Working Paper

The Experimental Design of an Observational Network: Optimization Algorithms of the Exchange Type

V. V. Fedorov

October 1986
WP-86-62

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

The Experimental Design of an Observational Network: Optimization Algorithms of the Exchange Type

V. V. Fedorov

October 1986
WP-86-62

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

Preface

For many years, designers of environmental monitoring systems have faced the problem of optimal allocation of resources for observational networks (see, for instance, Munn, 1981): where, how frequently and what characteristics have to be measured or observed in order to obtain data that will be sufficient for prognoses or warnings. From the 1970's to the early 1980's, a number of heuristic approaches appeared in the "environmental" literature. Most of them are based on the analysis of space and time correlation structures (usually historical time series are used for their estimation) of the observed entities with subsequent sieving to keep the less correlated (and hopefully most informative) observational points.

Different procedures for "sieving" have been used in applications: viz., forward and backward versions with various objective functions. These procedures have led to reasonably good results; however, no accurate mathematical analysis were undertaken.

In this present paper, a new approach to optimal allocation of an observational network is proposed and some iterative numerical procedures are considered. The approach is essentially based on the theory of the optimal design of regression experiments (Ermakov, ed., 1983). Using the classical results from the moment spaces theory, the author investigates the properties of optimal allocations and the convergence of the numerical procedures to optimal solutions.

Prof. M. Antonovsky
(Environment Program)

The Experimental Design of an Observational Network: Optimization Algorithms of the Exchange Type

V. V. Fedorov

1. Introduction

In this paper the numerical procedures of the "exchange" type for construction of continuous optimal designs with restricted measures (see definitions in Fedorov, 1986, Wynn, 1982) were considered. The "exchange" type procedures were based on the simple heuristic idea: at every subsequent step to delete "bad" (less informative) points and to include "good" (most informative) ones.

Before giving the accurate mathematical formulation of the problem and to illuminate the place of the results in experimental practice, let us start with two simple hypothetical examples. "Real" examples, where the considered approach seems to be appropriate can be found, for instance, in Munn, 1981.

Example 1. Let X be an area where N observational stations have to be located. An optimal (or at least, admissible) location depends upon models describing a system: "object under analysis – observational techniques".

The regression models:

$$y_i = \eta(x_i, \vartheta) + \varepsilon_i, i = \overline{1,N} \tag{1}$$

are commonly used in experimental practice. Here y_i is a result of an observation of the i–th station, $\eta(x, \vartheta)$ is an a priori given function, ϑ is a vector of parameters to be estimated and ε_i is an error which one believes to be random (more detailed specification will be given later). The optimal location of stations has to provide the minimum of some measure of deviation of estimates $\hat{\vartheta}$ from true values of ϑ.

For sufficiently large N the location of stations can be approximately described by some distribution function $\xi(dx)$ and one needs to find an optimal $\xi^*(dx)$. If X is not uniform, then one comes to the restriction that the share $N(\Delta X)/N$ of stations in any given part ΔX cannot exceed some prescribed level. In terms of distribution functions, it means that

$$\xi(dx) \leq \Psi(dx), \tag{2}$$

where Ψ is defined by an experimenter. Here is the crucial feature of the problem considered in this paper.

Example 2. Let some characteristic y_i be observed for members of a sample of size N. Every i–th member of this sample can be chosen from a group labelled by variables x_i. If the sampling is randomized, then the observed characteristic y_i can be described by some distribution $f(y / x_i, \vartheta)$.

In many cases, after some manipulations, the initial model can be reduced to (1), where $\eta(x_i, \vartheta)$ is an average characteristic of an i–th group and ε_i reflects a variation within this group. The size of any group (or number of units available for sampling) is normally bounded. When applied to a continuous version of the

design problem one can easily repeat the considerations of the previous example and come to model (1), (2).

In what follows, it will be assumed that in model (1), (2):

- a response function is a linear function of unknown parameters, i.e. $\eta(x,\vartheta)=\vartheta^T f(x)$, $\vartheta\in R^m$ and functions $f(x)$ are given;
- errors ε_i are independent and $E[\varepsilon_i^2]=1$ (or $E[\varepsilon_i^2]=\lambda(x_i)$, where $\lambda(x)$ is known, this case can be easily transformed to the previous one).

As usual, some objective function Φ defined on the space of $m\times m$ information matrices

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

will describe the quality (or accuracy) of a design ξ ($M^{-1}(\xi)$ as a normalized variance–covariance matrix of the least square estimators of parameters ϑ.

The purpose of optimum design of experiments is to find

$$\xi^*=\inf_{\xi}\ \Phi[M(\xi)]\ ,\ \int_X \xi(dx)=1, \tag{3}$$

$$\xi(dx)\le\Psi(dx)\ ,\ \int_X \Psi(dx)=Q\ge 1 \tag{4}$$

Constraint (4) defines the peculiarity of the design problem with respect to standard approaches. Similar to the moment spaces theory (compare with Krein and Nudelmann, 1973 Ch. VII), a solution of (3) and (4) will be called "(Φ , Ψ)–optimal design". In practice, $\Psi(dx)$ restricts the number of observations in a given space element dx (see the examples).

Optimization problem (1) and (2) were considered by Wynn, 1982 and Gaivoronsky, 1985. To some extent, they translated a number of classical results from moment spaces theory to experimental design language. Gaivoronsky also analyzed the convergence of the iterative procedure for optimal design construction based on the traditional idea of steepest descent (see, for instance, Ermakov (ed), 1983, Wu and Wynn, 1976)

$$\xi_{s+1}=(1-\alpha_s)\xi_s+\alpha_s\bar{\xi}_s\ , \tag{5}$$

$$\bar{\xi}_s=Arg\min_{\xi}\Phi[(1-\alpha_s)M(\xi_s)+\alpha M(\xi)],$$

where ξ has to satisfy (4) and some additional linear constraints:

$$\int_X q(x)\xi(dx)\le C\ .$$

Wynn briefly discussed a number of heuristic numerical procedures based on some results from the moment spaces theory.

The main objective of this paper is to consider the iterative procedures of exchange type which extensively use the nature of optimal designs for problem (3), (4) and therefore promises to be more efficient than the ones mentioned above.

General properties of optimal designs are discussed in Section 2. Section 3 deals with formulation and basic analysis of the iterative procedure and its modifications.

2. Characterization of (Φ , Ψ)–optimal Designs

In this section, the properties of optimal designs will be discussed only to the extent sufficient for the analysis of the proposed iterative procedures. More details can be found in Wynn, 1982.

The set of assumptions used later is the following:

a) X is compact , $X \in R^l$;

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

c) $\Psi(x)$ is atomless;

d) there exists $c < \infty$ such that

$$\Xi_c(\Psi) = \{\xi : \Phi[M(\xi)] \leq c < \infty , \xi \in \Xi(\Psi)\} \neq \phi ,$$

where $\Xi(\Psi)$ is the set of designs satisfying (4);

e) $\Phi(M)$ is a convex function of M ;

f) $\Phi[(1-\alpha)M(\xi_1) + \alpha M(\xi_2)] = \Phi[M(\xi_1)] + \alpha \int_X \varphi(x, \xi_1)$

$\xi_2(dx) + o(\alpha)$, $\xi_1 , \xi_2 \in \Xi_c(\Psi) , c < \infty$;

c′) $\Psi(x)$ has a continuous density $\psi(x)$;

f′) derivatives $\frac{\partial \Phi}{\partial M} = \dot{\Phi}$ exist and are bounded for all designs satisfying (d).

Let $\overline{\Xi}(\Psi)$ to be a set of measures ξ which either coincide with Ψ or equal to 0.

Theorem 1. If assumptions (a) – (e) hold, then there exists an optimal design $\xi^* \in \overline{\Xi}(\Psi)$.

Proof. The existence of an optimal design follows from (d)–(e) and the compactness of the set of information matrices. The compactness of the latter is provided by (a) and (b). The fact that at least one optimal design has to belong $\overline{\Xi}(\Psi)$ is the corollary of Liapounov's Theorem on the range of a vector measure (see, for instance, Karlin and Studden, 1966, Ch. VIII, Wynn, 1982).

Note 1. Liapounov's Theorem leads to another result which can be useful in applications: for any design ξ there is a design $\bar{\xi} \in \overline{\Xi}(\Psi)$ such that $M(\xi) = M(\bar{\xi})$.

A function $\varphi(x, \xi)$ is said to separate sets X_1 and X_2 if there is a constant C such that $\varphi(x, \xi) \leq C$ $(a.e.\ \Psi)$ on X_1 and $\varphi(x, \xi) \geq C$ $(a.e.\ \Psi)$ on X_2 , $(a.e.\ \Psi)$ means "almost everywhere with respect to the measure Ψ".

Theorem 2. If assumptions (a)–(f) hold, then a necessary and sufficient condition that $\xi^* \in \Xi(\Psi)$ *is* (Φ, Ψ)*–optimal is that* $\varphi(x, \xi^*)$ *separates two sets:* $X^* = supp\, \xi^*$ *and* $X \setminus X^*$.

This theorem was first formulated by Wynn, 1982; but its proof was not perfect. Therefore, we give the newer one which is also more illuminative for the formulation and analysis of the numerical procedures.

Proof. Necessity.

Consider two designs: ξ^* and $\xi \in \overline{\Xi}(\Psi)$. Let

$$supp\, \xi^* = X^* \subset X \ , \ supp\, \xi = X \setminus D \cup E \ ,$$

$$E, D \subset X, E \cap D = \phi, \tag{7}$$

$$\int_D \Psi(dx) = \int_E \Psi(dx) = \delta > 0.$$

Assume that ξ^* is (Φ,Ψ)–optimal. Then for any design ξ (see (f)):

$$0 \le \int_X \varphi(x,\xi^*)\ \xi(dx)$$

$$= \int_{X^*} \varphi(x,\xi^*)\Psi(dx) + \int_E \varphi(x,\xi^*)\ \Psi(dx) - \int_D \varphi(x,\xi^*)\ \Psi(dx)\ .$$

From the definition of $\varphi(x,\xi)$:

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

and, therefore, for any E and D :

$$\int_E \varphi(x,\xi^*)\ \Psi(dx) \ge \int_D \varphi(x,\xi^*)\ \Psi(dx)\ . \tag{8}$$

This proves necessity.

Sufficiency. Consider designs ξ^* and $\xi \in \bar{\Xi}(\Psi)$ satisfying to (7) and (8) and assume now that ξ^* is nonoptimal, i.e.

$$\Phi\ [M(\xi^*)] \ge \Phi\ [M(\xi)] + \delta\ ,\ \ \delta > 0. \tag{9}$$

Let $\bar{\xi} = (1-\alpha)\xi^* + \alpha\ \xi$, $\alpha > 0$ and ξ is now $(\Phi\ ,\ \Psi)$–optimal. Then, the convexity of Φ leads then to the inequality:

$$\Phi\ [M(\bar{\xi})] \le (1-\alpha)\ \Phi\ [M(\xi^*)] + \alpha\ \Phi\ [M(\xi)] \tag{10}$$

$$\le (1-\alpha)\ \ \Phi\ [M(\xi^*)] + \alpha\ \{\Phi[M(\xi^*)] - \delta\}\ =\ \Phi\ [M(\xi^*)] - \alpha\ \delta\ .$$

Assumption (f) and inequality (8) lead to the inequality

$$\Phi\ [M(\bar{\xi})] = \Phi\ [M(\xi^*)] + \alpha \int_X \varphi(x,\xi^*)\ \ \xi(dx) + o(\alpha)\ = \tag{11}$$

$$= \Phi\ [M(\xi^*)] + \alpha\ \{\int_E \varphi(x,\xi^*)\ \ \Psi(dx) - \int_D \varphi\ (x,\xi^*)\ \ \Psi\ (dx)\}\ +\ o(\alpha)$$

$$\ge \Phi\ [M(\xi^*)]\ +\ o\ (\alpha)$$

where E and D describe the difference between the supporting sets for ξ^* and ξ.

When $\alpha \to 0$, the comparison (10) and (11) gives a contradiction. This completes the proof.

Note 1. If instead of (c), one uses (c'), then a necessary and sufficient condition can be formulated in the form of the following inequality:

$$\max_{x \in X^*}\ \varphi\ (x,\xi^*) \le \min_{x \in X \setminus X^*}\ \varphi\ (x,\xi^*) \tag{12}$$

Note 2. If (f) is complemented by (f), then

$$\varphi\ (x,\xi) = \gamma\ (x,\xi) - tr\ \ \dot{\Phi}\ (\xi)\ \ M(\xi)\ ,$$

where $\delta(x,\xi)\ =\ f^T(x)\ \ \dot{\Phi}\ (\xi)\ \ f(x)$, and (12) can be converted to

$$\max_{x \in X^*}\ \gamma\ (x,\xi^*) \le \min_{x \in X \setminus X^*}\ \gamma\ (x,\xi^*) \tag{12'}$$

3. Numerical Procedure of Exchange Type

Theorem 2 gives a hint on how to construct optimal designs numerically: if for some given design ξ one can find a couple of sets:

$$D \subset supp\ \xi \text{ and } E \subset X \setminus supp\ \xi\ ,$$
$$\int_D \varphi(x,\xi)\ \Psi(dx) > \int_E \varphi(x,\xi)\ \Psi(dx)\ , \tag{13}$$
$$\int_D \Psi(dx) = \int_E \Psi(dx)$$

then it is hoped that the design $\bar{\xi}$ with

$$supp\ \bar{\xi} = supp\ \xi \setminus D \cup E$$

will be "better" than ξ. The repetitions of this procedure can lead to an optimal design.

A number of algorithms based on this idea can be easily invented. In this paper one of the simplest algorithms is considered in detail and it is evident that thorough consideration of others from this cluster is routine technique.

In what follows, the fulfillment of (c') is assumed.

Algorithm. Let

$$\lim_{s\to\infty} \delta_s = 0\ ,\ \lim_{s\to\infty} \sum_{s=1}^{\infty} \delta_s = \infty \text{ and } \lim_{s\to\infty} \sum_{s} \delta_s^2 = k < \infty\ . \tag{14}$$

Step a. There is a design $\xi_s \in \bar{\Xi}(\Psi)$. Two sets D_s and E_s with equal measures:

$$\int_{D_s} \psi(x)\,dx = \int_{E_s} \psi(x)\,dx = \delta_s$$

and including, correspondingly, points:

$$x_{1s} = Arg \max_{x \in X_{1s}} \delta(x,\xi_s) \text{ and } x_{2s} = Arg \min_{x \in X_{2s}} \delta(x,\xi_s)\ , \tag{15}$$

where $X_{1s} = supp\ \xi_s$ and $X_{2s} = X \setminus X_{1s}$, have to be found.

Step b. The design ξ_{s+1} with the supporting set

$$supp\ \xi_{s+1} = X_{1(s+1)} = X_{1s} \setminus D_s \cup E_s \tag{16}$$

is constructed.

Iterative procedure (14)–(16) is based on the approximation ($\delta \to 0$):

$$\int_Q \varphi(x,\xi)\ \psi(x)\ dx \approx \varphi(\bar{x},\xi)\ \delta\ ,\ \bar{x} \in Q\ ,\ \int_Q \psi(x)\,dx = \delta\ ,$$

The analysis of iterative procedure (14)–(16) becomes simpler if

(g) for any design $\xi \in \bar{\Xi}(\Psi)$:

$$|M(\xi)| \geq \zeta > 0\ .$$

This assumption is not very restrictive. If, for instance, $\psi(x) \geq q > 0$ and the functions $f(x)$ are linearly independent on any open finite measure subset of X, then (g) is valid.

Most optimality criteria (g) lead to the fulfillment of the following inequalities:

$$\Phi \leq K_1 < \infty\ ,\ \frac{\partial \Phi}{\partial M_{\alpha\beta}} \leq K_2 < \infty\ ,\ \frac{\partial^2 \Phi}{\partial M_{\alpha\beta}\,\partial M_{\alpha\delta}} \leq K_3 < \infty\ , \tag{17}$$

$$1 \le \alpha , \beta , \delta , \delta \le M ,$$

for any $\xi \in \bar{\Xi}(\Psi)$. Otherwise (17) is supposed to be included in (g).

Theorem 3. If assumptions (a), (b), (c'), (e)–(g) hold, then

$$\lim_{s \to \infty} \Phi [M(\xi_s)] = \inf_{\xi} \Phi [M(\xi)] = \Phi^* \tag{18}$$

Proof. The approach is standard for optimization theory (in the statistical literature see, for instance, Wu and Wynn, 1978). Therefore, some elementary considerations will be omitted.

Expanding (see (g) and (17)) by a Taylor series in δ_s gives:

$$\Phi[M(\xi_{s+1})] = \Phi[M(\xi_s)] + \delta_s [\gamma(x_{2s} , \xi_s) - \gamma (x_{1s} , \xi_s)] + \delta_s^2 K_s , \tag{19}$$

where $|K_s| \le K_0 = K_0(K_1, K_2, K_3)$. Due to this inequality and (14) the sequence $S_{2s} = \{\sum_s K_s \delta_s^2\}$ converges. By definition:

$$\gamma (x_{2s} , \xi_s) - \gamma (x_{1s} , \xi_s) \le 0 ,$$

and, therefore, the sequence:

$$S_{1s} = \sum_s \delta_s [\gamma(x_{2s} , \xi_s) - \gamma (x_{1s} , \xi_s)]$$

monotonically decreases.

From (g) and (19):

$$K_1 \ge \Phi[M(\xi_{2+1})] = \Phi[M(\xi_0)] + S_{1s} + S_{2s} \ge \Phi^*$$

leads to the boundness of S_{1s}.

Subsequently, the monotonicity of $\{S_{1s}\}$ provides its convergence and the convergence of $\{\Phi[M(\xi_s)]\}$. Assume that

$$\lim_{s \to \infty} \Phi [M(\xi_s)] = \Phi^1 \ge \Phi^* + a , a > 0 . \tag{20}$$

Then, from Theorem 2 and assumptions (b), (c') it follows that

$$\gamma(x_{2s} , \xi_s) - \gamma(x_{1s} , \xi_s) \le b < 0$$

and

$$\lim_{s \to \infty} S_{1s} \le b \lim_{s \to \infty} \sum_s \delta_s = -\infty , \tag{21}$$

$$\lim_{s \to \infty} \Phi [M(\xi_s)] \le -\infty .$$

The contradiction between (20) and (21) proves the theorem.

Note 1. In (14)–(16), there is some uncertainty in the choice of D_s and E_s. Somehow, they have to be located around x_{1s} and x_{2s}. When $\psi(x) = \text{const}$ (and one arrives at this case by the transformation $dx = \psi(x)dx$), then x_{1s} and x_{2s} could be the "geometrical" centers of D_s and E_s.

Note 2. The iterative procedure can be more effective (especially in the first steps) if there is a possibility to easily find

$$D_s^* = Arg \max_{D_s \in X_{1s}} \int_{D_s} \gamma (x, \xi_s) \psi (x) dx$$

and (22)

$$E_s^* = Arg \min_{E_s \subset X_{2s}} \int_{E_s} \gamma (x, \xi_s) \psi(x) dx$$

subject to

$$\int_{D_s} \psi(x)\,dx = \int_{E_s} \psi(x)\,dx = \delta_s \ . \tag{23}$$

Note 3. When δ_s is sufficiently small and

$$\int_D f(x) f^T(x) \psi(x)\,dx \approx f(x_{1s}) f^T(x_{1s})\,\delta_s \ ,$$

$$\int_E f(x) f(x) \psi(x)\,dx = f(x_{2s})\ f^T(x_{2s})\,\delta_s$$

then, the calculations in (14)–(16) can be simplified if one use the following recursion formula (see, for instance, Fedorov, 1972)

$$(M \pm \delta f f^T)^{-1} = (I \mp \frac{\delta M^{-1} f f^T}{1 \pm \delta f^T M^{-1} f}) M^{-1}$$

$$= (I \mp \delta M^{-1} f f^T) M^{-1} + o(\delta^2) \ .$$

The modified version of the algorithms, presented in Note 2, gives a hint for the construction of

Algorithm 2.

Step a. The same as (22), but instead of (23)

$$\int_{D_s} \psi(x)\,dx = \int_{E_s} \psi(x)\,dx \tag{24}$$

(no constraints on the sizes of D_s and E_s!).

Step b. Coincides with step b of algorithm 1.

This algorithm seems to be rather promising for changing the structure of an initial design ξ_0 rapidly, but it allows some oscillation regimes, at least principally. The author failed to prove its convergence. Probably some combination of both considered algorithms (for instance, the majorization of (24) by some vanishing sequence δ_s) could be useful.

4. Exchange algorithm in the standard design problem.

The possibility of changing the algorithms similar to (14)–(16) for design problem (3) (without constraint (4)) was somehow overlooked in the design theory. Atwood (1973) proposed a very similar algorithm but based on (5) and therefore handling all supporting points in design ξ_s.

The simplest analogue of (14)–(16) can be formulated as follows:

Step a. There is a design ξ_s. Two points

$$x_{1s} = Arg\max_{x \in X_s} \varphi(x, \xi_s) \text{ and } x_{2s} = Arg\min_{x \in X} \varphi(x, \xi_s) \ , \tag{25}$$

where $X_s = suppX$, have to be found.

Step b.

$$\xi_{s+1} = \xi_s - \delta_s\ \xi(x_{1s}) + \delta_s\ \xi(x_{2s}) \ , \tag{26}$$

where $\xi(x)$ is a design with one supporting point x.

The sequence $\{\delta_s\}$ can be chosen as in (14). The convergence of the algorithm can be proven similarly to Theorem 3.

It is worthwhile noting that the convergence of procedures (25), (26), in the discrete case (when $\delta_s = K/N$, $\geq N^{-1}$, where N is the total number of observations) is questionable, because proof of Theorem 3 is essentially based on the fact that $\delta_s \to 0$.

REFERENCES

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

Ermakov, S. ed. (1983) Mathematical Theory of the Design of Experiments (in Russian), Moscow, Nauka, p. 386.

Fedorov, V. (1986) Optimal Design of Experiments: Numerical Methods, WP–86–55, Laxenburg, Austria, International Institute for Applied Systems Analysis.

Gaivoronsky, A. (1985) Stochastic Optimization Techniques for Finding Optimal Submeasures, WP–85–28, Laxenburg, Austria, International Institute for Applied Systems Analysis.

Karlin, S. and W.J. Studden (1966) Tchebycheff Systems: With Applications in Analysis and Statistics. New York: Wiley & Sons, p. 586.

Krein, M.G. and A.A. Nudelman (1973) Markov Moment Problem and Extremal Problems, Moscow, Nauka, p. 552.

Munn, R.E. (1981) The Design of Air Quality Monitoring Networks, Macmillan Publishers LTD, London, p. 109.

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

Wynn, H. (1982) Optimum Submeasures With Applications To Finite Population Sampling in "Statistical Decision Theory and Related Topics III", **2**, Academic Press, New York, pp. 485–495.