COMPUTER AIDED DESIGN OF EXPERIMENTS

By

G. K. KRUG*

Department of Automation, Energetic University, Moscow

(Received January 9, 1975)

Introduction

A mathematical description of processes helps to reveal the optimum conditions of their behaviour. A successful solution is largely dependent on the accuracy of the mathematical description available to engineers. In recent years, the experimental statistical methods of mathematical simulation have increased in popularity. The experimental designs which determine the research programs are based on different criteria of optimality. Orthogonal designs optimal in terms of simplicity of data reduction and rotatable designs that yield the same information at equal distances from the center are widely used.

For these designs, however, the optimality criterion is not sufficiently general. Furthermore, though these designs are compositional, their separate units have a rigid structure.

Therefore it is desirable to develop an identification method that would be sequential and based on designs that lead to minimization of the generalized dispersion or the dissipation ellipsoid volume for estimates of model coefficients.

The expansion of methods of experimental design on dynamic problems has a great interest. The problem is concluded in synthesis of such an input testing sequence that provides the best estimations of coefficients of pulse transient response decomposition.

An extrapolation problem arises when the investigator uses regression equation for predicting an objective function at a point (field) situated out of variation field, experimental extrapolation design must provide minimum of prediction variance.

Continuous designs that satisfy the above criteria will be considered below. Nowadays there are no algorithmical methods of synthesis of similar designs except for the simplest cases.

* On the basis of lecture read at the scientific session held on the 25th anniversary of Electrical Faculty of Technical University of Budapest.

The most efficient way of this problem solution is adoption of special calculating recurrent procedures permitting to synthesize desirable computer aided designs.

Algorithms

Suppose the functional form of the regression equation to be known

$$\eta_i = \sum_{j=1}^{K} \Theta_j f_{ij}(\mathbf{x}) \tag{1}$$

where $f_{ij}(\mathbf{x})$ are known functions of the input parameters $x_1, x_2, \ldots, x_m$. The vector $\mathbf{x} = (x_1, x_2, \ldots x_m)$ is a vector of the input parameters.

Random disturbances result in the magnitude observed by the experimenter

$$y_i = \eta_i + E_i \tag{2}$$

assumed to be distributed by the normal law with the mathematical expectation $M\{y_i\} = \eta_i$ and the dispersion σ^2.

The experimenter has to find the estimates of the coefficients in Eq. (1).

Introducing the notation: $\mathbf{f}^T = (f_{i1}, f_{i2}, \ldots f_{iK})$ a vector that determines the set of the functions f_{ij} in the *i-th* observation; $\hat{\Theta} = (\hat{\Theta}_1, \hat{\Theta}_2, \ldots \hat{\Theta}_K)$ is a k-dimensional vector of the desired estimates; $\mathbf{y}^T = (y_1, y_2, \ldots y_N)$ is an N-dimensional vector of observations.

The regression equation coefficient estimates are determined from the set of normal equations:

$$G\hat{\Theta} = F^T F \hat{\Theta} = F^T \mathbf{y}. \tag{3}$$

The solution to this set is

$$\hat{\Theta} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{y} = \mathbf{C}\mathbf{F}^T \mathbf{y} = \mathbf{G}^{-1}\mathbf{F}^T \mathbf{y}. \tag{4}$$

For a D-optimum design $\mathbf{F}_*$ relationship (4) is valid:

$$\det [\mathbf{F}_*^T \mathbf{F}_*]^{-1} = \min \det [\mathbf{F}^T \mathbf{F}]^{-1}, \tag{5}$$

i. e. the covariant matrix determinant minimal. The dispersion of the esti-mated regression function prediction is:

$$d_* = \max_{\mathbf{x}} . [\mathbf{f}^T(\mathbf{x})\, \mathbf{G}_*^{-1} \mathbf{f}(\mathbf{x})] = \frac{K}{N}. \tag{6}$$

Also, the D-optimum designs have been shown to be invariant to changes in the scale of independent variables.

The quantity

$$\delta_K = \frac{Nd_* - K}{K} \tag{7}$$

can be taken as the index of the difference between the design in question and the D-optimum design. The number of experimental points in the design

$$h \leq \frac{K(K+1)}{2}.$$

Approximate or continuous D-optimum plans are completely determined by setting a finite number of the points in the design space and observation repetition frequency in these points. Thus they are not designs with a fixed number of observations. The number of observations is selected by the experimenter regardless of the design structure so that the observation repetition frequency be as close as possible to the value specified by the D-optimum design.

Certain methods to calculated D-optimum designs in particular cases were described in [2]. A more general approach is to be presented with a continuous design of an experiment [1]. This method locates the point of the maximum information on the process at each stage of continuous experimentation with the calculations by recurrent formulae [1]:

$$\mathbf{f}_*^T(\mathbf{x}_*)\, \mathbf{C}(N)\, \mathbf{f}_*(\mathbf{x}_*) = \max_{\vec{\mathbf{x}} \in X} \mathbf{f}(\mathbf{x})\, \mathbf{C}(N)\, \mathbf{f}(\bar{\mathbf{x}}) \tag{8}$$

$$g_{\alpha\beta}(N+1) = g_{\alpha\beta}(N) + f_{\alpha *}(\mathbf{x})\, f_{\beta *}(\mathbf{x})$$

where $g_{\alpha\beta}$ is an element of the information matrix $\mathbf{G}$. The formulae were obtained with the assumption that measurement effectiveness is constant throughout the region X and equal to 1.

The calculation by formulae (8) permit to select at each stage a point $\mathbf{x}_L^*$ which minimizes the determinant of the covariant matrix $\mathbf{C}$. Selecting an arbitrary initial design and using recurrent formulae (8) to calculate the importance of the initial non-optimality will shrink the increasing N and for $N \to \infty$ the design obtained will be close enough to a D-optimal one.

Formulae (8) are seen not to include the output value y or the parameters of its distributions therefore the design can be calculated before the experiment. These formulae impose no constraints on the shape of the design domain X or the decomposition function vector $\mathbf{f}$.

The algorithms of making D-optimal plans with continuous design consist in:

1. The determination of the points where a D-optimal design is concentrated:

a) an arbitrary non-degenerate initial design is selected with the informational matrix $\mathbf{G}_{\Delta}$.

b) Eq. (8) yields the point $\mathbf{x}$ such that quadratic form $\mathbf{f}^T(\mathbf{x})\,\mathbf{G}^{-1}(N)\,\mathbf{f}(\mathbf{x})$ has a global maximum over the area X.

A search for a quadratic form global maximum is based on repetitive application of the local search from random points of the space X and subsequent selection of the maximum value from among the local maxima.

c) the global maximum point is included in the design and the matrix $\mathbf{G}$ is corrected in the following way;

d) the calculations by recurrent formulae (8) are continued to completion of the given number of cycles. At the first stage a number of cycles was practically found, two to three times the maximal number of the points where the D-optimal design is concentrated.

2. Determination of the observation repetition frequences in each point:

a) the initial information matrix is formed on the basis of a design which includes once every point that was determined at the first stage;

b) Eq. (8) yields the point $\mathbf{x}_*$ where the quadratic form

$$\mathbf{f}^T(\mathbf{x})\,\mathbf{G}^{-1}(N)\,\mathbf{f}(\mathbf{x})$$

is greater than in other points of the design. If this has equal values in several points of the design, any one may be selected;

c) the matrix $\mathbf{G}$ is corrected by Eq. (8);

d) the calculations by b) and c) are continued until the stoppage rule is satisfied. The stoppage occurs after the quantity δ reaches a specified value. The observation repetition frequency in the l-th point of the design is determined by the formula

$$\xi_l = \frac{\gamma_l + 1}{n + h} \tag{9}$$

where γ_l — is the number of times the global maximum hits the l-th point of the design;

n — is the number of cycles by recurrent formula (8);

h — is the number of points in the initial design.

By means of the algorithm above a computer compiled a catalog of D-optimal designs.

By this time, several papers have appeared on the dynamic object dentification according to experimental data as regards restoring pulse transient response ordinates with decomposing coefficients in the system of basis functions (by Lager, Chebyshev).

The regression model of linear dynamic object which connects input and output values in discrete moments can be written by the following formula:

$$y[n\Delta t] = \sum_{j=0}^{K} \Theta_j \sum_{m=0}^{l-1} f_j[m\Delta t]\, x\,[(n-m)\,\Delta t]\,\Delta t + E[n\Delta t] \tag{10}$$

where $f_j[m\Delta t]$ — values of some basic functions, $j = 0, \ldots, K$, $l = \frac{T_n}{\Delta t}$ where T_n — time of object memory, $E[n\Delta t]$ — value of an uncorrelated error which occurs in the output of a dynamic object.

Assume that

$$M\{E[n\Delta t]\} = 0$$

and

$$M\{E[i\Delta t]E[j\Delta t]\} = \begin{cases} \sigma^2, & i = j \\ 0, & i \neq j. \end{cases}$$

The problem of identification comes to the determination of Θ_j on the basic results obtained by observing input and output. Special test signals should be given if possible. In contradistinction to the case of pseudorandom binary signals (M-sequence) now a sequence will be synthesized providing us with a D-optimal criterion in regard to estimation of Θ_j. Such a sequence can be synthesized on the base of algorithm (8). As distinct from the model (1) the model (10) has values of the input signal in discrete moments — $x[n\Delta t]$, $x[(n-1)\Delta t], \ldots, x[(n-l+1)\Delta t]$, functions $f_j(\vec{x})$ at estimated coefficients are linear combinations of values of the input signal with some weighing coefficients.

In many experimental investigations it is necessary to estimate values of object or process output in these points of input parameter space where a direct measurement is impossible or difficult to realize in practice. A successful solution of the extrapolation problem is possible if the form of the regression equation doesn't change while passing from the field of investigation to field of input parameters. In the point $\mathbf{x}_e^L$ of extrapolation a variance of prediction according to regression equation is the following:

$$d(x_e) = \mathbf{f}^T(\mathbf{x}_e)\,\mathbf{C}(N)\,\mathbf{f}(\mathbf{x}_e) \tag{11}$$

where $\vec{\mathbf{x}}_e$ doesn't coincide with design space X. Optimal continuous extrapolation design providing us the minimum of predictive variance of an objective function in the given point can be synthesized on the base of the algorithm (8):

$$\mathbf{f}_*^T(\mathbf{x}_{*e})\,\mathbf{C}(N)\,\mathbf{f}_*(\mathbf{x}_{*e}) = \max_{\vec{\mathbf{x}} \in X} \mathbf{f}^T(\mathbf{x}_e)\,\mathbf{C}(N)\,\mathbf{f}(\mathbf{x}_e) \tag{12}$$

where $\mathbf{x}_e$ belongs to the field Z which does not coincide with the design space X in general case. Experiments proved that in most cases spectra of extrapolation designs perfectly coincide with spectra of D-optimal designs for corresponding forms of regression equation. This enables us to use the spectra of corresponding D-optimal designs as an initial approach to construct extrapolation designs.

Examples

1. For a polynomial of the form

$$\eta = \Theta_0 + \Theta_1 x_1 + \Theta_2 x_2 + \Theta_{11} x_1^2 + \Theta_{12} x_1 x_2 + \Theta_{22} x_2^2 \tag{13}$$

the algorithm for the construction of D-optimum designs was used in the case of an arbitrary design domain X with the appropriate change of that part of the program which organizes a search at the boundary of the domain.

For the polynomial in Eq. (13) a design was found for a randomly selected domain. The design is concentrated in six points located as shown in Fig. 1. The observation repetition frequency is the same in all points and equal to $\zeta_i = \dfrac{1}{6}$, $i = 1, 2, 3, 4, 5, 6$.

2. In accordance with (8) D-optimal sequence was synthetized to identify linear objects that have pulse transient response approximated by expression

$$\omega(\tau) = e^{-\tau} \sum_{i=1}^{3} \Theta_i \tau^i, \qquad \Delta t = \Delta\tau = 1\,. \tag{14}$$

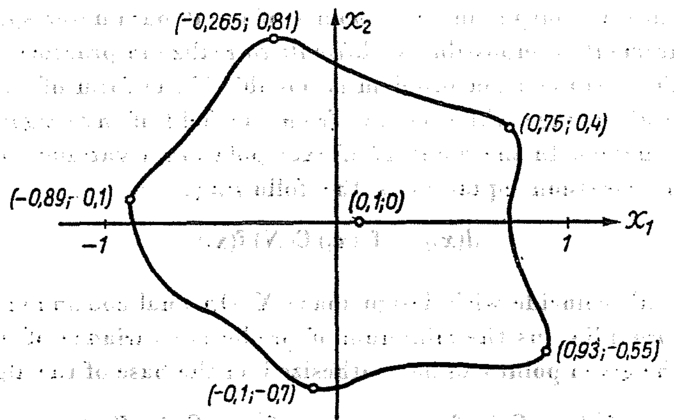

Fig. 1

Values $x[n-1]$, $x[n-2]$, $x[n-3]$, $x[n-4]$ were taken as factors of design. Spectrum of the D-optimal design for $-1 \leqslant x[n-i] \leqslant 1$ and the frequency of observations estimated by the number of cycles $N = 136$ are given in Table 1.

Table 1

Spectrum of D-optimal design				Frequencies
$x[n-1]$	$x[n-2]$	$x[n-3]$	$x[n-4]$	
+1	+1	+1	+1	0.31618
−1	+1	+1	+1	0.0808
+1	+1	−1	−1	0.308
+1	−1	−1	+1	0.294

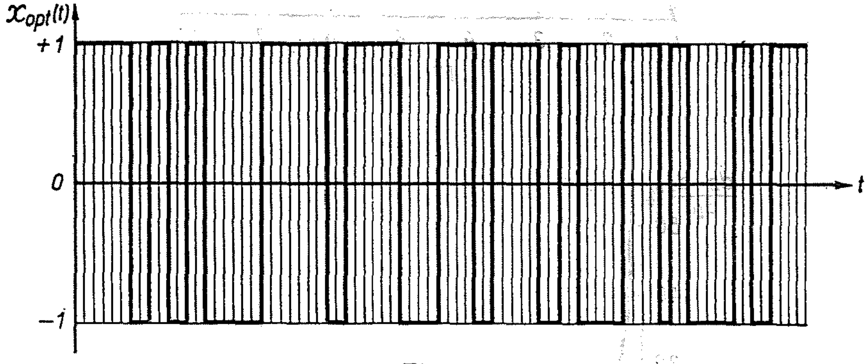

Fig. 2

An inital part of D-optimal test signal corresponding to the first 79 moments is shown in Fig. 2.

3. Continuous optimal extrapolation design the spectrum of which is obtained in points

$$\mathrm{I}(x = -1),\ \mathrm{II}(x = 0),\ \mathrm{III}(x = +1)$$

for polynomial model

$$y = \Theta_0 + \Theta_1 x + \Theta_2 x^2 \tag{15}$$

and admissible range of variable quantities $-1 < x < 1$ according to the algorithm (12). Fig. 3 shows the dependence of the frequencies of experiments on the depth of extrapolation x_e in these points. An advantage of optimal extrapolation design over the continuous D-optimal design as regards predictive variance of an objective function in the point x_e is given in Fig. 4. (d_0 — predictive variance calculated according to the D-optimal design, d_e — predictive variance calculated according to the extrapolation design).

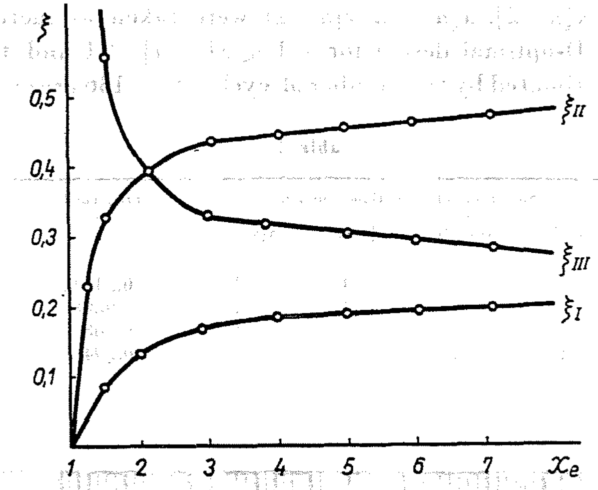

Fig. 3

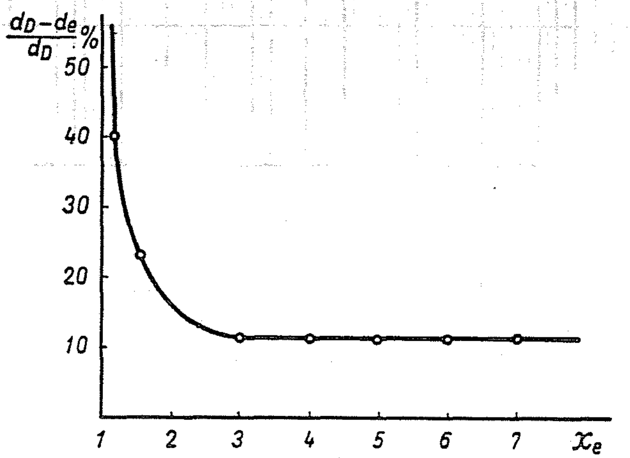

Fig. 4

Conclusion

Thus it is seen to be possible to synthetize enough complex experimental computer aided design satisfying different optimal criteria.

An investigator can calculate the design before doing the experiment if some algorithms and programs for different districts of data change and different kinds of models are available and thus raise the efficiency of solution of standard problems: identification and extrapolation.

Summary

The problem of synthesis of continuous experimental computer aided designs that consists in calculation of design spectrum and frequencies of experiment repetition in the spectrum points has been formulated.

Computational algorithms for synthesis of design in the problem of static and dynamic identification and extrapolation have been constructed. Some examples of synthesized designs for different optimal criteria and some additional conditions such as a form of model and a field of variable change have been given.

References

1. KRUG, G. K. Optimum Design of Experiments in Identification Problems. Reprints of Papers for IFAC Kyoto Symposium on Systems Engineering Approach to Computer Control, 1970.
2. KIEFER, J. Optimum designs in regression problems II. Ann. Math. Statist., 1961. 32, No 2.

Dr. G. K. KRUG, Energetic University, Moscow; 14 Krasnokazarmennaya Street. SSSR