition from (22) to (23) is not valid now. Therefore, one has to work with (31) directly, and that is much more complicated then to compute using (23). In (31) the optimization has to be made in the space of probability measures $\xi$, while in (23) one looks for a minimum over $X$.

Nevertheless, assume also that problem (31) can be solved. Then one can use all results of the previous section starting from Lemma 1. For instance, the following theorem takes place:

**THEOREM 7.** If (a)–($e_0$), (b$'$), (d$'$) hold then for $\{\alpha_s\}$ defined by rule (2) the iterative procedure (i$'$), (ii$'$) converges:

$$\lim_{s\to\infty} \Psi[M(\xi_s)] = \Psi^*.$$

The proof is identical to the proof of Theorem 5 with the obvious substitution

$$\min_x \psi(x,\xi_s) \text{ by } \int \psi(x,\xi_s)\xi_s(dx).$$

Thus one does not face any difficulties in optimal design construction if optimization problem (31) can be easily solved. Unfortunately, it is not the case in the majority of experimental situations. Some simplification can be obtained when one applies Note 1 (Section 1.3). Then (34) can be reduced to a finite dimensional problem:

$$\xi^s = \left\{ \begin{array}{c} x_1^s \ldots x_{l+1}^s \\ p_1^s \ldots p_{l+1}^s \end{array} \right\} = \arg \min_{p_j, x_j} \sum_{j=1}^{l+1} p_j \ \psi(x_j,\xi_s) \tag{36}$$

$$s.t. \sum_{j=1}^{l+1} p_j \phi(x_j) \le 0, \qquad \sum_{j=1}^{l+1} p_j = 1.$$

This problem can be considered as practically solvable when the number of constraints and the dimension of $X$ is reasonably small. Some numerical methods to solve (33) based on the cutting–plane technique are considered by Gaivoronski, 1986. They assume extra iterating within every s–th iteration, and the algorithm becomes impractical.

Obviously some corrections or amendments can be made to simplify the iterative procedure (i$'$), (ii$'$).

For instance, when an initial design $\xi_0$ is inside of $\Xi_q$, i.e.

$$\int \phi(x)\xi_0(dx) \le -\delta, \qquad \delta_\alpha > 0, \qquad \alpha = 1,\ldots,l \tag{37}$$

then the iterations (i), (ii) defined in the previous section can be used. Naturally one has continuously to verify that (34) is satisfied.

The convergence of the iterative procedure still takes place when instead of solution (31) or (33) one would look for any design $\xi^s$ such that

$$\int \psi(x,\xi_s)\xi^s(dx) \ge \gamma_s > 0 \text{ and } \int \psi(x)\xi^s(dx) \le 0,$$

where $\{\gamma_s\}$ is some diminishing sequence.

The equivalence theorem (Theorem 2) contains a hint for another modification of the iterative procedure. Particularly part 4 of this theorem states that the dual function

$$q(x, u, \xi) = \psi(x, \xi) + u^T \phi(x)$$

achieves zero almost everywhere in *supp* $\xi^*$.

Therefore at stage (i′) one can do the following:

There is a design $\xi_s \epsilon \Xi_q$. Let *supp* $\xi_s = \{x_1^s, \ldots, x_{n_s}^s\}$.

Step (a). Solve the linear programming:

$$u_s = \arg \max_u \; u_{l+1},$$

$$\psi(x_i^s, \xi_s) + u^T \phi(x_i^s) - u_{l+1} \geq 0, \; i = 1, \ldots, n_s$$

Step (b). Find

$$\xi^s = \xi(x_s) = \arg \min_x \quad q(x, u_s, \xi_s).$$

### 2.3 Linear Constraints. Lagrangian Approach.

On an intuitive level it is clear that minimizing function

$$\Phi(\xi) = \Psi(\xi) + \lambda^T C(\xi) \tag{38}$$

where $\lambda_\alpha > 0, \alpha = 1, \ldots, l$ and $C(\xi) = \int \phi(x)\xi(dx)$, one guaranties that $C(\xi) \leq c_\lambda$., where $c_\lambda$ become known after minimization of (35).

The problem consists of finding a vector $\lambda$ such that $c_\lambda$ is close to the origin, i.e. the $c_\lambda^T c_\lambda \simeq 0$. The following theorem can be useful for this purpose.

**THEOREM 8.** If (a)–($e_0$), (b′) hold, then a design:

$$\xi_\lambda = \arg \min_\xi [\Psi(\xi) + \lambda^T C(\xi)] \tag{39}$$

is a solution of the constrained optimization problem (7), (8) with

$$c_\lambda = C(\xi_\lambda).$$

**Proof.** Let $\bar{\xi} = (1 - \alpha)\xi + \alpha\bar{\xi}$ where $\xi \epsilon \Xi(q)$ and $\bar{\xi} \epsilon \Xi$. It is easy to check that

$$\lim_{\alpha \to 0} [\partial \Phi(\bar{\xi})/\partial \alpha] = \int \psi(x, \xi)\bar{\xi}(dx) + \lambda^T[\phi(x) - C(\xi)] = \int \tilde{\psi}(x, \lambda, \xi)\bar{\xi}(dx)$$

and therefore assumption (e) takes place for $\Phi(\xi)$ if it does for $\Psi(\xi)$. Similarly one can check the validity of all assumptions (a)–(d) for $\Phi(\xi)$ if they are valid for $\Psi(\xi)$. Thus all the conditions of the equivalence theorem for the unconstrained case (see Theorem 1) hold.

Thus the necessary and sufficient condition for a design $\xi_\lambda$ to be optimal (i.e. providing minimum for $\Psi(\xi)$) is fulfillment of the following inequality

$$\min_x \ \tilde{\psi}(x, \lambda, \xi_\lambda) \geq 0.$$

But $\tilde{\psi}(x, \lambda, \xi_\lambda)$ coincides with the function $q(x, u, \xi_\lambda)$ defined in Theorem 2 if one sets $u = \lambda$ and $\phi'(x) = \phi(x) - c_\lambda$ , $c_\lambda = C(\xi_\lambda)$. Thus it follows that

$$\min_x \ \max_u \ [\psi(x, \xi_\lambda) + u^T \phi'(x)] \geq$$

$$\geq \ \min_x \ [\psi(x, \xi_\lambda) + \lambda^T \phi'(x)] \geq 0$$

This inequality is sufficient to assert that

$$\xi_\lambda = \arg \ \min_\xi \ \Psi(\xi),$$

$$s.t. \ \int \phi(x)\xi(dx) \leq c_\lambda,$$

and it completes the proof.

Theorem 7 provides an approximate solution for (7), (8) if one manages to find out the vector $\lambda$ such that all components of vector $c_\lambda$ are small enough.

It seems not very much can be done here analytically, Cook & Wong, 1992, considered the case $\lambda \epsilon R^1$ in detail. The following approach looks very promising for most applications.

It will be assumed that for any given $\lambda$ problem (36) can be numerically solved at no significant expense. The likeliness of this assumption is confirmed by the experience reported by various authors (see Fedorov, 1975 & 1986, Mitchell, 1974, Nachtsheim, 1987) who had experimented with numerous modifications of the algorithms considered in section 2.1. Additionally we shall assume that a practitioner will be satisfied if the vector $c_\lambda = C(\xi_\lambda)$ would be “close” to 0 (but not exactly equal), for instance, a squared distance

$$y_\lambda = c_\lambda^T A c_\lambda \ , \ A > 0$$

would be small.

Thus, one can consider the “empirical” optimization problem:

$$\lambda^* = \arg \ \min_\lambda \ c_\lambda^T A c_\lambda, \tag{40}$$

$$s.t. \ \lambda_\alpha \geq 0 \ , \ \alpha = 1, \ldots, l,$$

Considering $y_\lambda$ as a response function depending upon $\lambda$, one can apply to the empirical optimization technique which is the famous milestone in the experimental design.

## 2.4 Nonlinear Constraints

All the basic results of sections 2.2 and 2.3 can be generalized for the case with nonlinear constraints, see (11), (12).

Assumption (e$'$) must be replaced by the more restrictive one:

$$\Phi[\underline{\xi}(\alpha)] = \Phi(\xi) + \alpha \int \phi(x,\xi)\bar{\xi}(dx) + \gamma[\underline{\xi}(\alpha)],$$

where $\underline{\xi}(\alpha) = (1-\alpha)\xi + \alpha\bar{\xi}, \xi\epsilon\Xi_q, \underline{\xi}\epsilon\Xi$, and $\gamma[\underline{\xi}(\alpha)] \leq \alpha^2 Q_q, Q_q > 0$ uniformly for all $\underline{\xi}(\alpha)\epsilon\Xi(q)$.

By linearizing $\Phi(\xi)$ in the "vicinity" of optimal design (compare with Section 1.4) one can check that all results of Section 2.2 stay valid if the vector above defined $\phi(x,\xi)$ replaces everywhere the corresponding vector defined in the comment to (8).

The approach proposed in Section 2.3 looks less elegant from a mathematical point of view, but it is more practical. In this case instead of the original problem (11), (12) one has to consider a sequence of optimization problems defined by (37) for a "compound" criteria of optimality:

$$\xi_\lambda = \arg \min_\xi [\Psi(\xi) + \lambda^T \Phi(\xi)],$$

$$\lambda_\alpha > 0, \alpha = 1, \ldots, l,$$

where very frequently $\Phi(\xi)$ is a vector of auxiliary optimality criteria (see Section 1.4).

## 2.5 Nonlinear Response Function

Together with the equivalence theorem (Theorem 3) for infinitely many constraints the approach from Section 2.3 gives a new insight on the problem of optimal experimental design for the nonlinear parameterization case.

Let

$$M(\xi,\theta) = \int f(x,\theta) f^T(x,\theta)\xi(dx),$$

where $f(x,\theta) = \partial\eta(x,\theta)/\partial\theta$ and $\theta\epsilon\Omega \subset R^l$.

If one wishes to ensure that for any $\theta$ some optimality criteria does not exceed a given level, then the following optimization problem can be considered:

$$\xi^* = \arg \min_\xi \ \Psi[M(\xi,\theta_0)], \tag{41}$$

$$s.t. \ \Psi[M(\xi,\theta)] \leq c \quad \text{for all } \theta\epsilon\Omega,$$

where vector $\theta_o$ could be, for instance, some prior guess about the values of the parameters to be estimated.

From Theorem 3 it follows that there exist $(l+1)$ vectors $\theta_k^*$ such that (38) can be reduced to the following finite dimension problem:

$$\xi^* = \arg \min_{\xi} \ \Psi[M(\xi, \theta_0)], \tag{42}$$

$$s.t. \ \Psi[M(\xi, \theta_k^*)] \leq c, \quad k = 1, \ldots, l+1$$

Theorem 3 tells us that the "compound" optimization problem (see the two previous sections):

$$\xi_\lambda^* = \arg \min_{\xi} \ \Psi[M(\xi, \theta_o)] + \sum_{k=1}^{l+1} \lambda_k \ \Psi[M(\xi, \theta_k^*)]$$

is closely related to (38). If one knows $\theta_k^*$ then all the results of section 2.3 can be used without any changes. Unfortunately $\theta_k^*$ must often be searched for and the authors failed to find any reasonable numerical procedure to do this. On the common sense level it is clear that $\theta_k^*$ ought to be some "worst" point in $\Omega$ corresponding to a "smallest" information matrices.

If there is no prior preference to any point from $\Omega$ one may consider another very similar problem instead of (38):

$$\xi^* = \arg \min_{\xi} c, \tag{43}$$

$$s.t. \ \Psi[M(\xi, \theta_k^*)] \leq c, k = 1, \ldots, l+1,$$

which, of course, can be considered as a particular case of (11), (12). For instance, (40) can be solved by multiple application to the solution of the "compound" optimization problem:

$$\xi^* = \arg \min_{\xi} \sum_{k=1}^{l+1} \lambda_k \Psi[M(\xi, \theta_k^*)],$$

where again no strict recipe how to find out set of $\theta_k^*, k = 1, \ldots, l+1$ are known to the authors.

We can conclude this section mentioning that (40) is equivalent to the following minimax problem:

$$\xi^* = \arg \min_{\xi} \ \max_{\theta} \ \Psi[M(\xi, \theta)].$$

## References

[1] Atwood, C.L. (1973). Sequences Converging to D–optimal Designs of Experiments. *Ann. Stat.*, **1**, 342–352.

[2] Atkinson, G.L. and Donev, A.N. (1992). *Optimum Experimental Designs*, Clarendon Press, Oxford, 327.

[3] Atkinson, A.C. and Fedorov, V.V. (1975). The Design of Experiments for Discriminating Between Two Rival Models. *Biometrika*, **62** 57–70.

[4] Atkinson, A.C. and Fedorov, V.V. (1988). The Optimum Design of Experiments in the Presence of Uncontrolled Variability and Prior Information. *Optimal Design and Analysis of Experiments*, Ed. Y. Dodge, V. Fedorov, H. Wynn, North-Holland, 327–344.

[5] Chernoff, H. (1972). Sequential Analysis and Optimal Design. *SIAM*, Philadelphia.

[6] Cook, R.D. and Nachtsheim, C.J. (1989). Computer–aided Blocking for Factorial and Response–Surface Designs. *Technometrics*, **31**, 339–346.

[7] Cook, R. D. and Thibodeau, L.A. (1980). Marginally Restricted D–optimal Designs. *JASA*, **75**, 366–371.

[8] Cook, R.D. and Wong, W.K. On the Equivalence of Constrained and Compound Optimal Designs. Submitted for publication.

[9] Ermakov, S.M. (1983). *Mathematical Theory of the Design of Experiments* (in Russian), Nauka, 386.

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

[11] Fedorov, V.V. (1986). Optimal Design of Experiments: Numerical Methods. WP–86–55, IIASA, Laxenburg, Austria, 10.

[12] Fedorov, V.V. (1989). Optimal Design with Bounded Density: Optimization Algorithms of the Exchange Type. *J.S.P.I.*, **22**, 1–13.

[13] Fedorov, V.V. and Gaivoronski, A. (1984). Design of Experiments Under Constraints, International Institute for Applied System Analysis. WP–84-8, Austria, Laxenburg.

[14] Fedorov, V.V. and Khabarov, V.I. (1986). Duality of Optimal Designs for Model Discrimination and Parameter Estimation. *Biometrika*, **73**, 191– 199.

[15] Fedorov, V.V. and Malyutov, A.B. (1972). Design of Experiments in Regression Problems. *Mathemat. Operationsforschung und Statist.*, **3**, 281–321.

[16] Fedorov, V.V. and Upensky, A.B. (1975). Numerical Aspects of Design and Analysis of Experiments. Moscow State University, Moscow, 167.

[17] Huang, M.L. and Hsu, M.C. (1993). Marginally Restricted Linear–optimal Designs. *JSPI*, **35**, 251–266.

[18] Gaivoronski, A. (1986). Linearization Methods for Optimization of Functionals Which Depend on Probability Measures. *Mathematical Programming Study*, **28**, 157–181.

[19] S. Karlin and W.J. Studden (1966). *Tchebycheff Systems: with Applications in Analysis and Statistics*, Wiley and Sons, New York.

[20] Kiefer, J. (1959). Optimal Experimental Design. *Journal of the Royal Statistical Society*, B, **21**, 272–319.

[21] Krein, M. and Nudelman, A. (1977). The Markov Moment Problem and Extremal Problems. *Translation Mathem. Monographs 50*, American Mathematical Society, Providence, EI.

[22] Laurent, P.J. (1972). Approximation et Optimization, Hermann, Paris.

[23] Lee, C.M.–S. (1988). Constrained Optimal Designs. *J.S.P.I.*, **18**, 377–389.

[24] Läter, E. (1976). Optimal Multipurpose Designs for Regression Models. *Math. Operationsforschung und Statist.*, **7**, 51–68.

[25] Mitchell, T.J. (1974). An Algorithm for the Construction of D–optimal Experimental Designs. *Technometrics*, **16**, 203–211.

[26] Nachtsheim, C.J. (1987). Tools for Computer– aided Design of Experiments. *Journal of Quality Technology*, **19**, 132–160.

[27] Pshenichnyi, B.N. (1971). *Necessary Conditions for an Extremum*, Marcel Dekker, New York.

[28] Pukelsheim, F. (1980). On Linear Regression Designs Which Maximize Information. *J.S.P.I.*, **4**, 339–364.

[29] Silvey, S.D. (1980). *Optimal Design*, Chapman and Hall, London, 1980.

[30] Whittle, P. (1973). Some General Points in the Theory of Optimal Experimental Design. *Journal of the Royal Statistical Society*, Ser. B, **35**, 123–130.

[31] Wu, C.F. and Wynn, H. (1978). The Convergence of General Step–Length Algorithms for Regular Optimum Design Criteria. *Ann. Statist.*, **6**, 1273–1285.

[32] Wynn, H. (1982). Optimum Submeasures with Applications to Finite Population Sampling. *Statistical Decision Theory and Related TopicsIII*, Vol. 2, Academic Press, New York, 485–495.